Smart, Automated, and Reliable Security Service Platform for 6G

Deliverable D2.3
Final ROBUST-6G Architecture and ROBUST-6G Dataspace

ROBUST-6G project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101139068.

Date of delivery:	30/04/2026	Version:	1.0
Project reference:	101139068	Call:	HORIZON-JU-SNS-2023
Start date of project:	01/01/2024	Duration:	30 months

Document properties:

Document Number:	D2.3
Document Title:	Final ROBUST-6G Architecture and ROBUST-6G Dataspace
Editor(s):	Main editor(s) of the document
Authors:	Contributors and their organisations are listed below
Contractual Date of Delivery:	30/04/2026
Dissemination level:	PU[1]/SEN
Status:	Final
Version:	1.0
File Name:	ROBUST-6G D3.3_v1.0

Revision History

Revision	Date	Issued by	Description
0.1	12-01-2026	ROBUST-6G WP2	Initial ToC
0.2	20-03-2026	ROBUST-6G WP2	First complete draft
0.3	14-04-2026	ROBUST-6G WP2	Internal and external review
0.4	27-04-2026	ROBUST-6G WP2	Final complete draft after review
1.0	30-04-2026	ROBUST-6G WP2	Final version

Abstract

This document presents the Final ROBUST-6G Architecture and ROBUST-6G Dataspace. It describes an end-to-end, holistic security framework designed to support the scale, intelligence, and flexibility of future 6G systems by combining programmable pervasive monitoring, secure and distributed data management, autonomous zero-touch security management, trustworthy and sustainable AI services, and closed-loop physical-layer protection. A flexible Exposure Framework makes these capabilities available to external consumers through secure, well-defined APIs. The ROBUST-6G Dataspace, realized through the Data Management Platform, provides interoperable and privacy-preserving data sharing via a data fabric, data catalog, and data security services, enabling controlled access to telemetry and measurements across the edge–cloud continuum. By integrating AI-driven detection and prediction with automated orchestration and rapid actuation, the architecture enables proactive protection and fast mitigation across domains, from the radio interface to application-level security services.

Keywords

ROBUST-6G architecture, distributed AI-driven security, physical layer security, zero-touch security management, data management and governance, requirement assessment

[1] SEN = Sensitive, only members of the consortium (including the Commission Services). Limited under the conditions of the Grant Agreement

PU = Public

Disclaimer

List of Contributors

Participant	Short Name	Contributors
Ericsson Araştırma Geliştirme ve Bilişim Hizmetleri A.Ş.	EBY	Ömer Faruk Tuna, Mustafa Riza Akdeniz, Betül Güvenç Paltun
Telefónica Innovación Digital	TID	Lucia Cabanillas Rodriguez, Riccardo Nicolicchia, Ignacio Domínguez, Diego R. López
Universita Degli Studi Di Padova	UNIPD	Stefano Tomasin
École Nationale Supérieure de l'Électronique et de ses Applications	ENSEA	Arsenia Chorti, Sara Berri
Gohm Elektronik ve Bilişim Sanayi Ticaret Limited Şirketi	GOHM	Cem Ayyıldız, Fatih Emre YILDIZ
University College Dublin	UCD	Bart Siniarski
Nextworks	NXW	Pietro G. Giardina, Marco Ruta
THALES SIX GTS FRANCE SAS	THALES	Louis Cailliot, Dhouha Ayed
Universidad de Murcia	UMU	Alberto García Pérez, José María Jorquera Valero, Manuel Gil Pérez
AXON LOGIC IDIOTIKI KEFALAIOUXIKI ETAIREIA	AXON	Chih-Yang Pee, Wei Chuen Yau, Su Fong Chien, Charilaos C. Zarakovitis
Chalmers University of Technology	CHA	Masoom Rabbani, Azadeh Tabeshnezhad, Tommy Svensson
EURECOM	EUR	Marios Kountouris, Ioannis Pitsiorlas
Linkopings Universitet	LIU	Nikolaos Pappas, Eunjeong Jeong

List of Reviewers

Participant	Short Name	Contributors
Telefónica Innovación Digital	TID	Diego R. Lopez
THALES SIX GTS FRANCE SAS	THALES	Louis Cailliot, Dhouha Ayed

Executive Summary

This deliverable (D2.3) presents the final ROBUST-6G end-to-end security architecture and the ROBUST-6G Dataspace, specifying how future 6G systems can be protected at scale across distributed compute and network domains. Building on the project's requirements and previous architecture work, it consolidates a holistic framework that couples pervasive observability, secure and governed data sharing, AI-enabled security intelligence, and automated orchestration to enable proactive protection and rapid mitigation across multiple administrative domains.

At its boundary, the Exposure Framework provides secure, developer-friendly APIs that expose security capabilities such as threat analytics, SSLA-driven automation, and physical-layer security functions so that verticals and applications can consume security as a programmable service. Underneath, the Programmable Pervasive Monitoring layer continuously collects telemetry, incidents, alarms, and performance data from network functions and infrastructure, feeding the Data Management Platform that implements the ROBUST-6G Dataspace. The Dataspace combines a data fabric, catalog/knowledge graph, and data security services to enable interoperable, privacy-preserving data exchange, enforcing identity and policy-based access control while supporting both batch and streaming pipelines.

The architecture operationalizes AI nativeness through a Trustworthy & Sustainable AI Services layer that delivers privacy-preserving and decentralized learning, adversarial robustness, explainability, and energy-aware AI integrated with lifecycle management and model governance. These AI capabilities supply predictions, classifications, and confidence/explainability artifacts to the Zero-Touch Security Management layer, which translates high-level security intent into deployable controls and coordinates multi-domain security closed loops. The result is a security operating model that reduces human intervention while improving timeliness, consistency, and auditability of detection and response actions.

Finally, ROBUST-6G extends closed-loop automation down to the radio interface via a Physical Layer Security closed loop that monitors signal-level features, detects anomalies such as spoofing and jamming, and actuates fast countermeasures close to the RAN. Overall, the final architecture and Dataspace defined in this document provide a coherent blueprint for implementing smart, automated, and reliable security services for 6G—enabling controlled exposure of security capabilities, trustworthy data sharing, and AI-driven, closed-loop protection from the physical layer to applications.

Table of Contents

List of Tables

List of Figures

Acronyms and abbreviations

Term	Description
3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
AoA	Angle of Arrival
AoD	Angle of Departure
API	Application Programming Interface
CL	Closed Loop

CSV	Comma-Separated Values
DMP	Data Management Platform
ETSI	European Telecommunications Standards Institute
FL	Federated Learning
GSMA	Groupe Spéciale Mobile Association
JSON	JavaScript Object Notation
KER	Key Exploitable Result
KPI	Key Performance Indicator
LCM	Life-Cycle Management
LDAP	Lightweight Directory Access Protocol
MTL	Mapping Template Language
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
NEF	Network Exposure Function
NetSecaaS	Network Security as a Service
NF	Network Function
PLS	Physical Layer Security
PMP	Programmable Monitoring Platform
RDF	RDF Mapping Language
SEAL	Service Exposure Abstraction Layer
SOAR	Security Orchestration, Automation, and Response
SSLA	Security Service Level Agreement
XAI	Explainable AI
UE	User Equipment
ZTSM	Zero-Touch Security Management
ZTSP	Zero-Touch Security Platform

1 Introduction

As part of the next wave of mobile network innovations, 6G networks will radically reshape how digital services are connected, managed, and secured. Beyond higher performance, 6G is expected to introduce new classes of services and mission-critical applications. At the same time, this evolution broadens the attack surface and increases the complexity of security management, requiring security mechanisms that can operate seamlessly across heterogeneous infrastructures, multiple stakeholders, and highly dynamic service deployments. The ROBUST-6G project contributes to this transformation by developing an end-to-end (E2E), holistic security framework capable of supporting the extreme scale, intelligence, and flexibility of future 6G systems.

ROBUST-6G lays the foundation for a future where security is autonomous, adaptive, and embedded across every layer of the 6G network, ensuring that next-generation digital infrastructure is secure, resilient, and trustworthy from the ground up.

The goal of Robust-6G is to realize autonomous, zero-touch security for 6G, powered by distributed trustworthy AI, secure data management, pervasive monitoring, and exposure of security services to verticals and applications.

2 Final Version of the ROBUST-6G System Architecture

The ROBUST-6G architecture brings together programmable monitoring, secure data handling, autonomous security orchestration, trustworthy AI, and closed-loop physical-layer protection into a cohesive, end-to-end security solution tailored for 6G networks.

By integrating these capabilities under a flexible Exposure Framework, the project ensures that advanced security intelligence becomes accessible to verticals, developers, and applications across the 6G ecosystem.

The ROBUST-6G architecture is built around modular, interoperable enablers that together deliver proactive protection, fast mitigation, and cross-domain trust. Figure 1 presents a detailed overview of the project's functional architecture.

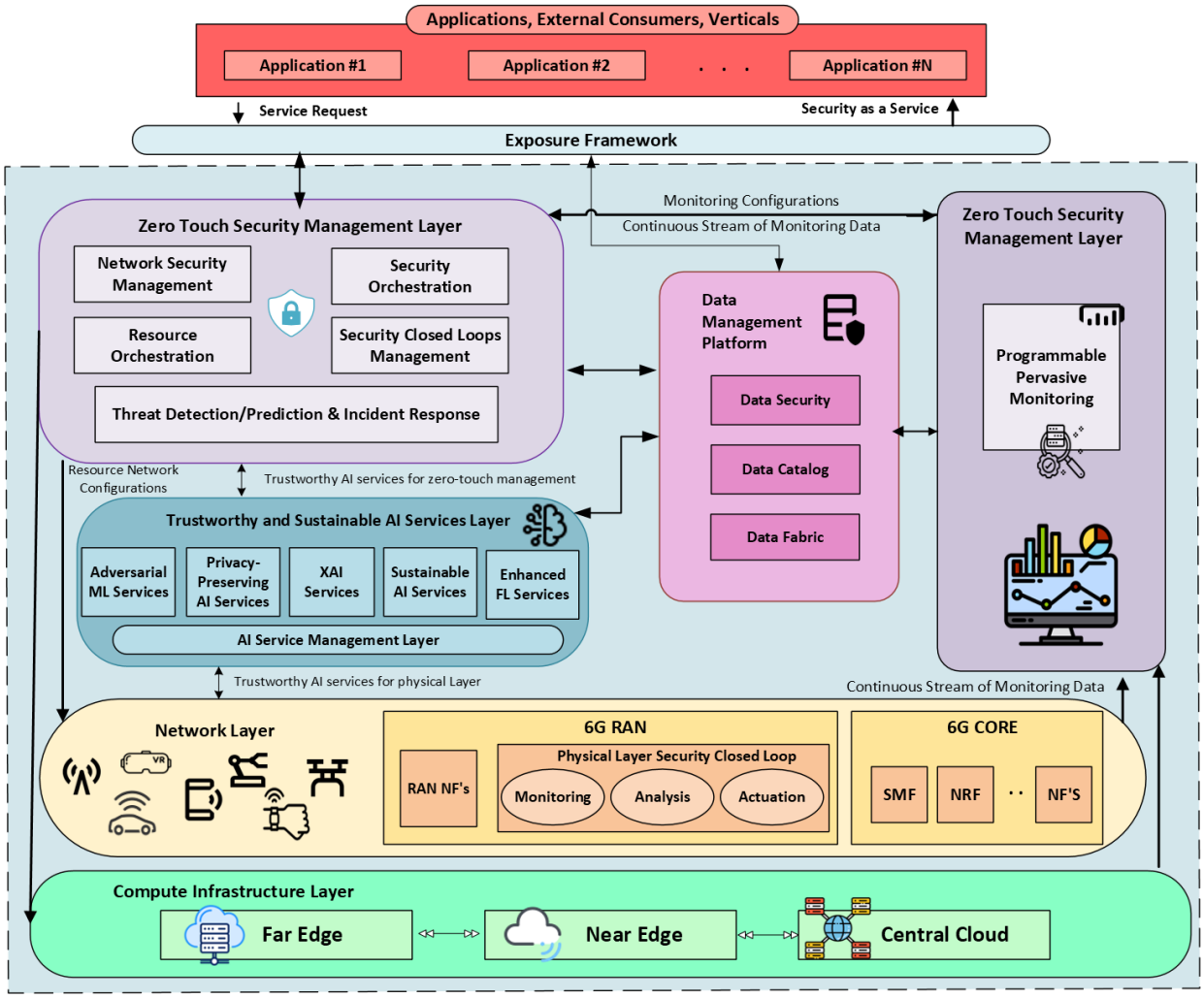


Figure 1: Functional Architecture of Robust-6G

2.1 Summary of ROBUST-6G Enablers

ROBUST-6G proposes various layers and enablers in the architecture. The Exposure Framework offers single contact point to external applications; Programmable Pervasive Monitoring handles continuous data collection mechanisms; Data Management Platform provides unified management capabilities in the data collected in the network or the data from external sources; Zero-touch Security Management Layer implements automation intelligence through resource orchestration and security closed loop management; Trustworthy and Sustainable AI Services Layer harmonizes AI services and model management for other layers in the architecture; and Physical Layer Security Closed Loop offers the domain-specific and agile intelligence for physical layer security. Each layer is summarized below, and detailed descriptions are given in Section 5.

2.1.1 Exposure Framework: The Entry Point to 6G Security Services

The Exposure Framework forms the external-facing interface of ROBUST-6G. It connects the 6G network's internal security capabilities with external consumers such as vertical industries, developers, applications, and other domains through secure and well-defined APIs.

Its key functions include:

- Exposing security services such as threat insights, trust scores, anomaly reports, and policy management tools.
- Handling service requests such as invoking mitigation procedures or requesting security policy updates.
- Enabling customization, for example integrating new services or onboarding domain-specific applications.

• Supporting use-case integration across industry verticals.

Because of its openness and extensibility, the Exposure Framework ensures that ROBUST-6G's advanced security intelligence is not isolated within the network, but actively usable by applications that require dynamic and trustworthy security guarantees.

2.1.2 Programmable Pervasive Monitoring: Continuous Awareness Across the Network

The Programmable Pervasive Monitoring Layer acts as the foundation of situational awareness for ROBUST-6G. It continuously observes the underlying infrastructure whether running at the far edge, near edge, or central cloud by collecting:

- Fault and performance measurements.
- Incident reports.
- Operational statistics.
- Alarms or anomaly indicators.

This layer (enabler) retrieves measurements directly from network functions (NFs) and infrastructure components, applying programmable monitoring logic that can adapt to new threats or emerging KPIs. All collected data is forwarded to the Data Management Platform, enabling Zero Touch Security Management Layer to detect early warning signs, refine AI models, and automate mitigation.

2.1.3 Data Management Platform: Unified, Secure, and Distributed Data Control

The Data Management Platform (DMP) serves as the central nervous system for data within ROBUST-6G. It aggregates, secures, organizes, and distributes data collected from multiple layers (components) of the 6G system, including:

- Physical-layer measurements.
- RAN and Core network telemetry.
- Application-level metrics.

Its responsibilities include:

- Managing secure data flow between sources and authorized consumers.
- Enabling distributed data management across diverse 6G environments.
- Enforcing privacy, integrity, and access controls.
- Supporting metadata organization and discovery.

The DMP comprises three core components:

- Data Fabric — ensures interoperability across heterogeneous data environments.
- Data Catalog — enables structured data discovery and metadata management.
- Data Security — governs authentication, authorization, and confidentiality.

It acts as the data provider for key functional layers (components) such as Zero-Touch Security Management, Trustworthy & Sustainable AI Services, and Physical Layer Security Closed Loops.

2.1.4 Zero-Touch Security Management Layer: Orchestrating Autonomous Protection

The Zero-Touch Security Management (ZTSM) Layer is the architectural "brain" coordinating security across the 6G system. It transforms traditional manual security management into an automated, intelligence-driven process capable of predicting and mitigating threats in real time.

Its main components include:

- Network Security Management – defines and enforces security policies and SSLAs.
- Security Orchestration – deploys and coordinates security functions across distributed nodes.
- Resource Orchestration – allocates compute/network resources for security operations.
- Security Closed-Loop Management – automates detection, analysis, decision-making, and actuation.
- Threat Detection/Prediction & Incident Response – uses AI/ML to identify, classify, and respond to attacks.

Through multiple autonomous closed loops, ZTSM can detect deviations, predict potential risks, and initiate corrective measures, all without human intervention. This automation is central to the zero-touch philosophy envisioned for 6G.

2.1.5 Trustworthy & Sustainable AI Services Layer: Making AI Reliable for 6G Security

AI is essential to the intelligence and automation of 6G, but AI-based security models must themselves be protected and trustworthy. The Trustworthy & Sustainable AI Services Layer ensures that all AI models used across ROBUST-6G are secure, interpretable, privacy-preserving, and resource-efficient.

It provides the following AI services:

- Enhanced Federated Learning (FL) — enabling training across distributed datasets without sharing raw data.
- Adversarial ML Services — protecting models from manipulation or adversarial attacks.
- Explainable AI (XAI) — ensuring transparency and interpretability of decisions.
- Privacy-Preserving AI — employing mechanisms such as differential privacy and secure computation.
- Sustainable AI Services — enabling low-complexity model design, energy-aware training and inference, and resource-efficient AI pipelines across the edge–cloud continuum.

By combining trustworthiness, privacy, robustness, and sustainability, this layer provides the foundation for secure, responsible, and future-proof AI operations supporting Zero-Touch Security Management and Physical Layer Security in 6G networks.

2.1.6 Physical Layer Security Closed-Loop Layer: Protecting the Air Interface

6G introduces new threats to the physical layer, making Physical Layer Security an essential element of the ROBUST-6G architecture. This layer implements a local closed-loop system to protect the air interface, integrating:

- Monitoring — gathering signal-level data from UEs and base stations (BSs).
- Analysis — applying AI/ML to detect physical-layer anomalies like spoofing, jamming, or abnormal channel patterns.
- Actuation — enforcing rapid mitigation, such as power adjustments, beam adaptation, or localized blocking.

Deployed close to the RAN, either within the base station or the edge node, this layer ensures ultra-fast detection and response to threats originating from the radio environment.

2.2 Security Capability Exposure

Security Capability Exposure is a key component of the ROBUST-6G architecture. It shows how external applications can benefit from ROBUST-6G security features by using Network-Security-as-a-Service capabilities via standardized, developer-friendly APIs.

This paradigm shift transforms traditional network security from an intricate, infrastructure-dependent service into a more accessible, programmable capability. Application developers and enterprises can integrate it seamlessly into their solutions without requiring in-depth security expertise.

Based on the foundation established in D2.2 [R6G24-D22], the NetSecaaS framework has evolved to provide a comprehensive abstraction layer between the security-specific mechanisms developed within ROBUST-6G and external consumers. Drawing inspiration from the GSMA Open Gateway initiative [GSM26], the framework extends its scope to include AI-driven security services, zero-touch management capabilities, and the security mechanisms developed across the project's technical work packages.

The NetSecaaS architecture uses a multi-tiered approach to capability exposure. At the northbound interface, Service APIs offer intuitive access to security functionalities for capability discovery, data exposure and security function access. These APIs abstract the underlying complexity of the ROBUST-6G platform, presenting security services in terms that are familiar to application developers by focusing on outcomes rather than implementation details.

To enable effective and secure exposure of security capabilities, the component is integrated with the Data Management Platform, which defines access policies for southbound ROBUST-6G capabilities and data, thereby serving as an enforcement layer. In terms of data management, NetSecaaS interacts with the platform's Data Fabric component to enable two-way data flows. This supports the ingestion of high-level security queries from consumers and the delivery of security data, analytics and threat intelligence back to them.

These integrations ensure that exposed capabilities remain based on up-to-date network and security data while maintaining strict governance and access control policies.

The Transformation Function within this integration layer plays a pivotal role in formally mapping the northbound (NBI) to southbound (SBI) interfaces. It bridges the semantic gap between the high-level security requirements expressed through CAMARA-style APIs and the low-level configurations required by ROBUST-6G components. This function orchestrates workflows involving one or more internal services, such as coordinating with the zero-touch security management platform for orchestration, while still presenting a unified, simplified interface to external consumers.

As previously mentioned, the development of security-focused APIs follows the open-source methodology of the CAMARA project, ensuring interoperability and alignment with industry standards. ROBUST-6G contributes novel API definitions that expose capabilities unique to 6G security requirements, including:

- AI-driven Threat Analytics API: providing access to explanation and classification services powered by the platform's trustworthy AI models.
- Zero-touch Automation API: enabling users to define Security Service Level Agreements (SSLAs) with high-level queries, triggering automated detection and mitigation workflows.
- Physical Layer Security API: exposing authentication, key agreement and anomaly detection capabilities operating at the physical layer.

Each API is designed according to security-by-design principles and incorporates OAuth-based authentication, fine-grained authorization policies and comprehensive audit logging.

2.3 Trustworthiness in AI-Nativeness

In ROBUST-6G Trustworthiness is mainly enabled through Trustworthy & Sustainable AI Services Layer and its interactions with other layers. This layer provides the methodological and architectural foundations required to ensure that AI-driven security functionalities are robust, privacy-preserving, sustainable, explainable, and effective. These domain-agnostic services and KPIs are integrated for the consumption of ZTSM and Physical Layer Security Closed Loop. This layer addresses the Objective 3 of ROBUST-6G:

"To develop methodologies for ensuring that AI-driven security functionalities are robust, sustainable (in terms of energy efficiency), explainable, effective (in terms of performance) and preserving privacy in all aspects."

From a distributed intelligence perspective, this layer delivers a federated and **decentralized learning framework** that enables collaborative AI model training across heterogeneous 6G domains without sharing raw data. As detailed in D3.3 [R6G26-D33], the framework supports centralized, hierarchical, and fully decentralized FL configurations, allowing security intelligence to be trained and updated close to the data sources across the edge–cloud continuum. This directly satisfies the KER-1 requirement for secure and privacy-preserving distributed intelligence in multi-tenant 6G environments and is reflected architecturally in the Enhanced Federated Learning services exposed by the Trustworthy & Sustainable AI Services Layer.

To ensure **privacy** and **trustworthiness**, this layer integrates privacy-preserving mechanisms such as differential privacy, secure aggregation, and encrypted model updates into the federated learning lifecycle. These mechanisms are coordinated through a dedicated privacy management function, ensuring that collaborative learning remains compliant with data protection constraints while still enabling effective model convergence. This design is fully aligned with the Data Management Platform and Privacy-Preserving AI services in the final architecture, which jointly enforce data confidentiality, access control, and secure data/model exchange.

Services offered by this layer also address the **robustness** and the security of AI models by incorporating adversarial-aware learning and robustness mechanisms against poisoning and manipulation attacks. As described in D3.2 [R6G25-D32] and D3.3 [R6G26-D33], the framework employs robust aggregation strategies and anomaly detection on model updates to reduce the impact of malicious participants in collaborative learning. These capabilities are architecturally linked to the Threat Detection and Prediction functions of the Zero-Touch Security Management layer, enabling closed-loop, AI-driven mitigation actions.

With respect to **sustainability**, sustainable AI services introduce energy-aware AI design principles and mechanisms that reduce the computational and energy footprint of security analytics. This includes energy-efficient model architectures, post-training optimization techniques, and adaptive client participation strategies in federated learning. These contributions directly support the project's sustainability targets and are realized in the architecture through the Sustainable AI Services that operate across far-edge, near-edge, and cloud environments.

Finally, **explainability** and **accountability** are embedded in the design from the outset. Explainable AI mechanisms, including uncertainty estimation and concept-level explanations, are integrated into both training and inference phases to support transparency of security decisions. This ensures that AI-driven threat detection and response actions remain interpretable and auditable, in line with O3 and KER-1 requirements. In the architecture, these capabilities are exposed via dedicated XAI services and consumed by ZTSM closed loops to support trustworthy autonomous decision-making.

2.4 Zero touch management

The minimisation of human intervention is one of the main goals of ROBUST-6G. With respect to the automatic guarantee of security levels, the zero-touch automation is reached through a dedicated (zero-touch) security management layer (ZTSML) in charge of orchestrating and automating security across different network layers and segments. This security layer is designed to address objective 4 of ROBUST-6G:

"Automatic, zero-touch, security, and resource management for trusted and certified services among multiple stakeholders in distributed dynamic scenarios"

To fulfil the objective, the ZTSML defines a set of functionalities that represent the pillars of the Zero-Touch security Platform i.e., the implementation of the ZTSML, reported in WP4 deliverables (D4.1 [R6G24-D41], D4.3 [R6G25-D43], and D4.4 [R6G26-D44]) and briefly summarised in Section 5.3:

- **Security Orchestration.** Refers to the orchestration of security services properly defined through specific information models.
- **Resource Orchestration.** Orchestration of network and computing resources for security purposes.
- **Network Security Management.** Determines configuration to guarantee network security.

Security and Resource Orchestration is the *first pillar* of the Zero-Touch Security Platform (ZTSP).

- **Programmable Pervasive Monitoring.** Provides the capabilities to collect heterogeneous data from multiple data sources belonging to different network segments and layers. Historical and real-time data exposure are also crucial for detecting and predicting anomalies and deriving security insights from the underlying system.

Programmable Pervasive Monitoring is the *second pillar* of the ZTSP.

- **Threat Detection/Prediction & Incident Response.** A set of mechanisms for threat detection, prediction, and remediation (incident response) that consumes the monitoring data provided by the Programmable Pervasive Monitoring. Threat prediction is usually AI-driven, while detection and response can be both rule/policy-based. The AI services can be provided by the Trustworthy and Sustainable AI Service Layer.

This set of mechanisms is the *third pillar* of ZTSP.

- **Security Closed-Loop Management.** Provides management functionalities for the security closed-loops i.e., the core concept on which the security automation is based. Security closed-loops (CL) are built by using data from Programmable Pervasive Monitoring (second pillar), analysing it through detection/prediction mechanisms, and remediating in case of anomaly (third pillar). In ROBUST-6G, CLs are orchestrated and part of the security services (first pillar). The CL management encompasses mechanisms to manage the lifecycle of multiple CL entities.

The closed-loop-based security automation is the *fourth pillar* of ZTSP.

2.5 E2E security architecture

To achieve its objectives, the end-to-end (E2E) security architecture of ROBUST-6G leverages Zero-Touch Management (ZTM) by integrating security requirements from the earliest phases of deployments. This architecture addresses two critical challenges: horizontal complexity, originating from the diversity of orchestrated domains (including core networks, edge computing, and radio access networks, or RAN), and vertical complexity, coming from the hierarchical structure of meta-orchestrators responsible for resources, monitoring, and security.

The proposed solution for managing vertical complexity focuses on the Zero-Touch Security Management Layer (ZTSML), which decomposes high-level security needs into deployable configurations or software components for lower-layer orchestrators. The main component of this layer is the Zero-Touch Security Orchestrator (ZTSO), which distributes these decomposed requirements to domain-specific orchestrators, allowing each one of them to translate and implement security measures according to its domain context. To ensure continuous compliance, security monitoring requirements are forwarded to the Programmable Monitoring Platform (PMP), a cross-domain observation framework.

For horizontal complexity, where each orchestrated domain enforces heterogeneous networks, software, and IT technologies, the architecture introduces an interoperable communication OpenC2 framework. In this way, each domain resource orchestrator is responsible for translating security services requirements into domain-specific deployments, while the ZTSML ensures the security posture maintenance. In the other way around, each domain is responsible for monitoring its own security functions, reporting status updates and events to the PMP. This multi-domain visibility enables the ZTSO to maintain the security posture compliance after deployments.

Security violations are mitigated through a closed-loop process that combines Cyber Threat Intelligence (CTI), Security Information and Event Management (SIEM), and Security Orchestration, Automation, and Response (SOAR) frameworks. When anomalies are detected, these systems analyse events against the established security policies to compute and deploy countermeasures, either as reconfigurations or additional security functions, via the ZTSO in the same way as described above. This iterative cycle of deployment, observation, analysis, and mitigation forms the backbone of ROBUST-6G's E2E security architecture, ensuring adaptive and resilient protection across all orchestrated domains.

Combining these solutions, the E2E architecture fulfils the objectives of providing E2E security guarantees across the heterogeneity and dynamicity of technologies envisaged in 6G and enabling autonomous approaches to security deployment in an environment with several controlled closed-loops.

To address both vertical and horizontal complexity, the architecture integrates a suite of AI/ML technologies designed to enhance adaptability, automation, and intelligence in security management. For vertical complexity, the Zero-Touch Security Orchestrator (ZTSO) employs a combination of semantic AI and symbolic reasoning. This component dynamically computes the optimal decomposition of security services to fulfill high-level policies while adapting the heterogeneity of lower-layer orchestrators. By constructing an abstract representation of key concepts, such as security activities, functions, and target environments, linked through functional and non-functional capabilities, the system decouples these abstractions from specific tools, platforms, or frameworks. This approach establishes a generalized, extensible foundation for defining and managing security services within the ZTSO, ensuring flexibility and scalability across diverse orchestration layers.

Within the Zero-Touch Security Management Layer (ZTSML), a generative AI service enables the automated and dynamic creation of both remediation playbooks and monitoring rules. Through contextualized prompt engineering, this service generates SIGMA rules, derived from security policies submitted to the ZTSO, and forwards them to the Programmable Monitoring Platform (PMP) for deployment across orchestrated domains. Additionally, the generative AI dynamically produces CACAO playbooks in response to real-time security alerts raised by the PMP, ensuring swift and adaptive incident response. Those playbooks can immediately be enforced by SOAR framework to mitigate security threats.

To further strengthen proactive security, machine learning (ML) models for threat mitigation and prediction are embedded within the security functions catalog of the Zero-Touch Security Platform (ZTSP). These models, ready for deployment as countermeasures or policy components, analyze telemetry and network traffic data captured by the PMP and stored in the data fabric. By retrieving and processing this data, the models generate real-time analytical reports for the ZTSO, enabling data-driven decision-making and improving the mean time to respond (MTTR) to security threats.

The implementation of these AI technologies dedicated to the E2E security architecture fulfills the objectives of the development of techniques, protocols within a novel 6G security architecture for the integration of AI/ML in multiple security related scenarios, providing adaptive response to incidents, and improving AI security.

Proactive security, combined with continuous monitoring and testing of physical infrastructure, significantly reduces the risk of system breaches. This approach implies existing countermeasures, designed and implemented for a wide range of physical platforms. Within the end-to-end (E2E) architecture, addressing this challenge requires a well-defined communication protocol and standardized interfaces between the Zero-Touch Security Platform (ZTSP) and physical components. To achieve this, prior discussions have been led during the project to address this matter in the form of a translation mechanism that converts OpenC2 commands into device-specific languages compatible with the physical controllers embedded in hardware components. These controllers are responsible for implementing and executing security functions at the physical layer. By establishing OpenC2-based interfaces, where the hardware acts as OpenC2 consumers, the ZTSP can deploy mitigations directly to the physical infrastructure.

This solution of the E2E architecture of the project fulfills the objective of defending against potential threats at the physical layer (rogue terminal/network identification) and cyberattacks (eavesdropping, jamming) to secure both users and networks.

2.6 Physical Layer Security

The ROBUST-6G architecture incorporates Physical Layer Security (PLS) through a structured framework known as the Physical Layer Closed Loop (PLCL). This system functions as a continuous feedback cycle that integrates physical radio data with high-level security management to ensure network integrity.

The framework is organized into three primary functional groups:

Monitoring: Data Gathering and Environment Sensing -- The process begins with the collection of diverse data points from both the hardware and management levels. On one hand, the system tracks real-time radio environment metrics, such as channel conditions and signal characteristics. On the other hand, it incorporates broader contextual data from security management layers, including localization details and orchestration alerts. This phase is dedicated to establishing a baseline of the physical environment and generating the datasets necessary for deeper investigation.

Analysis: Evaluation and Threat Detection -- Once the data is gathered, the system moves into an assessment phase. Here, the architecture evaluates the overall trustworthiness of the physical layer. It specifically looks for anomalies and identifies potential electromagnetic threats, such as jamming or unauthorized signal interference. By measuring information leakage and sensing reliability, the system determines whether the current state of the network meets the required security standards.

Actuation: Response and Resource Control -- The cycle concludes with decisive actions based on the preceding analysis. The system dynamically adjusts security features, enabling or disabling specific authentication and encryption protocols depending on the detected threat level. Beyond security settings, this phase also manages technical radio parameters—such as power allocation and modulation—to maintain performance. These updates are then fed back into the network, closing the loop and allowing for continuous, automated adaptation to new security challenges.

The Physical Layer Closed Loop serves as the core architectural engine that translates the high-level security ambitions of the project into a tangible, self-healing network environment. It acts as a continuous bridge between raw physical phenomena and the intelligent management layer, ensuring that the 6G network remains resilient against a wide array of sophisticated radio-level threats while meeting the strict demands of low latency and energy efficiency.

A full description of the close loop is available in D5.2, with details on the novel solutions developed in ROBUST-6G for each block. This part of the architecture implements the objectives of ROBUST-6G.

The implementation begins within the monitoring phase, which directly supports the project's goal of identifying and classifying diverse attacks such as jamming, false base stations, and malicious pilot contamination. By gathering real-time radio metrics and hardware signatures, the framework creates a rich dataset that enables the AI-driven techniques specified in the project scope. This granular data collection is essential for establishing a baseline of sensing integrity, allowing the system to distinguish legitimate signal variations from intentional injection or synchronization attacks. In doing so, the monitoring phase provides the necessary raw material for the AI/ML models to achieve the high detection accuracy required for 6G trustworthiness.

Once the data is ingested, the analysis phase fulfils the objective of providing measurable security guarantees and general trustworthiness. At this stage, the system employs the smart techniques mentioned in the project goals to evaluate the overall security state of the physical layer. By measuring parameters like information leakage and detecting generalized anomalies across both physical and hardware layers, the architecture can determine if the current connection meets the safety requirements for distributed MIMO and infrastructure integrity. This analytical process is the foundation for the resilient, trust-based environment the project aims to build, ensuring that privacy is maintained even at the signal layer through advances such as channel charting.

The final stage of the loop, actuation, is where the project's novel physical layer security schemes are deployed in real-time. This phase directly addresses the need for low-latency and low-footprint protocols by dynamically adjusting security features like authentication and encryption at the signal level. By leveraging advanced radio technologies such as Reconfigurable Intelligent Surfaces and Massive MIMO, the actuation phase can mitigate eavesdropping or jamming through precise power allocation and beamforming. This automated response capability is critical for supporting challenging use cases like the Industrial IoT, ultimately facilitating the zero-touch security solutions that characterize the broader vision of the architecture. Through this integrated cycle of sensing, thinking, and acting, the framework provides a comprehensive solution for privacy and resilience at the lower layers of the 6G stack.

2.7 Unified, Secure, and Distributed Data Management

The ROBUST-6G Data Management Platform provides the core capabilities required to enable centralized, integrated, and governed access to the datasets exposed by data sources within the ROBUST-6G ecosystem. As a key component of the ROBUST-6G architecture, the platform enables interoperable data exchange among heterogeneous data providers and consumers, effectively breaking data silos and supporting the implementation of the project's diverse use cases.

The platform is designed to deliver unified access to a wide variety of data types, including structured and semi-structured datasets. It supports data consumption by applications of different nature, ranging from visualization tools and Machine Learning (ML) models to operational applications that underpin the ROBUST-6G platform.

To meet these requirements, the proposed architecture adopts a semantically driven data management approach based on knowledge graph and semantic web technologies. The knowledge graph acts as the semantic backbone of the platform, providing structured and meaningful representations of interconnected data that are ingested, integrated, and managed across the platform. In addition, the Data Management Platform incorporates data engineering mechanisms to support the ingestion, transformation, integration, and consumption of datasets.

At the core of the platform, the ROBUST-6G Knowledge Graph serves as a metadata layer that orchestrates data integration pipelines while simultaneously enforcing data governance strategies and policies defined within the project.

Based on these requirements, we define high-level architecture for the ROBUST-6G Data Management Platform. The architecture consists of several functional building blocks that interact with one another through the ROBUST-6G Knowledge Graph. The main building blocks are described below:

- **Data Governance**: Provides capabilities for data discovery and security management to ensure compliance with the project's governance strategies. Data discovery is enabled through a data catalog service that organizes datasets within the knowledge graph and links them to business concepts and relevant metadata. Data security services manage access control policies, ensuring that only authorized consumers can access datasets under defined constraints. By combining policy definitions with catalog metadata, access decisions can account for privacy, sensitivity, and usage restrictions.
- **Data Ingestion:** Handles the collection of datasets from data sources through dedicated connectors tailored to the protocols supported by each source. Depending on the source capabilities, data ingestion can be performed using pull-based mechanisms (batch ingestion) or push-based mechanisms (streaming ingestion).
- **Data Processing:** Provides mechanisms for transforming, integrating, and cleansing ingested data to make it suitable for consumption. Raw structured and semi-structured datasets can be normalized into a common data model, enabling interoperability, cross-dataset integration, and efficient reuse in downstream processing and data exchange.
- **Data Access:** Enables data consumers to access datasets managed by the platform. While some consumers may directly interact with the integrated data through the ROBUST-6G Knowledge Graph, most applications require data to be exposed through interfaces that match their specific formats, data models, and access protocols.

The ROBUST-6G Data Management Platform directly contributes to the fulfilment of Objective 2 by enabling secure, governed, and end-to-end data flows across heterogeneous 6G environments. By providing centralized yet flexible data management, semantic integration through the ROBUST-6G Knowledge Graph, and controlled data access mechanisms, the platform ensures that security-related data can be reliably ingested, processed, shared, and consumed.

3 Differentiated Service Examples

3.1 Explainable AI

The ROBUST-6G architecture provides an end-to-end framework for ensuring and enhancing the trustworthiness of AI/ML algorithms deployed in autonomous 6G security and management services. As AI becomes a core enabler of zero-touch network operation, AI-driven decisions and actions must be reliable, transparent, robust to distribution shifts and adversaries, privacy-preserving, and continuously verified across the full model lifecycle (data collection → training → deployment → monitoring → adaptation).

To improve transparency and interpretability, the ROBUST-6G architecture integrates XAI capabilities and mechanisms, which provide interpretable insights and evidence behind model predictions and decision-making processes. This includes feature-attribution explanations (e.g., SHAP), counterfactuals ("what minimal change would flip the decision"), and, where appropriate, intrinsically interpretable policy models (rules/monotonic constraints). Explanations are attached to security decisions (anomaly alerts, slice-level mitigation, access control) to support operator validation, policy compliance audits, and root-cause analysis during incident response.

ROBUST-6G further enhances reliability through confidence-aware inference that quantifies prediction uncertainty and produces calibrated trust metrics (e.g., ensembles/MC-dropout uncertainty, prediction intervals) for AI outputs. These metrics enable the identification of unreliable or high-risk predictions, supporting automated decision-making processes and reducing false alarms in security monitoring systems.

To continuously verify model validity under non-stationary 6G conditions, ROBUST-6G incorporates distribution-shift monitoring and conformal prediction-based safeguards. Conformal mechanisms provide coverage guarantees (under stated assumptions) and yield actionable signals when those guarantees begin to fail, which is critical for detecting concept drift, data drift, and degradation from incremental/online learning. Complementing conformal monitoring, the architecture tracks data quality and integrity (feature completeness, schema changes, sensor faults), KPIs such as precision/recall for threats, time-to-detect, time-to-mitigate, etc., and calibration drift. When degradation is detected, the system can trigger autonomous adaptation workflows within closed-loop security control, such as safe retraining with curated replay buffers, model rollback to a previously validated version, or switching to a robust fallback policy. Finally, ROBUST-6G operationalizes trustworthiness via governance and secure MLOps mechanisms integrated into the architecture: dataset/model versioning, lineage tracking, reproducible training pipelines, continuous evaluation against adversarial and stress-test suites, and secure deployment controls (signed artifacts, attestation, runtime policy enforcement). These mechanisms ensure that explainability and confidence signals are not merely informative but are used as control inputs to enforce safe decision-making, auditable behaviour, and dependable 6G autonomy.

Integrating Explainable AI (XAI) into 6G networks is crucial for building transparent, interpretable, and trustworthy AI-driven systems. Traditional AI models function as opaque "black boxes," making their reasoning difficult to understand which is a significant concern in critical applications like cybersecurity, autonomous systems, and network anomaly detection. To ensure AI performs reliably as network conditions change, confidence-aware evaluation approaches are necessary.

Within the ROBUST-6G architecture, XAI serves as a foundational component for enabling reliable, sustainable, and secure AI operations. It forms an essential part of the Trustworthy AI framework, supporting both robust and privacy-preserving AI capabilities. By improving transparency in decision-making, XAI strengthens the ability to identify adversarial attacks and maintain resilient security defences.

The XAI module comprises four main services that support this architecture. First, incoming requests are processed through the ROBUST-6G network layers, where explainability requirements are defined and operational context is established for subsequent model deployment and assessment. Second, a model selection process determines optimal architectures by balancing interpretability constraints with performance requirements. Third, model training is conducted using transparency-preserving algorithms and regularisation techniques, with context-aware optimisation ensuring that explainability is embedded throughout the learning process. Fourth, model outputs are generated as primary inference results, which then feed into an

explainability assessment pipeline for comprehensive interpretability evaluation. Finally, the Explainability Assessment component orchestrates a multi-faceted evaluation by coordinating various analytical services. XAI also enhances threat detection and response by providing interpretable insights into potential risks, which can inform adaptive incident response strategies, reduce false positives in alarm systems, and improve overall detection accuracy. Furthermore, it can assist in analysing Denial of Service (DoS) attacks and anomaly alerts while strengthening authentication mechanisms by identifying adversarial access attempts.

3.2 Sustainable AI

A core objective of the Trustworthy & Sustainable AI Services Layer is to minimize the energy footprint of AI-driven security operations across the 6G edge–cloud continuum. In this regard, semantics-aware task scheduling introduces goal-oriented intelligence into the orchestration of distributed learning tasks at the network edge.

In conventional federated learning (FL) deployments, participating agents (e.g., edge devices, base stations, or network functions) contribute model updates at regular intervals, regardless of how informative those updates are. This uniform participation leads to unnecessary computation and communication, consuming energy without proportionally advancing the learning objective.

Semantics-aware task scheduling addresses this inefficiency by enabling agents in a multi-agent system to decide what and when to transmit, and consequently what and when to update, produce, or actuate. These decisions are driven by the significance of the information to be communicated and the circumstances under which the communication takes place, jointly capturing the timeliness, relevance, and marginal value of each agent's contribution toward the collective learning goal.

Through this approach, the system can quantify and prioritize all energy-consuming activities at the edge, including local model training, gradient computation, and uplink transmission, according to their expected contribution to the predefined learning objective. By selecting only the actions that advance the goal faster and more accurately, semantics-aware scheduling suppresses redundant or low-value transmissions, thereby reducing overall energy consumption while maintaining or even improving model convergence and accuracy.

The following studies, conducted within the scope of ROBUST-6G, advance semantics-aware scheduling for energy-efficient FL:

a) Battery-aware cyclic scheduling: A framework that groups energy-harvesting FL clients by their available energy and schedules them sequentially, reducing redundant computations through deferred local training.
b) Selective participation for energy-efficient FL: An extension that allows clients to skip non-essential training opportunities, approximately halving computation costs without compromising convergence, with formal convergence guarantees.
c) Lightweight semantics-aware client selection: A scheduling policy that assesses the informational value of each client's potential update using a single forward pass, filtering out redundant participants before expensive training is initiated.

The energy-aware cyclic scheduling logic is implemented within the training interface of the DFL framework, where it governs client participation and resource-aware orchestration of distributed training rounds. Trained models produced through this process are stored in the GMR, making them accessible to other architectural components for integration into training or inference pipelines targeting specific event detection problems.

A fundamental feature of 6G networks is that the envisioned integrated and pervasive AI is sustainable and energy-efficient directly from the ML model design. Towards this direction, ROBUST-6G investigated spiking neural networks (SNNs), a form of neural computing closely mimicking how the human brain works. While the energy consumption of these NNs can be orders of magnitude lower than traditional ANNs, SNNs currently perform well by processing only event-driven inputs. Luckily, these kinds of signals are often of interest in modern networks, for security (e.g., attacks detection), and more in general, event detection purposes. During the ROBUST-6G project span, UNIPD worked on

a) SNN architecture selection, to improve model's performance (e.g., classification tasks).

b) Surrogate gradient training implementations, with the library snnTorch.
c) A novel custom training method based on the alternating direction method of multipliers (ADMM).

The SNN simulator, specifically, point b) listed above (optimized also with point a)) is reachable in the ROBUST-6G architecture thanks to the training interface provided by the DFL framework, into which an optimized SNN model is provided, ready for training. Trained models are also available in the global model repository (GMR), and the code (available on GitHub) can be used by other modules/layers to integrate training/inference targeting specific event detection problems.

3.3 Zero touch orchestration

The Zero-Touch Security Management Layer of ROBUST-6G (mentioned in Section 2.4) interacts with the Exposure Framework through well-defined northbound interfaces that expose security capabilities in an intent-driven and service-oriented manner. Verticals, service providers, and system administrators interact with the platform by submitting Security Service Level Agreements (SSLAs) and security policies, which express desired security outcomes at the business level. These intents are exposed via the exposure framework without requiring users to understand infrastructure-specific details. Internally, the Zero-Touch Security Orchestrator (ZTSO) consumes these inputs, translates SSLAs into Security Level Objectives (SLOs), and produces orchestration plans that drive the deployment of security services and automations across the cloud–edge continuum.

Through direct interaction with the Catalogue Management module, end consumers can onboard their infrastructures, execution environments, and proprietary security functions such as firewalls or monitoring tools. Authorised users may also access the security orchestration ontology and knowledge graph, enabling advanced governance tasks such as extending security concepts, updating relationships, or aligning security knowledge with multi-domain scenarios. This exposure allows users to understand what security capabilities are available and how they can be composed, while the Zero-Touch Security Management Layer retains responsibility for selecting, configuring, and orchestrating the appropriate resources.

The interaction with the exposure framework allows the Zero-Touch Security Management Layer to expose security posture, compliance status, and incident awareness. Data collected by the Programmable Monitoring Platform, semantic-aware anomaly detection modules, rule-based threat detection, and AI-driven prediction and mitigation components is processed internally through security closed loops. The resulting outcomes, such as detected anomalies, predicted threats, SSLA compliance indicators, and applied mitigation actions, are surfaced to end users as high-level security events and status views. This provides transparency and explainability while shielding users from the underlying monitoring tools, analytics models, and closed-loop execution logic.

The integration of GenAI4SOAR further extends the interaction with the exposure framework by introducing AI-assisted, explainable remediation workflows. When security policies are onboarded or when anomalies and predictions are raised, GenAI4SOAR dynamically generates CACAOv2-compliant remediation playbooks that can be exposed to consumers through SOAR interfaces or conversational tools, enabling review, validation, modification, or approval before enforcement. This supports human-in-the-loop oversight, as required by ETSI ZSM principles and regulatory frameworks such as the EU AI Act, while preserving the benefits of zero-touch automation.

Overall, the interaction between the Zero-Touch Security Management Layer and the exposure framework exposes security as a programmable, transparent, and policy-driven service. End consumers are provided with mechanisms to define security intent, discover and onboard capabilities, observe security posture and incidents, and oversee AI-generated remediation actions. At the same time, the Zero-Touch Security Management Layer autonomously handles orchestration, closed-loop automation, resource deployment, and continuous adaptation, ensuring scalable and intelligent security management without exposing unnecessary operational complexity.

3.4 Physical Layer Security as a Service

As 6G networks move toward highly directional transmissions, distributed MIMO, reconfigurable intelligent surfaces, and extreme edge deployments, the physical layer becomes a first-class attack surface rather than a passive communication medium. Classical security mechanisms operating at higher layers are insufficient to protect against threats that exploit spatial properties of radio signals, such as beam leakage, angular spoofing, rogue reflectors, and passive eavesdropping. In this context, physical layer security (PLS) must evolve from static design assumptions into a service-driven, observable, and enforceable capability that can be requested, monitored, and dynamically enforced during runtime. To address this gap, we introduce Physical Layer Security as a Service (PLSaaS) for 6G by defining a new stakeholder, the Security Service Provider (SSP), and its functional interaction with the RAN and end consumers. As illustrated in Figure 2, the SSP acts as an intermediary that translates high-level, consumer-defined spatial-security expectations into concrete, RAN-enforceable mechanisms, while continuously monitoring spatial integrity and orchestrating mitigation actions.

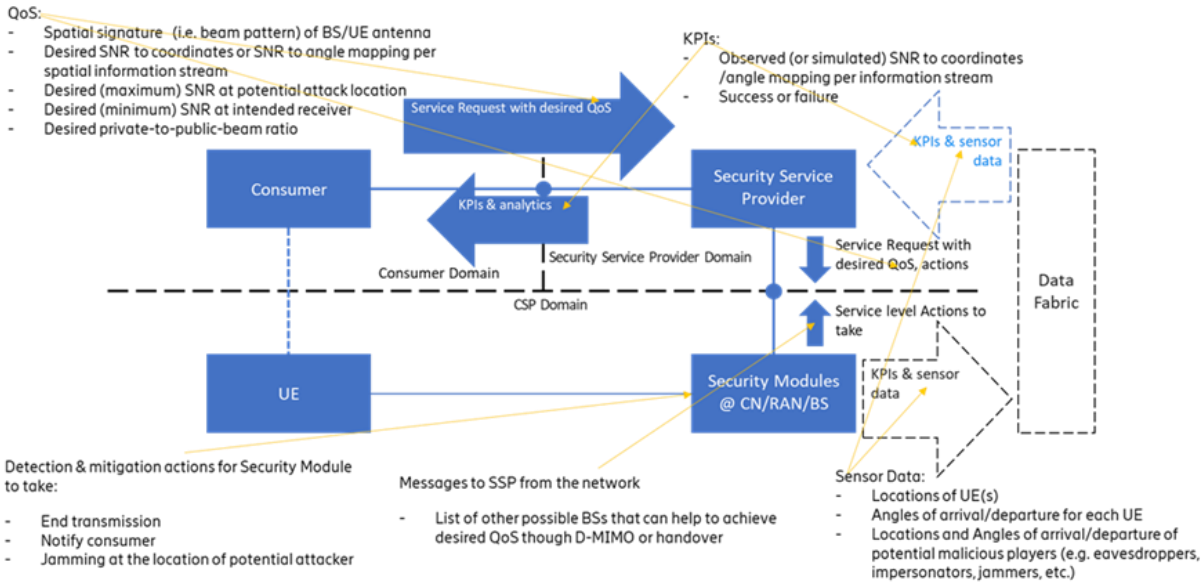


Figure 2: Security Service Provider (SSP) entity

The associated signalling procedures and information exchange required to support this closed-loop operation are shown in Figure 3.

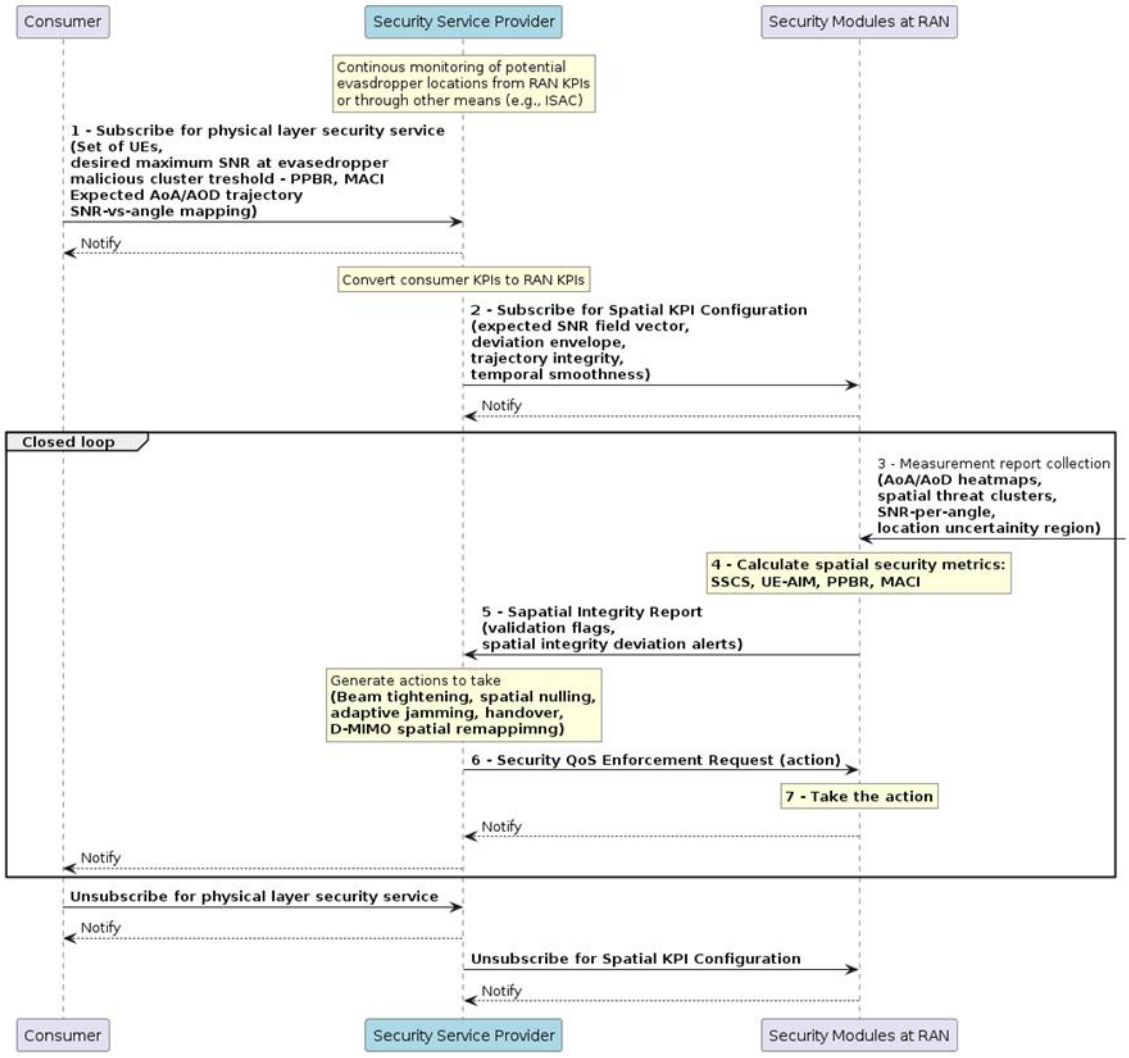


Figure 3: Signalling required for physical layer security as a service

The service lifecycle begins when a consumer or application explicitly requests a communication service with Secure Spatial QoS requirements. Unlike traditional QoS profiles limited to latency or throughput, this request includes constraints on beam stability, angular SNR distribution, allowable spatial leakage, and tolerance to anomalous energy patterns. The SSP interprets these behavioural requirements and initiates a closed-loop spatial-security process by converting them into measurable conditions.

The SSP then derives a set of spatial-integrity KPIs that the RAN can continuously observe. These include the expected SNR field vector across angles, allowable angular deviation envelopes for legitimate UE motion, trajectory integrity metrics, and temporal smoothness of angle-of-arrival evolution. These KPIs, delivered through dedicated configuration signalling, establish a baseline definition of “secure spatial behaviour” that does not exist in current 3GPP systems.

To evaluate compliance with these KPIs, RAN Security Modules collect high-resolution spatial sensor data beyond conventional measurements. This includes AoA/AoD heatmaps, per-angle SNR fields, UE uncertainty regions, and persistent spatial energy clusters. Based on this data, the RAN computes four dedicated spatial-security metrics: the Spatial SNR Consistency Score (SSCS), UE Angle-of-Arrival Integrity Metric (UE-AIM), Private-to-Public Beam Ratio (PPBR), and Malicious AoA Cluster Index (MACI). Together, these metrics capture complementary aspects of spatial integrity, confidentiality, and adversarial presence.

The computed metrics are periodically reported to the SSP, which performs multi-metric spatial-integrity analysis. Rather than relying on single thresholds, the SSP correlates spatial, temporal, and geometric deviations to detect complex threats such as coordinated angular spoofing combined with beam leakage or persistent malicious clusters. When violations of the requested Secure Spatial QoS profile are detected, the SSP generates Security QoS Enforcement Commands, instructing the RAN to apply targeted countermeasures.

These enforcement actions are executed directly at the PHY and MAC layers and include beam tightening, spatial null-forming, adaptive directional jamming, secure handover triggering, and D-MIMO spatial remapping. The actions reshape the electromagnetic environment in real time while maintaining service continuity for legitimate users. A continuous feedback loop then monitors post-mitigation stability, allowing the SSP to relax or escalate actions as conditions evolve.

Overall, this framework establishes the closed-loop, service-oriented spatial-security architecture for 6G. By introducing new entities, KPIs, signalling, and enforcement mechanisms, it transforms spatial behaviour from an implicit by-product of radio design into a measurable, controllable, and consumer-driven security dimension. This enables effective protection against adversarial reflectors, angular spoofing, passive eavesdropping, and impersonation attacks, threats that cannot be detected or mitigated by existing 3GPP mechanisms, thereby providing a fundamentally new security capability for future 6G systems.

4 Relationship to Other Architecture Studies

4.1 AIML Life-Cycle Management

4.1.1 Relation to Standardizations

In standards and open-source communities, there is an effort to define AI/ML frameworks in recent years. In 3GPP, AI/ML Enablement (AIMLE) layer focuses on enabling vertical AI applications through a communication network [3GP26b].

The AIMLE Service provides a unified framework for enabling, coordinating, and managing AI/ML operations across 3GPP networks (cloud, edge, and device layers). It exposes these capabilities to Vertical Application Layer (VAL) applications via standardized APIs, supporting distributed, federated, and split learning with tight integration to network analytics and control functions. These services can be covered under ML model services, client services, FL services, and distributed ML services and mobility support.

ML Model services focus on model storage, discovery, and retrieval as well as training, update, and performance monitoring. The model discovery can be based on filters such as domain, vendor interoperability, accuracy, etc. The model retrieval can be based on filter criteria such as model type, domain, performance metrics, etc. Training services allow dataset assignment, training objective definitions, and training progress notifications. These services can be in the form of request/response and/or subscribe/notify. Notifications include new model availability, changes in model support, performance metrics (accuracy, latency, etc.), performance degradation, or retraining needs.

AIMLE client services allow clients to register with profiles containing supported model types, supported operations (training, inference, model split, offload), availability schedule, location constraints, dataset availability, compute capability & energy preferences. Then service consumers can discover (and subscribe to information on) these clients based on required model types & operations dataset requirements, location, mobility, QoS thresholds, and compute capabilities. AIMLE can utilize NWDAF, NEF, or other SEAL services for location and QoS analysis.

AIMLE also provides some basic FL services to the entities in the 3GPP system such as registering themselves based on FL role (client or server) they can have along with models they support and their temporal and spatial availability. Then the service consumers can subscribe to changes in this information. The services can help the consumer to form a group of clients who can participate in FL. It also has support for feature alignment to enable VFL.

In addition to FL, AIMLE has some support for other kinds of distributed learning. To enable split learning, it has services to manage data pipelines and for per-stage model assignment. It also has services for transferring heavy ML tasks to edge or cloud nodes with information including intermediate model states, iterations, or resource usage. Moving client context also addresses continuity of operations under UE mobility.

The AIMLE service acts as a standalone AI/ML enablement platform within 3GPP networks. It supports ML model discovery and lifecycle operations to federated learning, split inference pipelines, task transfer, and

edge-to-cloud coordination — all exposed through standardized SEAL-based APIs. The service is designed to support large-scale, distributed, and network-aware AI/ML applications across vertical industries such as automotive, robotics, XR, industrial IoT, and more.

ROBUST-6G Trustworthy and Sustainable AI Layer borrows solutions and services already studied in the standards, and it offers extra solutions and services enabling explainable AI, privacy preserving distributed learning, and sustainable AI.

In addition to AIMLE, 3GPP also studies AI/ML life-cycle management (LCM) of AI/ML workloads in the network [3GP26c]. Sustainable AI is in the study phase for the next release [3GP26d]. In open-source community, CAMARA has Model-as-a-Service APIs in development [CAM26b]. Use cases and solutions studied in ROBUST-6G can help these standardization and development activities to extend their scope if needed.

4.1.2 Comparison of AI/ML Life-Cycle Management in ROBUST-6G, VERGE, and HEXA-X II Architectures

This section positions the AI/ML life-cycle management capabilities and architectural principles of the ROBUST-6G, VERGE, and HEXA-X II research projects. The comparison highlights the degree and manner in which each project incorporates AI/ML capabilities into its architectural framework, focusing on lifecycle management, automation, distributed intelligence, trustworthiness, and orchestration support in next generation 6G ecosystems.

4.1.2.1 ROBUST-6G Architecture: Integration of Trustworthy AI into Existing MLOps Frameworks

The ROBUST-6G architecture does not aim to define or redesign AI/ML lifecycle management mechanisms from scratch. Instead, it assumes the availability of AI/ML lifecycle and MLOps capabilities already present in future 6G networks, such as model training pipelines, deployment mechanisms, monitoring, and orchestration frameworks.

Within this assumption, ROBUST-6G focuses on augmenting existing MLOps frameworks with trustworthy AI functionalities, particularly in the context of security and resilience. Key architectural characteristics include:

- Explicit reliance on pre-existing AI/ML lifecycle mechanisms, which are treated as part of the underlying network and service management infrastructure.
- Integration of trustworthy AI enablers such as robustness against adversarial attacks, explainability, accountability, privacy preservation, and secure model execution into the operational phases of the AI lifecycle. [SSW+24]
- Alignment with zero-touch and closed-loop operational models, where AI components participate in monitoring, decision-making, and remediation while leveraging existing lifecycle automation.
- Support for distributed and federated learning paradigms, not by redefining their lifecycle, but by enhancing them with security and trust mechanisms that ensure safe collaboration and model evolution across administrative domains.
- Focus on security-centric AI usage, where lifecycle stages (e.g., deployment updates, retraining triggers, inference validation) are tightly coupled with network security functions and policies.

In summary, ROBUST-6G positions itself as an architectural integrator of trustworthy AI capabilities into existing MLOps and AI lifecycle frameworks, ensuring that AI-driven security and management functions in 6G networks remain reliable, explainable, and resilient.

4.1.2.2 VERGE Architecture: Edge-Native Lifecycle Unification and Closed-Loop Automation

The VERGE architecture addresses AI/ML lifecycle management primarily through the lens of edge-cloud continuum orchestration and closed-loop automation. [VER25-D22]

- Life Cycle Management (LCM) and closed-loop automation are foundational elements of the “Edge for AI” (Edge4AI) pillar, which seeks to unify management of cloud-native applications, Multi-access

Edge Computing (MEC) services, and distributed workflows across heterogeneous processing resources.

- The architecture embeds lifecycle operations within a modular edge continuum, facilitating deployment, orchestration, and dynamic placement of AI tasks and workflows in response to application and network KPIs.
- "AI for Edge" (AI4Edge) introduces AI-driven orchestration and optimization techniques leveraging real-time telemetry to manage resources and adapt lifecycle state transitions automatically.
- Although not explicitly framed as an AI lifecycle platform, VERGE's architecture manifests lifecycle management capabilities through closed-loop mechanisms that span training, deployment, operation, and contextual adaptation of distributed AI assets.

Thus, VERGE architecture emphasizes distributed lifecycle orchestration over an edge-cloud continuum, with strong integration between lifecycle state transitions and closed-loop automation that supports high performance edge AI services.

4.1.2.3 HEXA-X II Architecture: Intelligent Networking with Embedded AI Frameworks

The HEXA-X II architecture adopts a broad system-level perspective**,** where AI/ML is a fundamental enabler of intelligent network behavior rather than a standalone managed asset. [HEX24-D53]

Relevant characteristics include:

- Introduction of AI-as-a-Service (AIaaS) and Federated Learning-as-a-Service (FLaaS) as architectural frameworks that support training, inference, and analytics across multiple domains.
- Lifecycle-related functions such as model instantiation, execution, monitoring, and updates are embedded within these service frameworks and orchestrated as part of the overall network control and management plane.
- Emphasis on large-scale intelligent automation, where AI components support self-configuration, self-optimization, and self-healing capabilities of the 6G system.
- Limited explicit treatment of AI trustworthiness, governance, or lifecycle-wide security controls, which are largely implicit or left to implementation-specific solutions.

Consequently, HEXA-X II provides architectural enablers for AI integration and operation, but does not explicitly frame AI/ML lifecycle management or trustworthiness as a dedicated architectural concern.

4.2 Exposure Gateway

The NetSecaaS framework demonstrates strong alignment with existing standardization efforts, while also highlighting meaningful gaps in their current scope. On the borrowing side, the framework's heavy reliance on the GSMA Open Gateway [GSM26] initiative and the CAMARA [CAM26a] project's API methodology is evident, with the leveraging of their developer-friendly, outcome-oriented design philosophy and OAuth-based authentication patterns — areas where these bodies offer mature, well-tested blueprints that NetSecaaS adopts rather than reinvents. The architecture also borrows concepts from 3GPP [3GP26a] around Network Exposure Function (NEF) design and policy enforcement frameworks, particularly for the southbound interface governance model. Here, 3GPP's established work on capability exposure in 5G could strengthen the Transformation Function's interface mapping logic. However, the ROBUST-6G NetSecaaS architecture also highlights significant gaps in the scope of these standardization bodies. CAMARA and GSMA Open Gateway APIs are largely agnostic to security-as-a-service semantics, they focus on general network capability exposure without natively addressing AI-driven threat analytics, physical-layer security primitives or Security Service Level Agreement (SSLA) constructs. Similarly, the 3GPP NEF-centric exposure model does not account for zero-touch security orchestration triggered by high-level consumer queries nor does it define mechanisms for exposing trustworthy AI model outputs as a core API capability. In this sense, the NetSecaaS framework positions ROBUST-6G not merely as a consumer of standardization, but also as a contributor. It offers novel API definitions, particularly the AI-driven Threat Analytics, Zero-touch Automation and Physical Layer

Security APIs, which could extend the scope of the CAMARA API family. These definitions could also inform future 3GPP and GSMA work items on the exposure of security-specific capabilities in 6G networks.

4.3 RAN Functions

The ROBUST-6G architecture relies on some radio access functions such as telemetry inputs for the data management platform and physical layer security closed-loop layer:

- **RAN telemetry for data management platform:**
  The data related to performance metrics (e.g., radio measurements, resource usage stats) and event logs from base station components is forwarded into the data management platform for organization and cross-layer aggregation.
- **RAN telemetry for physical layer analysis:**
  Physical layer measurements (e.g., signal characteristics, spectrum occupancy, power levels) are gathered from user equipment and base station to feed the physical layer security closed-loop to detect physical-layer anomalies like spoofing or jamming.

Moreover, the zero-touch security management layer of the ROBUST-6G architecture can interface with the near-real-time RAN intelligent controller (Near-RT RIC) of the O-RAN Alliance to deploy security-oriented xApps that can predict threats and automatically trigger real-time mitigation actions, such as configuration changes, or isolation of compromised components, without manual intervention, which therefore reduces operational complexity and maintains stable and secure network operation even in highly dynamic 6G environments.

External consumers can leverage the ROBUST-6G architecture as an automated security and trust management platform integrated with their 6G infrastructures, either through direct deployment or by connecting to the platform via secure, well-defined APIs that expose its internal security capabilities. Rather than relying on manual monitoring and static security policies, the platform continuously collects telemetry from network components (e.g., including both RAN and core network elements), analyses this data using AI/ML models and automatically orchestrates appropriate mitigation actions. For instance, external architectures can utilize the physical-layer security closed-loop mechanisms to detect threats originating from the radio environment, such as jamming or spoofing attempts, and trigger rapid mitigation through automated actuation.

5 ROBUST-6G Enablers for Smart Security Services

This section describes the four principal pillars of ROBUST-6G architecture: i) Data Management, ii) Trustworthy AI Services, iii) Zero Touch Security Orchestration, and iv) Physical Layer Security. In this vein, the following subsection presents how these enablers are decomposed into simpler components, functions, or theoretical models, as well as how they fit the initial requirements defined in D2.2 [R6G24-D22].

5.1 Data Management Platform

The ROBUST-6G platform introduces a trusted data sharing framework designed to enable secure, interoperable, and policy-driven data exchange across distributed infrastructures and multi-domain environments. The architecture, as shown in Figure 4, is structured around two complementary layers: the **Data Fabric**, which provides the technical capabilities required to ingest, transform, integrate, and expose data across heterogeneous sources, and the **Data Governance**, which ensures that data sharing follows clear governance policies and access control mechanisms.

The Data Fabric enables seamless integration of heterogeneous datasets and services by combining semantic technologies, distributed data processing mechanisms, and standardized interfaces. In parallel, the Data Governance guarantees that data is discoverable and accessed only by authorized entities under well-defined policies. Together, these layers provide a robust foundation for trusted data exchange and interoperability within the ROBUST-6G ecosystem.

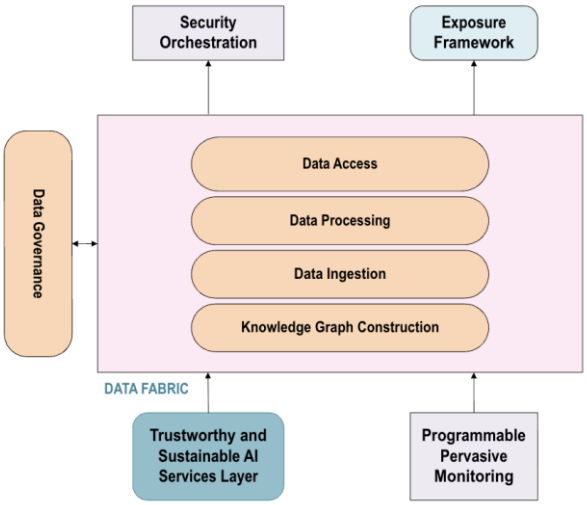


Figure 4: Data Management Platform

5.1.1 Data Fabric

The Data Fabric layer provides the technical infrastructure responsible for the ingestion, transformation, integration, and exposure of data across the ROBUST-6G platform. It enables data from heterogeneous systems to be normalized, semantically integrated, and delivered to consumers in a consistent and interoperable manner.

The main building blocks of the Data Fabric include the **Knowledge Graph**, **Data Ingestion**, **Data Processing**, and **Data Access** components.

5.1.1.1 Knowledge Graph

The Knowledge Graph represents the central component of the ROBUST-6G Data Fabric. It acts as a semantic integration layer that stores both data and metadata describing datasets, services, and relationships between system entities. By relying on semantic models, the Knowledge Graph enables interoperability between heterogeneous data sources while supporting advanced reasoning and querying capabilities.

In addition to serving as a metadata hub for the platform, the Knowledge Graph facilitates the transformation of raw datasets into structured semantic representations aligned with shared ontologies and taxonomies. This process enables the normalization and integration of heterogeneous data into a unified representation suitable for analytics, automation, and AI-driven services.

The Knowledge Graph implementation relies on standards defined by the **Semantic Web technology stack** [AH11], [BHL01], including:

- **RDF (Resource Description Framework)** – a graph-based data model used to represent information as triples (subject–predicate–object), enabling flexible integration of heterogeneous datasets.
- **RDFS (RDF Schema)** – a vocabulary for defining classes, properties, and hierarchical relationships within RDF graphs.
- **OWL (Web Ontology Language)** – a language for defining complex ontologies with rich semantics and logical constraints.
- **SPARQL** – the standard query language for retrieving and manipulating RDF data within knowledge graphs.

Together, these standards provide the foundation for semantic interoperability and machine-interpretable data integration. Figure 5 shows an example stack of layers for such integration.

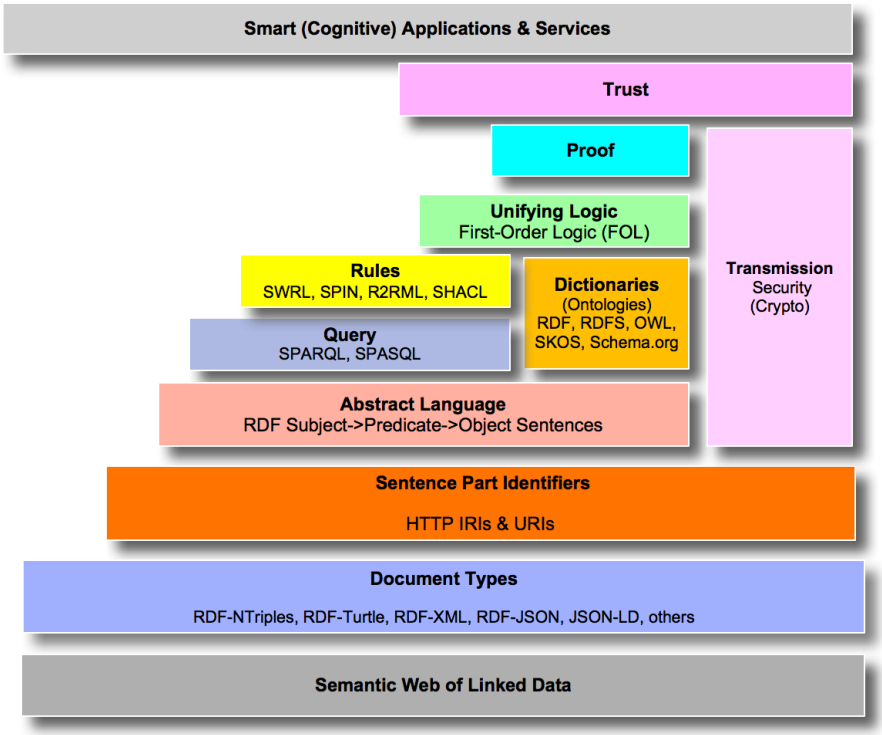


Figure 5: Semantic Web layer. (Source: [Ide26])

To ensure semantic consistency and interoperability across datasets, ontology engineering practices are adopted as a fundamental component of the Knowledge Graph design. Ontologies define the core concepts, entities, relationships, and properties that describe domain knowledge within the ROBUST-6G ecosystem. By providing a shared semantic model, ontologies enable heterogeneous data sources to be integrated under a common representation, facilitating interoperability between systems that may otherwise use different schemas, formats, or terminologies. In this context, ontologies play a critical role in enabling semantic alignment between datasets originating from different domains, organizations, or network components.

The development of ontologies follows established methodologies that support systematic design, reuse, and long-term maintenance. As shown in Figure 6, One such methodology is **Linked Open Terms (LOT) [PFF+22]**, an industrial approach for building ontologies and vocabularies in a structured and iterative manner. The LOT methodology has been successfully applied in the development of the **ETSI SAREF** [GLP+23] ontology and its domain-specific extensions. LOT adopts best practices from agile software development, such as iterative sprints, continuous integration, and collaborative development between domain experts and ontology engineers.

The methodology is structured around a lifecycle composed of several iterative phases. The first phase, **Ontology Requirements Specification**, focuses on systematically collecting and documenting the requirements that the ontology must satisfy. Techniques such as competency questions, natural language descriptions, and tabular requirement specifications are used to capture the expected functionality and intended use cases of the ontology. This stage typically involves close collaboration between domain experts and technical stakeholders to ensure that the resulting semantic model accurately reflects real-world concepts and relationships.

The second phase, **Ontology Implementation**, translates the specified requirements into a formal semantic model using languages such as **Web Ontology Language (OWL)** or **RDF Schema (RDFS)**. During this stage, ontology engineers define classes, properties, and relationships that represent domain knowledge. A key principle in this process is the reuse of existing ontologies and vocabularies whenever possible, which promotes interoperability and reduces duplication of modelling efforts across projects.

The third phase, **Ontology Publication**, ensures that the ontology is made available in both machine-readable and human-readable formats through persistent and dereferenceable URIs. Publishing ontologies according to

Linked Data principles allows them to be reused, extended, and integrated by other systems, thereby supporting interoperability across distributed data ecosystems.

Finally, the **Ontology Maintenance** phase addresses the continuous evolution of the ontology over time. As domain requirements change or new datasets are integrated into the platform, the ontology may require updates to incorporate new concepts, relationships, or constraints. Maintenance activities typically include bug fixing, refinement of definitions, adaptation to evolving domain requirements, and the incorporation of feedback from users and developers.

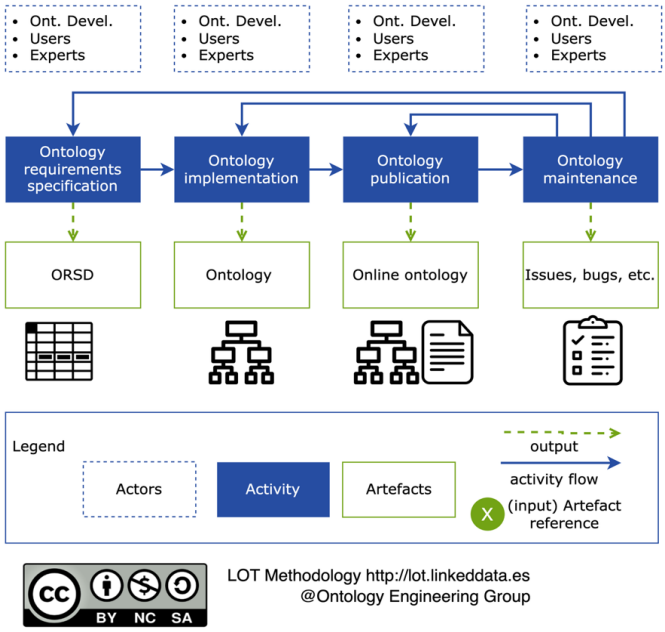


Figure 6: LOT methodology. (Source: [PFF+22])

To support the implementation and storage of the Knowledge Graph, several triple store technologies can be used. One prominent option is **GraphDB** [Ont26], a scalable RDF database designed for high-performance semantic applications. GraphDB [Ont26] provides full support for RDF 1.1 and **SPARQL**, along with advanced reasoning capabilities based on RDFS and OWL semantics. It also offers features such as semantic inference, full-text search integration and graph exploration tools, making it suitable for large-scale knowledge graph deployments.

In addition to the storage layer, client libraries are required to enable programmatic interaction with the Knowledge Graph. Libraries such as **RDFLib** [RDF26] for Python and **Apache Jena** [Apa26a] for Java provide comprehensive APIs for creating, manipulating, and querying RDF graphs. These libraries support multiple serialization formats, including Turtle, RDF/XML, and JSON-LD, and enable applications to execute SPARQL queries, manage RDF datasets, and integrate semantic data processing capabilities within the ROBUST-6G platform.

5.1.1.2 Data Ingestion

The **Data Ingestion** building block is responsible for collecting datasets from distributed data services operating within the ROBUST-6G ecosystem. Its main objective is to enable the seamless integration of heterogeneous data sources by providing connectors and interfaces that support a wide range of communication protocols, data formats, and service interfaces.

Through these connectors, the platform can retrieve datasets from multiple systems, including monitoring platforms, network components, analytics services, or external data providers. Each connector implements the specific access mechanisms required by the corresponding data source, ensuring compatibility with existing technologies and minimizing integration overhead.

Once collected, the ingested datasets are forwarded to the subsequent stages of the data pipeline for further processing, integration, and normalization. Depending on the use case, the ingestion layer may support both

batch-based ingestion workflows and **streaming data flows**, allowing the platform to handle static datasets as well as real-time data streams.

In addition to retrieving the raw datasets, the ingestion process may also capture relevant metadata describing the origin, structure, and characteristics of the data. This metadata can be registered within the platform's Knowledge Graph to support later stages such as data discovery, governance, and processing orchestration.

To support the implementation of ingestion pipelines, integration frameworks such as **Apache Camel** [Apa26b] provide a flexible and extensible solution for connecting heterogeneous systems. Apache Camel offers a large collection of pre-built connectors, known as *components*, which enable integration with a wide variety of protocols and platforms. These components support common technologies such as HTTP, FTP, messaging systems, relational databases, cloud services, and event streaming platforms.

Within integration pipelines, these components act as standardized endpoints that allow data to be retrieved, routed, and delivered across different stages of the data processing workflow.

5.1.1.3 Data Processing

The **Data Processing** building block provides the mechanisms required to transform ingested datasets into a structured and interoperable representation suitable for integration, analysis, and reuse. Typical processing operations include **data cleansing**, **format normalization**, and **data integration** across multiple sources.

A central capability of this component is **semantic lifting**, a transformation process through which raw datasets are converted into structured RDF graphs aligned with the ontologies used within the ROBUST-6G Knowledge Graph. During this process, as shown in Figure 7, the original data elements are mapped to semantic entities and properties defined in the ontology, enriching the dataset with explicit semantics.

Semantic lifting achieves two main objectives. First, it enables **data normalization**, as heterogeneous datasets are transformed into a shared semantic representation. Second, the graph-based structure facilitates **data integration**, allowing datasets from different sources to be connected and queried through common semantic relationships.

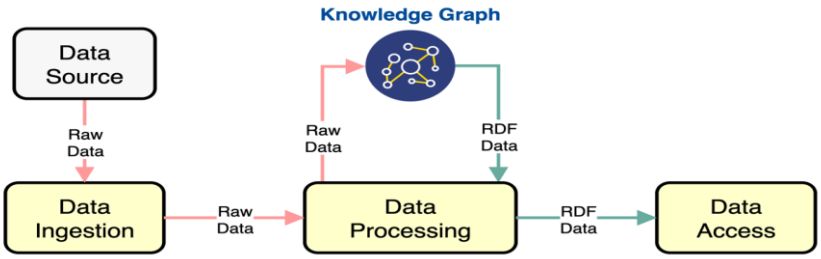


Figure 7: Semantic lifting

To implement semantic lifting in a scalable and reusable way, the platform adopts **declarative mapping approaches** [VDH+23]. Instead of embedding transformation logic directly in code, dataset providers define mapping rules that describe how elements of the source dataset correspond to concepts in the target ontology. A generic mapping engine then executes these rules to automatically generate the corresponding RDF graph.

This approach simplifies the integration of new data sources, as data owners only need to define the transformation rules, while the processing infrastructure handles the execution of the mappings.

Among the available declarative mapping languages, the platform adopts **RDF Mapping Language (RML)** [Igl+23]. RML is an extension of the W3C **R2RML** standard, which was originally designed for mapping relational databases to RDF.

RML extends the capabilities of R2RML by supporting a wider range of structured data formats, including CSV, JSON, XML, and TSV. This flexibility makes RML particularly suitable for environments where datasets originate from heterogeneous systems and follow different data representations.

Using **RML**, dataset providers can define declarative mappings that specify how data elements from the input source correspond to RDF triples in the target Knowledge Graph. These mappings are executed by the **Chimera framework** [GSC+23], an extension built on top of Apache Camel designed specifically to handle RDF data and enable semantic data transformation pipelines.

Chimera manages RDF data throughout the transformation process, supporting in-memory operations, temporary storage on the filesystem, or direct insertion into the platform's Knowledge Graph via **SPARQL endpoints** or dedicated APIs. Its integrated **RML processor** applies the mapping rules to generate semantically enriched datasets that conform to the shared ontological model.

Additionally, Chimera can leverage the **Mapping Template Language (MTL)** to perform complex transformations, enabling RDF datasets to be adapted to alternative formats or consumer-specific schemas.

5.1.1.4 Data Access

The **Data Access** building block provides the mechanisms that allow data consumers to retrieve the datasets generated by the data pipelines within the ROBUST-6G Data Fabric. This component is responsible for exposing the processed and integrated datasets to authorized consumers through standardized interfaces and communication protocols.

A key capability of this component is **semantic lowering**, which can be optionally enabled when data consumers are not able to directly handle semantic data formats such as RDF. In such cases, the platform transforms the semantically normalized dataset stored in the Knowledge Graph into the specific format and schema required by the consumer. As illustrated in Figure 8, this process converts RDF graph data into more commonly used data representations such as JSON, CSV, or other structured formats, allowing seamless integration with external systems and applications.

In addition to format transformation capabilities, the Data Access layer provides multiple delivery mechanisms that enable datasets to be exposed or streamed to consumers. These mechanisms support several communication protocols commonly used in distributed systems, including REST APIs, messaging platforms, and streaming technologies such as Kafka or MQTT. This flexibility allows the platform to support both request–response interactions and event-driven data distribution models.

Advanced consumers may also directly interact with the Knowledge Graph through its **SPARQL endpoint**, enabling complex semantic queries and advanced data analytics operations over the integrated datasets.

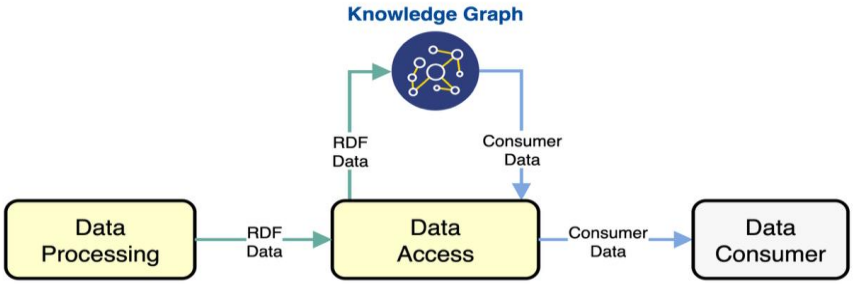


Figure 8: Semantic lowering

To implement the data delivery mechanisms of the **Data Access** layer, integration frameworks such as **Apache Camel** provide a flexible and modular solution for exposing datasets through multiple communication protocols. Camel offers a wide range of connectors supporting REST services, messaging systems, streaming platforms, and other integration interfaces, enabling seamless delivery of datasets produced by the data pipelines to diverse consumers.

In addition, the **Chimera framework** [SCG+24], which extends Apache Camel, adds advanced semantic transformation capabilities for the **denormalization** of RDF datasets into consumer-specific outputs. During this process, data stored in the Knowledge Graph is queried using **SPARQL**, and Chimera applies declarative mappings (RML) or template-based transformations (MTL) to generate the requested data format and schema, such as JSON, CSV, or other structured outputs. This approach ensures that the Data Access layer can provide

both semantically enriched RDF datasets and simplified representations tailored to the requirements of individual consumers, maintaining interoperability across the ROBUST-6G Data Fabric.

5.1.2 Data Governance

The **Data Governance** building block is responsible for defining and enforcing the policies and mechanisms that regulate how data is described, discovered, accessed, and trusted within the ROBUST-6G Data Fabric. Its primary objective is to ensure that datasets shared across the platform remain **discoverable, trustworthy, and securely accessible**, while complying with governance principles such as transparency, accountability, and controlled data sharing.

Data governance mechanisms ensure that datasets integrated within the Data Fabric follow the **FAIR principles** [Wil+16], meaning that data is **Findable, Accessible, Interoperable, and Reusable**. These principles are implemented through a combination of metadata management, dataset cataloguing, and secure access control mechanisms.

Within the ROBUST-6G architecture, the governance layer relies heavily on the **Knowledge Graph**, which stores metadata describing datasets, services and data relationships. As shown in Figure 9, This semantic metadata layer enables the platform to support dataset discovery, enforce governance policies, and facilitate interoperability across heterogeneous data sources.

The following sections describe the main governance capabilities provided by the platform.

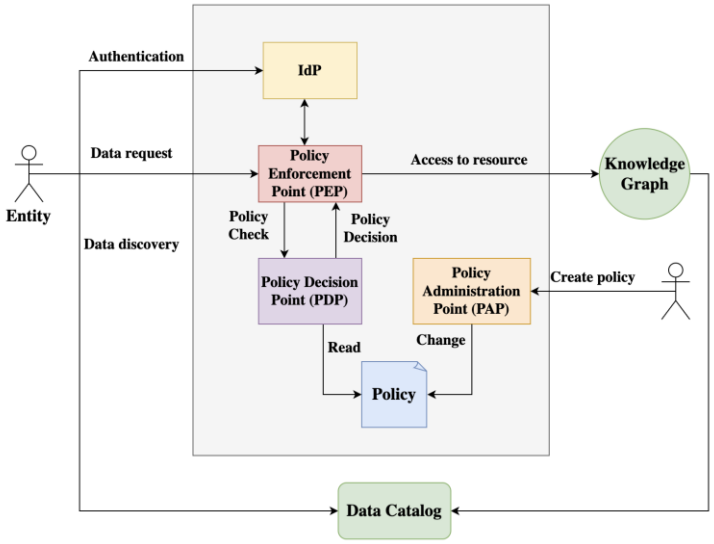


Figure 9: Data Governance

5.1.2.1 Data Catalog

The **Discovery** capability enables data consumers to identify and understand the datasets available within the ROBUST-6G Data Fabric. This functionality is implemented through a **data catalog** that maintains structured metadata describing datasets, data services, and their associated characteristics.

This catalog is integrated with the platform's **Knowledge Graph**, which acts as the semantic backbone of the Data Fabric. By storing dataset metadata in a graph-based representation, the system enables advanced search capabilities, semantic queries, and automated discovery of relationships between datasets.

Through this discovery mechanism, data consumers can identify datasets relevant to their use cases and specify how they wish to consume them, either by accessing the raw data directly or by triggering processing pipelines that transform the data into a suitable format.

Dataset metadata within the catalog is represented using **DCAT-AP [Com26]**, a specification based on the W3C **Data Catalog Vocabulary** designed to harmonize dataset descriptions across European data portals.

DCAT defines a common vocabulary for describing datasets, distributions, data services, and catalogs, enabling interoperable metadata exchange between different data platforms. DCAT-AP extends this model with additional constraints, controlled vocabularies, and properties tailored to European data ecosystem requirements.

5.1.2.2 Identity management and authorization

Secure data sharing requires robust mechanisms for verifying identities, enforcing authorization policies, and protecting access to resources. For this reason, the ROBUST-6G platform integrates an **authentication and authorization framework** that governs how users and services interact with protected datasets.

This framework establishes a **policy-driven access control architecture** that ensures every request to access data is authenticated, evaluated against governance policies, and authorized before the requested resource is delivered.

The architecture follows a modular design composed of several cooperating components:

- **Identity Provider (IdP)** responsible for authentication and identity federation
- **Policy Decision Point (PDP)** responsible for evaluating authorization policies
- **Policy Enforcement Point (PEP)** responsible for enforcing authorization decisions
- **Policy Administration Point (PAP)** responsible for defining and managing authorization policies

These components collectively ensure that access to datasets within the ROBUST-6G Data Fabric is governed by consistent and auditable policies.

5.1.2.3 Identity Management

Identity management within the platform is implemented using a centralized identity repository based on **OpenLDAP** [Ope26]. The LDAP directory acts as the authoritative source of identity information, storing user accounts, roles, and group memberships in accordance with standard LDAP schemas.

The directory structure organizes identities within hierarchical namespaces that include organizational units for users, groups, and roles. This structure enables centralized administration of identity attributes and simplifies the management of user permissions across multiple services. By relying on LDAP as the primary identity store, the architecture decouples identity management from application logic, allowing identities to be reused consistently across the entire platform. Figure 10 shows the interface for this directory.

User roles and access privileges are derived from group memberships defined within the directory. These groups form the foundation for role-based access control mechanisms, enabling administrators to assign permissions at the group level rather than individually for each user. This approach improves scalability and simplifies policy management in large deployments.

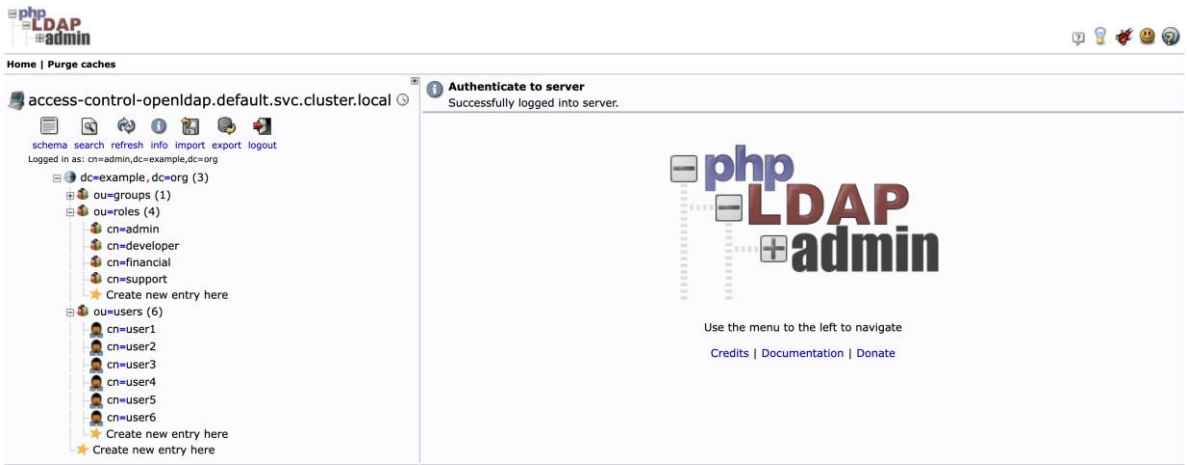


Figure 10: OpenLDAP Directory

5.1.2.4 *Authentication and Identity Federation*

Authentication and identity federation are handled by **Keycloak** [Key26], which serves as the platform's Identity Provider (IdP). Keycloak integrates with the LDAP directory through a federation mechanism that allows user identities to be stored externally while still being used for authentication and authorization within the platform.

Keycloak has been **customized to interact with the Policy Decision Point (PDP)**. Before completing authentication flows, Keycloak can consult the PDP to evaluate policies that govern the authentication process itself, such as which **authentication protocols are allowed** to a given user or service. This allows Keycloak not only to authenticate identities but also to enforce governance policies dynamically during the authentication step, ensuring that security rules are applied consistently across all access methods.

During authentication, Keycloak validates user credentials against the LDAP directory and issues signed **JSON Web Tokens (JWTs)** containing identity attributes and role information. These tokens are subsequently used by downstream services and the API gateway to enforce authorization decisions based on centrally defined policies. Keycloak supports OAuth2, OpenID Connect, and SAML protocols, enabling secure integration with APIs, web services, and distributed microservices. Figure 11 shows the interface of Keycloak.

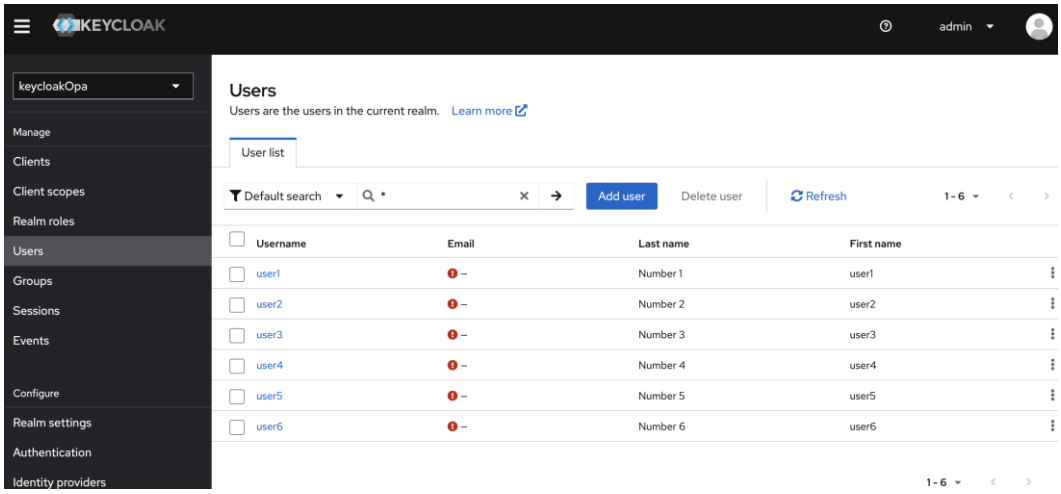


Figure 11: Keycloak integrated with OpenLDAP

5.1.2.5 *Policy-Based Authorization*

Authorization decisions within ROBUST-6G are performed using **Open Policy Agent (OPA)** [OPA26a], which serves as the **Policy Decision Point (PDP)** of the architecture. OPA evaluates incoming access requests using declarative **Rego policies** [OPA26b], providing a flexible and auditable mechanism for enforcing fine-grained authorization.

Decisions can consider multiple factors, including:

- User identity and role information extracted from **JWT tokens**
- The requested resource and action
- Contextual request information, such as **HTTP method** or service endpoint
- Attributes included in the request payload

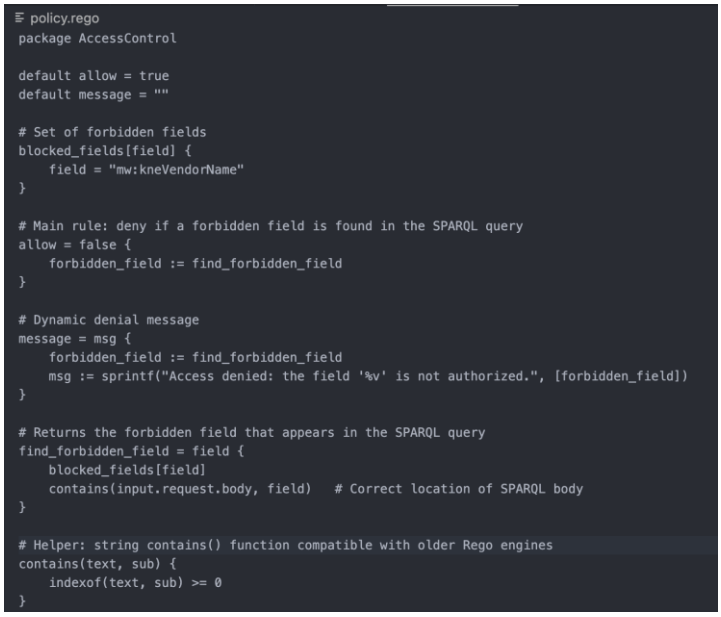
This dynamic evaluation enables advanced authorization models beyond standard role-based access control, supporting **attribute-based** and **context-aware policies** that adapt to runtime conditions.

For example, the following Rego policy (illustrated in Figure 12) enforces **field-level restrictions on SPARQL queries** targeting the Knowledge Graph:

- A **set of forbidden fields** is defined (blocked_fields), such as "mw:kneVendorName".
- The policy dynamically **denies access** if a SPARQL request contains any forbidden field.

- A **custom denial message** is generated, specifying which field caused the rejection.
- The SPARQL query body is analyzed directly from "input.request.body", ensuring enforcement at the query level.
- A helper contains() function ensures compatibility across Rego engine versions for string matching.

This example demonstrates how ROBUST-6G leverages OPA to implement **fine-grained, request-aware access control**, enforcing governance rules at the semantic level while keeping policies declarative, auditable, and adaptable to multi-domain data-sharing scenarios.



```
policy.rego
package AccessControl

default allow = true
default message = ""

# Set of forbidden fields
blocked_fields[field] {
    field = "mw:kneVendorName"
}

# Main rule: deny if a forbidden field is found in the SPARQL query
allow = false {
    forbidden_field := find_forbidden_field
}

# Dynamic denial message
message = msg {
    forbidden_field := find_forbidden_field
    msg := sprintf("Access denied: the field '%v' is not authorized.", [forbidden_field])
}

# Returns the forbidden field that appears in the SPARQL query
find_forbidden_field = field {
    blocked_fields[field]
    contains(input.request.body, field)   # Correct location of SPARQL body
}

# Helper: string contains() function compatible with older Rego engines
contains(text, sub) {
    indexof(text, sub) >= 0
}
```

Figure 12: Rego policy in OPA

5.1.2.6 Policy Enforcement

Authorization decisions are enforced by an API gateway serving as the **Policy Enforcement Point (PEP)**. ROBUST-6G uses **Apache APISIX** [Apa26c] to intercept all incoming requests to protected services.

APISIX responsibilities include:

- Forwarding authorization requests to OPA.
- Enforcing allow or deny decisions from OPA.
- Routing authorized requests to backend services.

Many data access requests involve **SPARQL queries to the Knowledge Graph**. In these cases, the request body contains the query, which determines the entities, concepts, and relationships being accessed. Standard API gateway forwarding of metadata is insufficient for policy evaluation.

ROBUST-6G introduces an **enhanced APISIX–OPA plugin** that forwards the **entire HTTP request body** to OPA. This allows OPA to evaluate **concept-aware policies**, enforcing restrictions based on the specific ontology concepts referenced in the SPARQL query. For instance:

- Policies can restrict access to sensitive entities or classes.
- Multi-domain data-sharing rules can be applied based on the query contents.
- Authorization can consider both the requested resource and the semantics of the query.

This enhancement ensures fine-grained, semantic-aware control over Knowledge Graph access while preserving a centralized, auditable enforcement model.

5.1.2.7 *Policy Administration*

Authorization policies in the ROBUST-6G architecture are centrally managed through the **Policy Administration Point (PAP)**. This layer provides unified mechanisms for creating, updating, and maintaining **Rego-based policies** that govern access to datasets, ensuring consistent and auditable enforcement across all platform services. Figure 13 shows APIs for the policy management.

To manage these policies, a dedicated service has been introduced. Implemented as a **FastAPI-based application**, this service acts as a client of the **Open Policy Agent (OPA)** and allows policies to be directly uploaded, retrieved, updated, or deleted.

This approach enables **centralized, flexible, and fine-grained control**, allowing Rego policies to be applied consistently across the platform without impacting the underlying services.

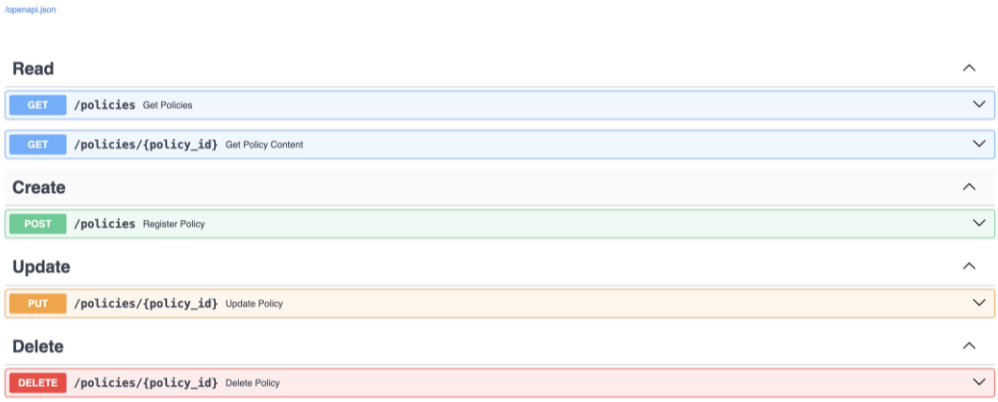


Figure 13: Policy Administration Point

The Data Management Platform integrates a **Data Fabric** and **Data Governance** framework to enable secure, interoperable, and data sharing across distributed environments. The Data Fabric orchestrates the full lifecycle of data, from ingestion and processing to access and delivery, leveraging semantic normalization, knowledge graph integration, and pipeline observability to ensure consistent, traceable, and reusable datasets. Complementing this, the Data Governance layer enforces policies for discovery, security, and compliance, combining federated identity management, policy-driven authorization, and enforcement to guarantee that only authorized consumers can access datasets in accordance with the policies. Together, these layers create a robust ecosystem where data can flow seamlessly across domains while preserving sovereignty, trust, and transparency.

Last but not least, Table 5-1 explains how the requirements associated with the Data Management domain have been addressed. Note that the requirements ID is the same as that defined in D2.2 [R6G24-D22].

Table 5-1: Data management requirements in the ROBUST-6G system

Req. ID	Requirement description	Fulfilled	Brief justification
Application Domain: DATA MANAGEMENT			
R2.1	The ROBUST-6G system shall define user-friendly APIs for external consumers to gain access to exposed capabilities.	Yes	All components that provide relevant data or functionalities of interest to third-party users via accessible APIs are securely exposed through the Exposure framework.
R2.2	The ROBUST-6G system shall provide secure API access for external consumers to the internals of transformation mapping between the service API and the network API.	Yes	All the external reachable API interfaces are protected by APISIX and enforced with OPA-based policies, combined with JWT authentication.

R2.3	The ROBUST-6G system shall have mechanisms to make these APIs discoverable to external consumers.	Yes	All the APIs planned to be accessible by a third-party user and defined within the project are discoverable using a single discovery endpoint that shows only the accessible capabilities to the user who make the query
R2.4	The ROBUST-6G system should provide information to the outside through APIs designed for secure information transaction.	Yes	All project APIs intended for third-party access are discoverable via a single endpoint, which displays only the capabilities accessible to the querying user.
R2.5	The ROBUST-6G system shall allow access to data only to authorised data consumers based on the permissions defined by data owners.	Yes	Access control is enforced through the Policy Decision Point (OPA) and Policy Enforcement Point (APISIX) as described in D2.3, Section 5.1.2. Only authorized consumers are permitted to retrieve datasets.
R2.6	The ROBUST-6G system shall include authentication mechanisms for accessing data.	Yes	Authentication is centrally managed via Keycloak integrated with the LDAP identity repository, as outlined in D2.3, Section 5.1.2.
R2.7	Defining data governance policies for data access shall account for sensitive data.	Yes	The Policy Administration Point (PAP) handles the creation, update, and management of access control policies, while the Policy Decision Point (OPA) enforces authorization by evaluating requests against these policies, ensuring that only authorized users can access sensitive data, as detailed in D2.3.
R2.8	Tracing the data's history throughout its life cycle should be necessary for instilling trust in the data.	Partially	Approaches based on data signing mechanisms modelled in YANG and protected using COSE signatures have been proposed to guarantee traceability and integrity of data throughout its lifecycle. However, this capability was not required by the project's use cases and, therefore, was not integrated into the lifecycle.
R2.9	The ROBUST-6G system must support mechanisms for collecting data in batch mode.	Yes	Section 5.1.1 of D2.3 details that the Data Fabric implements batch ingestion pipelines, leveraging Chimera for semantic transformation, converting raw datasets into RDF.
R2.10	The ROBUST-6G system must support mechanisms for collecting data from streaming data sources.	Yes	According to D2.3, 5.1.1, real-time ingestion pipelines in the Data Fabric allow continuous collection and processing of streaming data.
R2.11	Data consumers within ROBUST-6G should require a means to discover available data.	Yes	As per D2.3, Section 5.1.1, a DCAT-AP-based metadata layer catalogs all datasets and services in the Knowledge Graph.
R2.12	Data management should be distributed to ensure scalability and adaptability in dynamic data exchange scenarios.	Yes	Section 5.1.1 in D2.3 explains that the Data Fabric's distributed architecture enables parallel processing and dynamic integration. Chimera can execute semantic transformations across multiple nodes,

			supporting scalable and resilient data pipelines.
R2.13	Data owners must be accountable for the data products created and exposed with the Data Fabric.	Yes	The Policy Administration Point (PAP) manages these policies, allowing administrators to upload, update, and maintain rules that govern access to data
R2.14	The ROBUST-6G system must enable the integration of heterogeneous data from data sources of different types.	Yes	As described in D2.3, Section 5.1.1, the Data Fabric supports the integration of structured data (e.g., relational databases, CSV files, JSON, XML, event streams).
R2.15	The ROBUST-6G system shall avoid duplicity of information, making efficient use of resources.	Yes	As described in D2.3, Section 5.1.1, the Knowledge Graph acts as the central semantic layer of the Data Fabric, where datasets and their metadata are uniquely identified and referenced.
R2.16	The ROBUST-6G system shall support the monitoring of network and security resources.	Yes	ROBUST-6G offered tools such as T-shark and Falco to monitor and supervise network data and security resources by using the Data Collection Module via Programmable Monitoring Platform.
R2.17	The ROBUST-6G systems should collect data across different layers of the 6G system: service, network, and infrastructure.	Yes	ROBUST-6G collected data at the PHY and application layer and combined them to identify misinformation attacks in VANETS. Besides, it also worked on gathering data via the Data Collection Module of the Programmable Monitoring Platform.
R2.18	The ROBUST-6G system should integrate tailored monitoring agents for far-edge and edge monitoring for selective distribution of monitoring data.	Yes	ROBUST-6G provided monitoring microservices to gather information though the Data Collection Module of the Programmable Monitoring Platform, using containerisation to adapt to different environments.
R2.19	The ROBUST-6G system should have a correlation mechanism for similar data collected from different environments.	Yes	ROBUST-6G used Cortex as correlation tool for cybersecurity incidents.
R2.20	The ROBUST-6G system should support new monitoring modules/tools to extend the capabilities without modifying the core of the platform.	Yes	ROBUST-6G developed Programmable Monitoring Platform with a modular structure to allow for future modifications without affecting the other modules already implemented.
R2.21	The ROBUST-6G system should be able to create new metrics from those previously monitored.	Yes	Data Aggregation and Normalisation Module of Programmable Monitoring Platform uses information from logs to create new numerical metrics.
R2.22	The ROBUST-6G system should aggregate information to preprocess early threat detection.	Yes	The Programmable Monitoring Platform allows numerical and log data to be aggregated using the Data Aggregation and Normalisation Module. The information can be grouped using the Machine_ID, a unique identifier that allows each entity in the environment to be identified.

R2.23	The ROBUST-6G system should be able to reconfigure the monitoring agents dynamically.	Yes	The Programmable Monitoring Platform includes a Configuration Manager for (re)configuring the tools used by each module. It also offers a REST API for receiving new configurations from the ZTSO in two formats. The first format is for Sigma rules and the second is for JSON, using environment variables.
R2.24	The ROBUST-6G system should have secure communication between its modules, with special emphasis on the transmission of the agents with the module in charge of aggregating and preprocessing the information.	Yes	The PMP ensures the security of the information exchanged between the Data Collection Module and the modules of the platform, such as the Data Aggregation and Normalisation Module, the storage modules (Long-term, Medium-term and Real-time) or the Flow Module, amongst others.

5.2 Trustworthy AI Services

The Trustworthy AI Services constitute the operational core of the ROBUST-6G AI-native security framework, translating WP3 innovations into modular, deployable capabilities within the Trustworthy AI Service Layer (see Figure 14). This layer integrates advanced AI functionalities: Adversarial AI, Enhanced Federated Learning (FL), Explainable AI (XAI), and Sustainable AI into a unified service environment that supports secure, distributed, and trustworthy intelligence across 6G systems.

These services are tightly coupled with the AI Service Management Layer, which orchestrates the full AI lifecycle (data ingestion, training, validation, deployment, and storage), and are exposed to external consumers through the Trustworthy AI Service Exposure interface, enabling seamless interaction with ZTSM, physical-layer security and other architectural components.

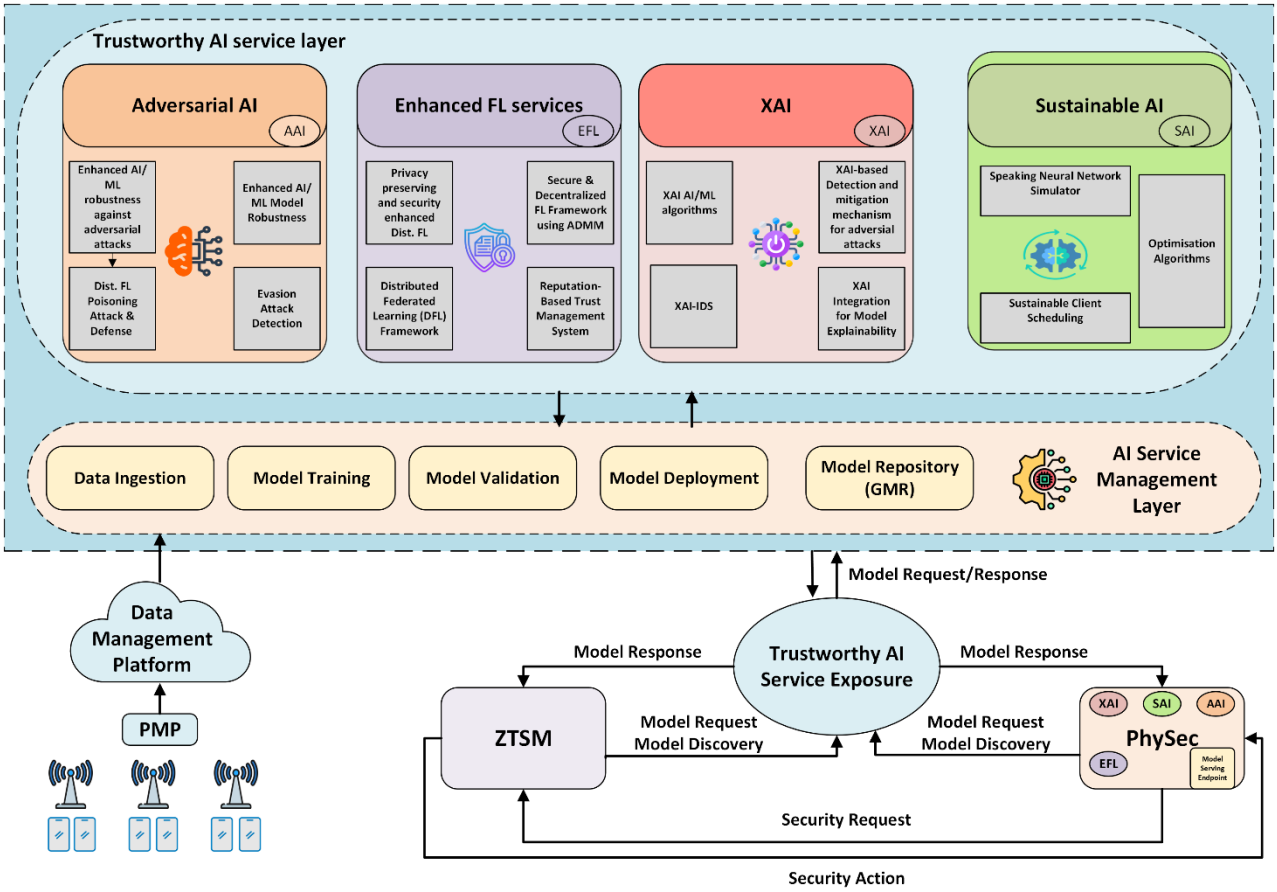


Figure 14 Trustworthy AI Service Layer

5.2.1 Architectural positioning and service exposure

As illustrated in Figure 14, the Trustworthy AI Services operate as an intermediate intelligence layer between:

- The Data Management Platform, which provides data ingestion and preprocessing capabilities,
- The AI Service Management Layer, responsible for lifecycle orchestration and model governance,
- The ZTSM and PhySec components, which consume AI-driven insights for security automation and enforcement.

All AI capabilities are exposed via the Trustworthy AI Service Exposure interface, which supports:

- Model discovery and selection
- Model request/response interactions
- Integration of AI outputs into closed-loop security workflows

This abstraction ensures that AI services are decoupled, reusable, and accessible across domains, aligning with the Exposure Framework principles of the ROBUST-6G architecture.

5.2.2 Core service modules

The Trustworthy AI Service Layer is structured around four complementary service domains: Adversarial AI, Enhanced Federated Learning, Explainable AI, and Sustainable AI, which collectively ensure that AI-driven security mechanisms are robust, privacy-preserving, interpretable, and efficient.

Adversarial AI Services focus on ensuring the resilience of AI models against malicious manipulation. These services embed robustness mechanisms directly into both the training and inference phases, enabling the detection and mitigation of adversarial behaviours such as poisoning and evasion attacks. By incorporating robustness-aware learning strategies and anomaly detection techniques, the system ensures that compromised or unreliable inputs do not degrade model performance. This approach reflects the WP3 principle of enforcing robustness during training rather than as a post-deployment safeguard.

Enhanced Federated Learning Services provide the foundation for distributed intelligence across the 6G continuum. Building on the decentralized federated learning framework developed in WP3, these services enable collaborative model training without sharing raw data, thus preserving privacy while supporting scalability. Security and trust are embedded through mechanisms such as secure aggregation and reputation-based trust management, while decentralized optimization techniques ensure efficient coordination across heterogeneous nodes. This design eliminates single points of failure and aligns with the distributed and multi-domain nature of 6G environments.

Explainable AI Services ensure transparency and interpretability of AI-driven decisions, which is essential for trust, auditability, and regulatory compliance. Explainability mechanisms are integrated into the AI lifecycle, producing interpretable outputs such as feature relevance, confidence levels, and uncertainty estimates alongside model predictions. These insights are leveraged by higher-layer components, particularly ZTSM, to support informed and trustworthy decision-making processes. As highlighted in D3.4, explainability artifacts are generated and stored together with model outputs, enabling traceability across the system.

Sustainable AI Services address the need for energy-efficient and resource-aware AI operation in 6G systems. These services introduce optimization mechanisms that reduce computational and communication overhead while maintaining model performance. By incorporating adaptive client participation strategies and energy-aware training processes, the system can dynamically balance accuracy and resource consumption. Sustainability is therefore treated as an inherent property of the AI workflow, continuously optimized during operation rather than evaluated post hoc.

5.2.3 AI Lifecycle Management and Governance

The Trustworthy AI Services are tightly integrated with the AI Service Management Layer, which orchestrates the complete lifecycle of AI models across the ROBUST-6G architecture. This lifecycle spans from data ingestion to model deployment and continuous monitoring, ensuring that AI services are consistently governed, traceable, and aligned with trustworthiness requirements.

The process begins with data ingestion from the Data Management Platform, where data collected across distributed 6G environments is prepared for training. Model training and validation are then executed using the distributed and privacy-preserving mechanisms provided by the Enhanced Federated Learning and Trustworthy AI services. These stages inherently incorporate robustness, privacy, and explainability, ensuring that trustworthiness is enforced throughout the learning process rather than evaluated after deployment.

Following validation, models are deployed across the edge-cloud continuum, where they can be dynamically accessed and utilized by different system components. The lifecycle is continuously monitored, allowing models to be updated, retrained, or replaced based on evolving conditions, threats, or performance requirements.

A central element in this process is the Global Model Repository (GMR), which acts as a cross-layer governance and storage component. The GMR maintains all model artifacts together with their associated metadata, including performance indicators, robustness characteristics, explainability outputs, and energy-related metrics. This enables full traceability and reproducibility of models across their lifecycle, while also supporting interoperability between different architectural layers. As highlighted in D3.4, the GMR provides a unified interface for storing, versioning, and retrieving models, ensuring that all AI assets remain accessible, auditable, and consistently managed across the system

5.2.4 Integration with ZTSM and Physical-Layer Security

The Trustworthy AI Services are designed to operate as an integral part of the ROBUST-6G closed-loop security framework, enabling seamless interaction with both the Zero-Touch Security Management (ZTSM) layer and physical-layer security (PhySec) mechanisms.

Through the Trustworthy AI Service Exposure interface, ZTSM and PhySec components can dynamically discover and request AI models, enabling flexible and on-demand access to AI-driven intelligence. These components submit model requests or discovery queries, to which the AI layer responds with predictions,

classifications, or risk assessments enriched with trust-related information such as confidence levels and explainability outputs.

The integration supports a continuous feedback loop in which AI-generated insights inform security decisions, while the outcomes of these decisions are fed back into the system to refine future model behaviour. In the case of ZTSM, this enables automated threat detection, prediction, and mitigation actions that are both adaptive and trustworthy. For PhySec, AI services enhance physical-layer protection by providing intelligent, data-driven mechanisms that respond to evolving threats in real time.

This interaction transforms the role of AI within the architecture from a passive analytical component into an active element of control, directly influencing security enforcement and system adaptation. By combining trustworthy AI outputs with automated orchestration, the ROBUST-6G architecture achieves a fully integrated, closed-loop security paradigm aligned with the requirements of future 6G networks.

Summary

Through the tight integration of adversarial robustness, privacy-preserving federated learning, explainability, and sustainability mechanisms, the Trustworthy AI Service Layer supports the deployment of distributed AI models across heterogeneous 6G environments while maintaining strong guarantees on data protection, resilience, and transparency. The integration with the AI Service Management Layer and the Global Model Repository further ensures full lifecycle governance, enabling traceability, reproducibility, and continuous model evolution.

Exposed via a unified service interface and seamlessly integrated with ZTSM and physical-layer security components, these services enable closed-loop, AI-driven security automation. This allows the system to dynamically detect, assess, and mitigate threats based on trustworthy and context-aware intelligence.

Finally, Table 5-2 explains how the requirements for the Distributed AI-Driven Security domain have been covered. Note that the requirements IDs are defined as we did in D2.2 [R6G24-D22].

Table 5-2: Distributed AI-Driven Security requirements in the ROBUST-6G system

Req. ID	Requirement description	Fulfilled	Brief justification
Application Domain: DISTRIBUTED AI-DRIVEN SECURITY			
R3.1	The ROBUST-6G system should align with international AI/ML ethics guidelines to ensure ethical considerations are embedded in the development and deployment processes.	Yes	The DFL Framework is capable of mitigating certain ethical and privacy risks because it keeps data at the source. It incorporates privacy, robustness, and explainability features into training processes.
R3.2	The ROBUST-6G decentralized learning framework should integrate solutions to reduce carbon emissions by wisely scheduling clients.	Yes	ROBUST-6G worked on the topic publishing several conference and journal papers. However, the solutions are currently not integrated in DFL framework and prototype P1, both related to Use Case 1 - Scenario 1.
R3.3	The ROBUST-6G AI solutions (both centralized and decentralized) should integrate ML models that are energy efficient by design at inference time.	Yes	ROBUST-6G studied and published conference papers on spiking neural networks. A part of the proposed solutions results in integration of an SNN model into the DFL framework, available for training and inference.
R3.4	The DFL framework developed in ROBUST-6G should be able to evaluate accountability, fairness, explainability and robustness in AI/ML models.	Yes	The DFL Framework outputs, via the Global Model Repository (GMR), all AI models generated during training by each node in the federation across the different rounds and epochs, along with the scenario configuration, logs, and output metrics for

			evaluating the parameters of accountability, fairness, explainability, and robustness.
R3.5	ROBUST-6G must employ robust Explainable AI (XAI) practices for threat detection, prediction, and mitigation, increasing transparency throughout these implementation processes.	Yes	The ShaTS (Shapley-based time series) and SHAP module have been implemented and integrated with the DFL Framework through the GMR, enabling time series AI models to produce metrics related to explainability that that are kept for consultation at the GMR.
R3.6	The ROBUST-6G system must provide trustworthiness and robustness enhancement capabilities for AI-driven autonomous adaptations of 6G.	Yes	A solution has been proposed to enhance AI/ML robustness against adversarial attacks within the AI-as-a-Service setup and has been disseminated in a conference paper.
R3.7	The ROBUST-6G DFL framework should supply techniques to prevent attacks that attempt to infer an AI/ML model from spoofing learning messages flowing between federation nodes.	Yes	The DFL framework mitigates these threats through a robust hybrid cryptographic approach. Initial node verification and secure key exchange are performed using RSA, which prevents malicious actors from spoofing federation nodes. Subsequently, all learning messages containing model updates are encrypted in transit using AES. This guarantees data confidentiality, ensuring that intercepted communications cannot be used by attackers to infer the underlying AI/ML model.
R3.8	The ROBUST-6G system should make use of secure communication channels during the process of assessing the trustworthiness of the AI/ML models and the physical and sensing layers.	Yes	Physical layer security as a service described in Section 3.4 can be used to increase the security of communication layers through location validation.
R3.9	The ROBUST-6G system should ensure the privacy of data by implementing Differential Privacy (DP), Secure Multiparty Computation (SMC) or Homomorphic Encryption (HE) techniques in the FL framework.	Yes	ROBUST-6G implements privacy-enhancing federated learning through homomorphic encryption, ensuring that model updates remain encrypted throughout the training process. As a result, privacy attacks—such as Deep Leakage from Gradients—are mitigated, since intermediate model updates are never shared in cleartext with other participating parties.
R3.10	The ROBUST-6G system must maintain detailed logs and audit trails for all stages of the AI/ML model lifecycle, including training, deployment, and updates, to ensure accountability.	Yes	ROBUST-6G successfully implements detailed logging mechanisms throughout the whole process. In case of the DFL Framework, comprehensive logs are generated across all nodes during both the federation initialization and the AI model training phases.
R3.11	The ROBUST-6G system must incorporate extensive robustness testing, including adversarial attack simulations, to ensure model resilience against various threats such as poisoning or evasion attacks.	Yes	Evasion attacks have been considered for image-based ISAC attacks. Poisoning attacks in anomaly detection mechanisms have been avoided only by controlling the data before re-training. Model poising attacks have been integrated into the DFL Framework for validation.

R3.12	The ROBUST-6G system shall use AI/ML to enhance system security by facilitating predictive threat/anomaly detection.	Yes	ML has been used to detect anomalies in industrial networks and in jamming attacks. Metrics and scoring mechanisms to quantify decision confidence in AI-enabled models for threat and intrusion detection have been developed.
R3.13	The ROBUST-6G system should introduce a Security-as-a-Service (SecaaS) based E2E AI/ML driven security framework for 6G.	Yes	The ZTSO enable the definition of Security Services, which are automatically deployed and managed, composed of multiple Security Functions (e.g. AI analytics algorithms) and the establishment of multiple S-CLs using these functions to establish a Security Posture.
R3.14	The ROBUST-6G system must generate and evaluate AI/ML models using a DFL framework for training shared models in a privacy-preserving manner by design.	Yes	The DFL Framework covers the generation of AI models in a privacy-preserving manner with trustworthiness capabilities to evaluate them.
R3.15	The fully DFL framework of ROBUST-6G must be agnostic of any application UC, applicable to any multiparty scenario where shared AI/ML models may be generated.	Yes	By providing the fully trained global models in standardized formats ready for external serving tools, such as BentoML, the DFL framework proves its applicability to any domain. This approach ensures that shared AI models generated in any multiparty environment are entirely decoupled from specific use cases and ready for universal deployment.
R3.16	The ROBUST-6G DFL framework should provide decentralized aggregation capabilities and local AI/ML model testing performed by multiple trusted entities, eliminating centralized aggregation processes that could cause bottlenecks and single point attacks.	Yes	ROBUST-6G provided a DFL framework as a piece of asynchronous software together with the most common aggregation baseline, namely, (decentralized) FedAvg. Two additional aggregation methods based on dual optimization were also incorporated.
R3.17	The AI/ML services associated with the DFL platform should provide a well-defined interface to obtain the results achieved by the AI/ML techniques used.	Yes	ROBUST-6G developed a web interface to monitor real-time metrics produced during the federation, displaying graphs, diagrams, and images of the trained models. These metrics can also be retrieved directly from the GMR via a custom REST API.
R3.18	The ROBUST-6G system must continuously monitor the performance of AI/ML models during the FL process to detect and mitigate any trustworthiness issues.	Yes	ROBUST-6G implements a monitoring solution within the DFL Framework to track the performance, resource consumption, and explainability metrics of the AI models generated during the federation process. To ensure trustworthiness and enhance resilience, it leverages robust aggregation strategies that detect and mitigate anomalous model updates.
R3.19	The ROBUST-6G system must implement mechanisms to ensure fairness in the FL process. This includes identifying and mitigating biases in training data and model updates to ensure that AI/ML models do not	Yes	A framework for auditing AI/ML models has been developed to assess their performance with respect to fairness compliance and could be integrated into the DFL platform.

	unfairly benefit or harm any specific user group.		
R3.20	The ROBUST-6G system shall provide a way to evaluate how inter-domain relationships behave using a reputation-based system approach.	Yes	A decentralized reputation-based mechanism has been designed and published in a journal paper [MMG+26] to evaluate interactions between domains using behavioural metrics. It helps detect and mitigate the impact of malicious nodes by adjusting how much they influence in AI model training.
R3.21	The ROBUST-6G system should provide user-centric controls, enabling users to manage their data and model preferences effectively.	Yes	ROBUST-6G provides user-centric controls through its Exposure Framework and Data Governance layer. Users can dictate security intents via Security Service Level Agreements (SSLAs), manage fine-grained data access using Rego policies at the Policy Administration Point (PAP), and specify model preferences, such as explainability, performance, and resource constraints.

5.3 Zero Touch Security Management

The Zero-Touch Security Management is implemented by the namesake layer in the ROBUST-6G architecture, reported in Section 2. In Figure 15 is shown a more detailed version of such an architecture, which also provide an overview of the high-level interactions between the different functionalities.

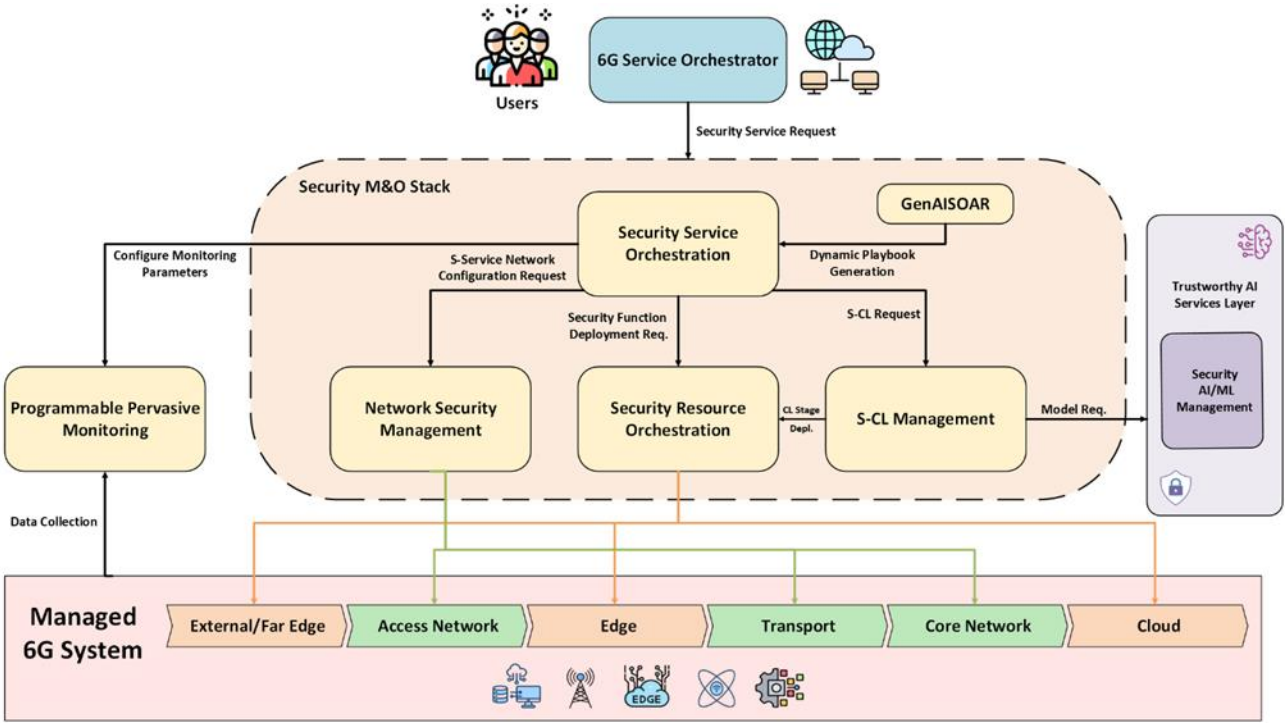


Figure 15 Zero-Touch Security Management layer high-level functional architecture

This architecture has been widely discussed and detailed in the different deliverables of WP4, responsible for the design and implementation of the zero-touch ai driven solutions for the security orchestration and automation. The scope of this section is to provide an overview on these two key functionalities i.e., security orchestration and automation, while, for the details in terms of design, implementation, integration and theoretical basis, the main reference remain the documentation related by WP4 (D4.1, D4.3, and D4.4).

5.3.1 Security Service Orchestration

The orchestration of a security service is realised by involving somehow all the modules represented in the architecture. The first step is the *request.* The request to orchestrate a security service can be performed by i) human users e.g., security administrators, ii) third party orchestrators, e.g., 5G/6G service orchestrators belonging to the Telco infrastructure, or iii) third-party applications. In the case (i), the requestors are supposed to be expert users, and they can be directly performed on the interface exposed by the Security Service Orchestrator. In (ii) and (iii) the request comes from external programs that may belong to other administrative domains (e.g. a third-party Operator). In this case, the request may happen through a dedicated abstraction layer, exposing security function and capabilities.

The Security Service Orchestrator (SSO, implementing the Security Service Orchestration) elaborates the request. In this phase, a number of operations are performed. The requests, coming in the form of a Security Service Layer Agreement (SSLA) is parsed and validated, then the build of the security service starts. Starting from the SSLA, the Security Service Orchestrator first identifies the security functionalities required and the metrics to be monitored, then the security functions capable of implementing the requested functionalities in the target environment. This is done by exploiting the semantic modelling based on a specific security ontology and a set of internal catalogues that provides information regarding the capabilities of the target environment and the security functions available.

At this point, the SSO request the GenAI4SOAR, to generate a security playbook (based on CACAO[CACAO23] specification). The generated playbook may need refinements and, once finalised, the SSO has built all the elements required for the orchestration of a security service: the set of security functions required, where to be placed, and the remediation plan.

The next step is to invoke the proper sub-orchestrator to actually provision the services: the Network Service Manager (NSM) and the Security Resource Orchestrator (SRO) implementing the respective management and orchestration functionalities. The NSO is responsible to enforce specific security configuration to the target network segments that could be RAN, Transport, Core Network. Transport and Core Network are not targeted by ROBUST-6G, while the orchestration on the RAN is limited to specific cases of the Physical Layer Security, as part of the integration in WP6.

The RSO is in charge of provisioning the security functions in the different cloud environments (public, edge, extreme edge). These security functions can be also part of the mechanism of security automation, as explained in the Section 5.3.2 below.

The SSO interacts also with the Programmable Monitoring Platform (PMP, implementing the Programmable Pervasive Monitoring functionality), to request the monitoring of the metrics provided by the semantic layer during the building of the security service starting from the SSLA. The value collected by the PMP can be used in different manners that includes the direct monitoring by e.g., an operator or to feed specific analysis functions and i) generate alert/notifications or ii) trigger predictive/reactive security automation process.

The orchestration workflow described in this section is detailed and demonstrated in the Deliverable D4.4.

5.3.2 Zero-Touch Security Automation

The security automation mechanism is built by exploiting the concept of Security Closed-Loop (S-CL) i.e. a CL as defined by ETSI ZSM [ZSM9-1] but focused on security automation. It may be noted from the Figure 15 that the S-CL is not present in the architecture while it can be finding a functionality called S-CL Management. This because in the ROBUST-6G view, the S-CL is implemented as a cloud application, build by specific security functions that represent its stages: *Monitoring, Analysis, Decision,* and *Execution.* These functions are orchestrable, making the S-CL composable on demand and tuneable for the automation needs of a given security service. The solution adopted makes the automation mechanism highly flexible. The reason why it is not part of the architecture is that the S-CL belongs to its security service: when the service is created, the S-CL is created and terminated when the service is terminated.

The S-CL Management function is responsible for the orchestration of the S-CL. Just like the NSM and the SRO, it receives a request to provision a S-CL for a given security service. The S-CL Management select the

proper CL stages and ask the SRO to deploy them in the target environment. Once deployed, the S-CL is directly manged by the S-CL Management that can start, pause, and terminate a S-CL as well as change its configuration e.g., the automation goal and retrieve information. Additionally, the S-CL Management can implement coordination logic to mitigate and/or avoid conflicts between S-CL that insist on the same set of resources.

The stages can incorporate their own functions or simply act as a proxy towards external services that do this. For example, a Monitoring stage can either collect data itself from the network or just implement a simple logic to interact with the PMP and retrieve the data required for the automation. Similarly, the Analysis stage could either integrate the analysis logic, e.g., an AI/ML algorithm or requests the inference to a third-party specialised AI-driven platform: the ROBUST-6G S-CLs make use of AI/ML in both Analysis and Decision stages. Thus, the S-CL can encompass very complex or very simple stages or a hybrid form: all of them are feasible with this orchestrable loops.

ROBUST-6G S-CLs, according to ETSI ZSM, integrate a fifth element, the Knowledge. The Knowledge is not a stage, is an entity that can store the S-CL configuration and allow communication between non-adjacent stages. In ROBUST-6G is also used to store CACAO playbooks. So, when it is the time to take a decision after the analysis, the Decision stage can i) do nothing or ii) select the proper playbook from the existing set stored in the Knowledge. The Execution stage will execute the playbook selected by the Decision. Considering that the Knowledge is configurable at runtime, the playbook can be refined and uploaded any time, making the remediation more precise and effective.

The CNXW01 is a security orchestrator software component that implements security policies through network, Information Technology (IT), and application services on edge infrastructure. It acts as the Zero-Touch Security Orchestrator (ZTSO) of the ROBUST-6G architecture, implementing its assigned requirements by dynamically translating high-level security policies, coming from CTHA01 component, into actionable configurations for heterogeneous security services across the edge and cloud environments (R4.10). Leveraging ontologies, infrastructure contexts and catalogues, it composes and deploys security functions. For zero-touch mitigation (R4.8), this component works with the CNXW03 component to automate the deployment and forward them to resource orchestrators (addressing R4.5). It also works with the CTHA01 component to integrate threat intelligence (R4.14) from external sources to refine proactive/reactive responses and optimizes resource allocation (R4.15). Finally, CNXW01 and CTHA01 work together to expose lifecycle management APIs (R4.3) to enable external systems (e.g., the 6G exposure framework) to submit security service requests, while coordinating with the Programmable Monitoring Platform (PMP) to validate policy enforcement through real-time monitoring, closing the loop between policy definition, deployment, and verification. Furthermore, the CTHA01 component is a policy-based security orchestrator software that implements security policies and requirements over the network, Information Technology (IT), and application domains. It works with the CNXW01 component to act as a central hub for threat intelligence integration (R4.14) and security investigation (R4.11). It ingests Security Service Level Agreements (SSLAs) from the 6G exposure framework, translating them into enforceable security policies and Key Performance Indicators (KPIs) while dynamically adapting workflows for closed-loop security responses. To fulfil R4.11 (point of investigation), CTHA01 aggregates and correlates security alerts from the PMP, providing security experts with a unified view of policy violations and automated responses. Additionally, it integrates external threat intelligence feeds (R4.14) to enrich its decision-making, enabling proactive adjustments to security policies based on evolving threat landscapes. In other words, the CTHA01 component focus on the SSLA ingestion, translation, and adaptive response workflows, while the CNXW01 complements this with semantic-based orchestration capabilities, ensuring a cohesive and automated approach to security closed-loops enforcement.

The CTHA02 acts as an intelligent incident response engine within the zero-touch security management layer. To fulfil R4.13 (*incident response plans management*), it dynamically ingests security alerts from the PMP and through the Security Orchestrator for alerts contextualization, then leverages generative AI services to detect policy violations and automatically generate adaptive remediation plans in CACAO format (a standardized playbook language for cybersecurity automation). These plans are forwarded to a Security Orchestration, Automation, and Response (SOAR) framework for enforcement, ensuring structured and machine-readable incident handling. For R4.12 (*optimal mitigation strategy*), CTHA02 evaluates remediation actions based on effectiveness and efficiency, balancing security objectives (e.g., threat containment, minimal

service disruption) with operational constraints (e.g., resource availability, compliance rules). The generated plans are also validated by cybersecurity experts, enabling human-in-the-loop oversight to refine AI-driven decisions. By integrating adaptive, AI-assisted response generation with standardized playbook execution, CTHA02 ensures that mitigation strategies are both automated (aligning with zero-touch principles) and optimized for real-world operational contexts.

The CENS01 component addresses R4.15 (Efficient Resource Allocation) by enabling i) the evaluation of Sensing Trustworthiness in CENS03 (e.g., based on the achievable localization accuracy for the estimated signal to interference and noise ratio and the number of antennas at the estimator); ii) the delivery of a lightweight Authentication and Key Agreement (AKA) solution optimized for resource-constrained edge devices, implemented in components CENS04 and CENS05 respectively, thus ensuring minimal computational overhead while maintaining resistance to eavesdropping, injection, jamming, tampering, and impersonation attacks . Importantly, CENS01 is part of the monitoring part of the Physical Layer Security Closed Loop (PLS-CL) and to R4.12 (Optimal Mitigation Strategy) by embedding security-by-design principles into the authentication and key exchange process by feeding the decision process for closing the loop the PLS-CL; in the current implementation this exemplified through power adaptation in case of jamming / tampering attacks detected. Thus, CENS01ensures that only trusted nodes participate in communication, reducing attack surfaces and enabling efficient, effective threat containment without excessive resource consumption.

The CUMU01 component is designed to provide a flexible and comprehensive monitoring capability across heterogeneous environments. It integrates a variety of tools within its Data Collection Module to gather information from data sources at service, infrastructure and network layers (R4.16), while a dedicated Communication Bus module enables sharing internal information in real-time with the modules and with the Near-real Time Data Retrieval API (R4.17). In addition, CUMU01 implements a Configuration Manager module with a REST API to receive configurations from the ZTSO and (re)configure the tools located within the CUMU01 modules (R4.18), supported by several REST APIs for extracting information at various stages of processing (R4.19). The Configuration Module accepts new configurations in Sigma Rule or JSON format (R4.20). In addition, a GUI supports interaction with the modules, allowing configuration and management (R4.21).

From a security and data management perspective, CUMU01 ensures controlled and scalable operations. It uses a role-based system to manage access to viewing and configuration, generating highly secure usernames and passwords for internal tools (R4.22), while the Configuration Module stores all configurations made via its API in the Configuration Data Storage (R4.23). The platform supports long-term, medium-term and real-time data storage for a variety of resources and applications using InfluxDB, MongoDB and RedisDB (R4.24), and enables data exposure through several REST APIs at different stages of processing (R4.26). Monitoring coverage across environments is achieved through a containerised instance of Tshark deployed on each network segment (R4.27), and visualization of historical data is provided via a Grafana-based GUI (R4.28). Furthermore, CUMU01 facilitates interoperability through a Historical Data Retrieval API for exporting long-term data (R4.29) and a Near Real-time Data Retrieval API using WebSocket for continuous streaming (R4.30), while its Data Aggregation and Normalisation Module includes an alarm system to generate alerts when unexpected behaviour is detected (R4.31).

Table 5-3 explains how the requirements for the Zero-touch security management domain have been tackled. It is worth mentioning that the requirements ID are the same we used in D2.2 [R6G24-D22].

Table 5-3: Zero-touch security management requirements in the ROBUST-6G system

Req. ID	Requirement description	Fulfilled	Brief justification
Application Domain: ZERO-TOUCH SECURITY MANAGEMENT			
R4.1	The ROBUST-6G platform must provide zero-touch integrated security management in multi-tenant distributed AI deployments.	Partially	The ZTSO is able to orchestrate distributed AI deployments in various environments, by selecting AI models in its catalog. The process is zero-touch, the selection and the orchestration being fully autonomous.

			However, even if the ZTSO does not have a built-in feature to manage multiple tenants to deploy AI, it is multi-tenant ready by its micro-service architecture, with which an additional fronted component for AAA capabilities can be connected to, handle authentications and RBAC management.
R4.2	The ROBUST-6G platform should be able to manage security service requests including Security Policies or Security Service Level Agreements (SSLAs).	Yes	The ZTSO has a component called SSLA Manager that handles the security policies in SSLA format by validating them, parsing them and managing them with a dedicated database.
R4.3	The ROBUST-6G platform must define and provide a set of APIs specific for the lifecycle management of security services.	Yes	The ZTSO exposes a set of APIs for the LCM of security services.
R4.4	The ROBUST-6G platform should be able to operate in multiple environments (edge, fog, cloud).	Yes	Thanks to the ZTSO capabilities to remain agnostic regarding the orchestrated environments (edge, fog, cloud), the ZTSP is able to operate in multiple environment. The security policies define requirements not regarding the environment, and the decomposition of the policy in security components to run is made thanks to the catalog, that can hosts lightweight or powerful components that can fit both constrained or permissive environments. The selection of adapted security components within the catalog is made thanks to the Context Manager, a ZTSO components which is aware of the different orchestrated environment. Finally, thanks to the resource orchestrators connected to the ZTSO, these components can seamlessly be orchestrated in various environments.
R4.5	The ROBUST-6G platform should have different orchestrators (Security, Resources, Network) all connected.	Partially	The ZTSP provides orchestrators for security (ZTSO) and resources (S-RO) that are interconnected.
R4.6	The ROBUST-6G platform must support multiple closed loops and avoid conflicting configurations via a priority mechanism.	Partially	The Closed Loop governance provides mechanisms for auto-coordination of S-CL via priority mechanisms, will be demonstrated in UC2.
R4.7	The ROBUST-6G system should implement and use the zero-touch platform for threat detection and alarm generation.	Yes	The PMP implements the Alert and Notification Module using Snort3 as a rule-based IDS network traffic detector and analyser to identify security alerts. The module is also responsible for notifying these alerts via the Communication Bus module and for storing in the Medium-term Data Storage, so that they are accessible to external components through respective REST APIs.

R4.8	The ROBUST-6G system should implement and use the zero-touch platform for deploying corrective/mitigation actions.	Yes	The ZTSP is capable of deploying S-CLs that can, in a zero-touch fashion, deploy corrective and mitigation actions.
R4.9	The ROBUST-6G system should be able to predict incidents and impose corrective actions according to the predicted threat.	Yes	A module has been developed to predict potential threats within the next five minutes based on current data. The model was evaluated using the CIC-ToN-IoT testing dataset and achieved an accuracy of 95.36%.
R4.10	The ROBUST-6G platform must be able to manage heterogeneous security services with different requirements and capabilities.	Yes	The ZTSO, using the security orchestrator ontology and the related knowledge graph, is capable of managing heterogeneous security functions and services.
R4.11	The ROBUST-6G platform should provide a point of investigation for security experts to visualise security events and eventually the automated response is taken.	Yes	The ZTSO has modules for the inspection of S-CL alerts and reports related to the mitigations applied.
R4.12	The ROBUST-6G platform should provide optimal mitigation in terms of effectiveness and efficiency, considering security objectives and constraints.	Yes	The proposed PHY-CL can be used to mitigate certain types of attacks, e.g., jamming, through adaptation of resource allocation. The ZTSO is capable of generating tailored CACAO remediation playbooks considering security objectives and constraints, including physical layer.
R4.13	The ROBUST-6G platform should provide the possibility to define, modify and delete incident response plans in an incident response playbook.	Yes	ROBUST-6G is composed with features to generate autonomously and dynamically, or by assisting a human to generate more easily, remediation playbooks in CACAO format.
R4.14	The ROBUST-6G platform should provide the ability to update and integrate from different sources, the Zero-Touch Security orchestrator threat intelligence.	Yes	The ZTSO threat intelligence is located in the ZTSO Knowledge graph and in the GenAI4SOAR. In the KG the threat intelligence is provided in terms of available activities and functions, and in the GenAI4SOAR, different CTI are integrated and used to generate CACAO Playbooks.
R4.15	The ROBUST-6G platform should be aware of target environment resources and should be able to suggest a resource allocation strategy to targets while responding a threat.	Yes	ROBUST-6G propose a physical-layer closed loop enabling adaptive resource allocation via feedback, alongside ZTSO-generated CACAO remediation playbooks based on available resources and security functions, and a risk-aware resource allocation framework incorporating nonlinear event evaluation and subjective perception, which can be integrated into the SO.
R4.16	The ROBUST-6G monitoring platform should have multiple types of collectors to ensure flexible monitoring from several heterogeneous data sources at runtime.	Yes	The PMP offers a variety of tools within its Data Collection Module to gather information from data sources at service, infrastructure and network layers.
R4.17	The ROBUST-6G monitoring platform should have a communication bus to forward the secure data parameters from collectors to the preprocessing modules.	Yes	The PMP has a Communication Bus module for sharing internal information in real-time with the modules and with the Near-real Time Data Retrieval API, which

			enables data to be made available to external components.
R4.18	The ROBUST-6G monitoring platform should have an entity in charge of interpreting the security requirements coming from the Security Orchestrator in order to deploy appropriate monitoring tools.	Yes	The PMP implements a Configuration Manager module with a REST API to receive configurations from the ZTSO and (re)configure the tools located within the modules.
R4.19	The ROBUST-6G monitoring platform shall provide a mechanism to enable external components or platform admin to interact with the Programmable Monitoring Platform (PMP).	Yes	The PMP offers several REST APIs for configuring tools and extracting information at various stages of processing.
R4.20	The ROBUST-6G monitoring platform should support the on-demand configuration of their internal modules such as the Data Aggregation, Communication Bus, or Data Collection.	Yes	The PMP Configuration Module ensables reconfiguration of tools within modules. The Configuration Module's API can accept new configurations in Sigma Rule or JSON format.
R4.21	The ROBUST-6G should enable the Platform Admin to do the configuration of the PMP via a GUI.	Yes	The PMP implements a GUI for interacting with the modules, allowing the configuration and management of the modules.
R4.22	The ROBUST-6G monitoring platform should have access control mechanisms to verify users trying to visualize data or add configurations have the privileges.	Yes	The PMP uses a role-based system to manage access to viewing and configuration. For internal tools, the system generates highly secure usernames and passwords.
R4.23	The ROBUST-6G monitoring platform should enable a database to store configuration parameters of monitoring tools or internal platform components.	Yes	The PMP Configuration Module stores all configurations made via the Configuration Module API by the ZTSO and the administrator in the Configuration Data Storage.
R4.24	The ROBUST-6G monitoring platform should enable a database to store collected and processed data in a scalable and secure manner. It also supports long-term storage for historical analysis.	Yes	ROBUST-6G proposed a cross-layer trustworthiness evaluation and applied it to misinformation attacks in VANETs. PMP is equipped to provide long-term, medium-term and real-time data storage for a variety of resources and applications. InfluxDB, MongoDB and RedisDB are the tools used for the aforementioned data storage modules.
R4.25	The ROBUST-6G monitoring platform should apply semantic techniques to understand the context, meaning, and significance of the monitored raw data, generating new features from the raw and correlated data that are more informative for prediction algorithms.	Yes	The semantics-aware scheduling framework evaluates the significance of model updates at the network edge using the Version Age of Information and a lightweight feature-based dissimilarity proxy. These semantic techniques extract informative features from raw model representations, enabling context-aware prioritization of contributions for prediction algorithms.

R4.26	The ROBUST-6G monitoring platform shall enable the export of data and report to external systems or for offline analysis.	Yes	PMP implements several REST APIs for data exposure at different stages of processing. These APIs are directly related to long-term, medium-term and real-time data stores.
R4.27	The ROBUST-6G monitoring platform should provide a closed set of monitoring tools to ensure the proper acquisition of security params from network segments such as extreme-edge, edge, and cloud.	Yes	The PMP Data Collector Module uses a containerised instance of Tshark as a trace analyser on each network segment, enabling great flexibility in deployment regardless of the environment.
R4.28	The ROBUST-6G monitoring platform should support visualization capabilities for its long-term data storage in order to observe historical data, patterns, or analyse potential plots.	Yes	The PMP Data Storage GUI provides a graphical user interface using Grafana to visualise historical data stored in the Long-term Data Storage.
R4.29	The ROBUST-6G monitoring platform shall be capable of sharing its long-term data with external ROBUST-6G modules such as Data Management, Analysis Engines, or Alerting and Notification to perform more sophisticated activities.	Yes	The PMP Historical Data Retrieval API is reponsible for exporting historical data from the Long-term Data Storage to the external modules.
R4.30	The ROBUST-6G monitoring platform shall support external modules to consume real-time data to perform quick reactions or actions for their internal objectives.	Yes	PMP's Near Real-time Data Retrieval API provide access to real-time data from the Real-time Data Storage system via WebSocket in order to continuously stream the latest information.
R4.31	The ROBUST-6G monitoring platform should generate alerts in case of unexpected behavior in aggregated information, stored in a Time Series Database of the PMP or the data pushed in the Data Fabric.	Yes	The PMP Data Aggregation and Normalisation Module includes an alarm system to generate alerts in case unexpected behaviour is detected in the aggregated numerical data.

5.4 Physical Layer Security

5.4.1 Physical Layer Security Closed Loop (deep dive)

The physical layer closed loop (PLS-CL) serves towards the automation of the trustworthiness and resilience of the physical and sensing layers, extending artificial intelligence-based evaluation mechanisms down to the infrastructure level. In this context, the PLS-CL represents a fundamental architectural element that enables continuous monitoring, evaluation, and adaptation of security mechanisms in real time. The architecture, illustrated in Figure 16, highlights the interaction between physical layer components, AI-driven analysis modules, and control mechanisms that collectively form a feedback loop.

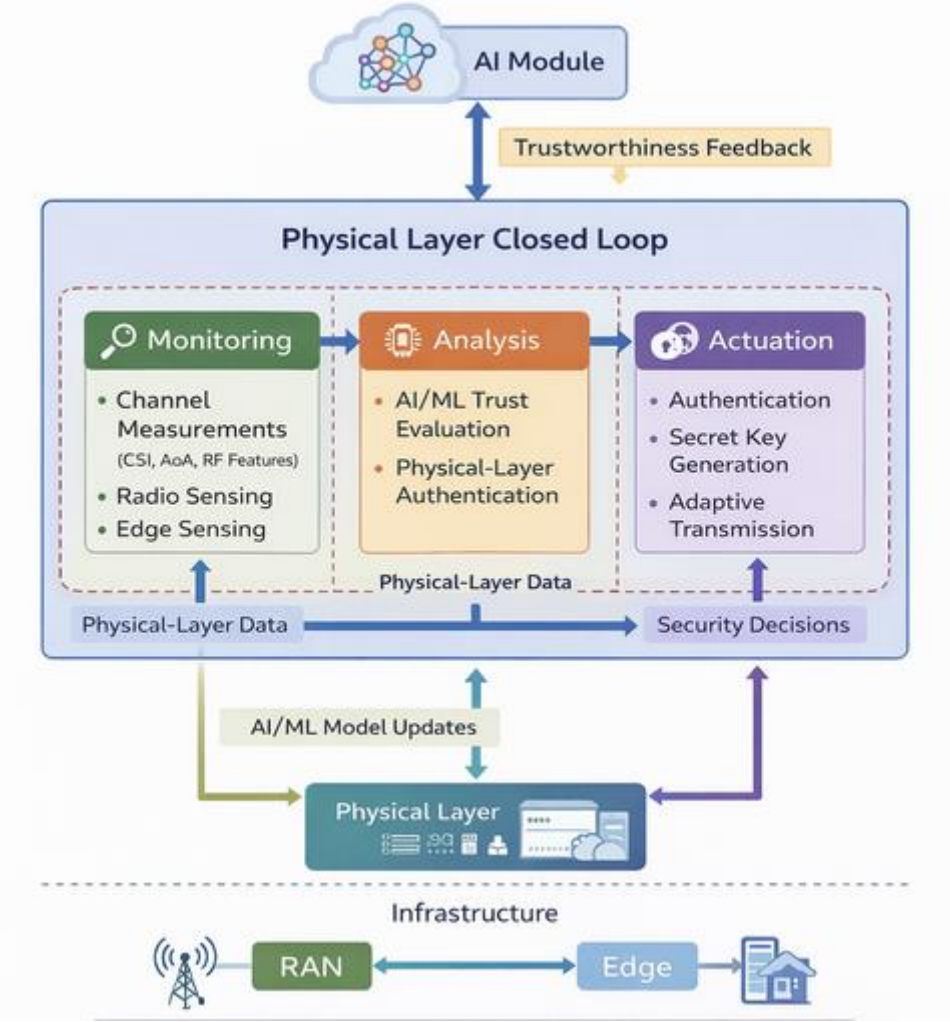


Figure 16: Physical Layer Closed Loop

The physical layer closed loop is structured into three main stages: monitoring, analysis, and actuation. These stages operate continuously and collaboratively to ensure that the system can detect anomalies, evaluate trustworthiness, and respond appropriately using physical layer security techniques.

The monitoring stage is responsible for collecting data from the physical environment and communication processes. This includes channel state information, radio frequency fingerprints, signal strength indicators, and spatial characteristics such as angle of arrival. Data is gathered through distributed sensing components embedded in the radio access network, including base stations, edge devices, and specialized sensing modules. In some cases, simulation tools are also used to complement real-world measurements, particularly during validation. The primary objective of this stage is to provide a continuous and reliable stream of data that reflects the current state of the wireless environment and the behaviour of participating devices.

Following data acquisition, the analysis stage processes the collected information to derive meaningful insights and trustworthiness metrics. This stage integrates machine learning and signal processing techniques to identify patterns, detect anomalies, and classify communication entities. For example, physical layer authentication mechanisms can be applied to verify device identities based on unique radio signatures, while anomaly detection models can identify suspicious behaviour indicative of attacks such as spoofing or signal manipulation. The analysis stage also computes trust scores that quantify the reliability of devices and communication links. These scores serve as the basis for subsequent decision-making processes. By combining data-driven models with domain-specific knowledge of wireless systems, this stage enables adaptive and context-aware evaluation of security conditions.

The actuation stage translates analytical outcomes into concrete actions at the physical and network levels. Based on the trustworthiness assessments, the system can trigger various security mechanisms, including authentication procedures, secret key generation, and adaptive transmission control. Physical layer security

techniques such as key generation from channel randomness can be employed to establish secure communication channels, while suspicious devices may be isolated or denied access. The actuation stage interacts directly with network control functions and hardware components, ensuring that decisions are implemented with minimal latency. This real-time responsiveness is critical for maintaining system integrity in dynamic 6G environments.

The flow of information within the closed loop follows a continuous cycle. Data collected during the monitoring stage is transmitted to the analysis stage, where it is processed and evaluated. The resulting decisions are then executed in the actuation stage, which modifies the system behaviour accordingly. The effects of these actions are subsequently observed again through the monitoring stage, completing the feedback loop. This iterative process enables the system to adapt dynamically to changing conditions and emerging threats.

The integration of the physical layer closed loop within the broader 6G architecture provides several advantages. It allows for low-latency security mechanisms that complement higher-layer approaches, reduces computational overhead by leveraging inherent physical properties of the wireless channel, and enhances overall system resilience through continuous adaptation. Furthermore, the use of AI-driven analysis ensures that the system can evolve over time, improving its ability to detect and respond to sophisticated threats.

In conclusion, the physical layer closed loop in Use Case 1.2 represents a comprehensive approach to security and trustworthiness in 6G networks. By combining monitoring, analysis, and actuation into a unified framework, it enables real-time evaluation and enforcement of security policies at the physical layer. This approach is essential for addressing the challenges of highly dynamic and distributed 6G environments, where traditional security mechanisms alone may not be sufficient.

ROBUST-6G developed multiple components and research contributions to address the defined security requirements. The work carried out is summarized below.

5.4.2 Envisioned Applications

This section provides a comprehensive overview of the key research contributions of the ROBUST-6G project to address the different requirements in terms of security, privacy and trust at the physical layer. It covers developed solutions spanning physical layer security, authentication and key agreement, trustworthy sensing, anomaly detection, and secret key generation, as well as advanced mechanisms for attacker detection and mitigation. In addition, it highlights the integration of machine learning techniques, including supervised and unsupervised learning approaches, to enable adaptive and data-driven security mechanisms.

Table 5-4 explains how the requirements associated with the Physical layer security domain have been addressed. Note that the Requirements ID is the same as that defined in D2.2 [R6G24-D22].

Table 5-4: Physical layer security requirements in the ROBUST-6G system

Req. ID	Requirement description	Fulfilled	Brief justification
Application Domain: PHYSICAL LAYER SECURITY			
R1.1	The ROBUST-6G system must provide PLS based security schemes for 6G leveraging massive Multi-Input Multi-Output (mMIMO), RIS, dMIMO.	Yes	This requirement is addressed through the development of physical layer security (PLS) solutions for massive MIMO systems and the investigation of the impact of reconfigurable intelligent surfaces (RIS) on the feasibility of attacks on angle of arrival (AoA)-based authentication.
R1.2	The ROBUST-6G system must provide low latency and low footprint authentication and key agreement protocols for the considered UCs.	Yes	This is supported by the design of protocols for authentication and key agreement of static nodes in UC1.2, in addition to the components CENS04 and CENS05.

R1.3	The ROBUST-6G system must include a technique for the identification of false base stations.	Partially	This can be supported through the use of 2D AoA-based authentication combined with time-of-flight measurements for BS to BS authentication in the wireless backhaul, which can allow identifying false BS. However, this requirement is only partially addressed, as experiments have not yet been performed due to the lack of suitable real datasets.
R1.4	The ROBUST-6G system must provide physical-layer based solutions for localization privacy.	Partially	Localization privacy preservation has been tackled using approaches based on channel charting and CSI pre-distortion to ensure differential privacy. In addition, obfuscation-based methods have been proposed to further protect location information while optimizing latency, also leveraging differential privacy techniques.
R1.5	The ROBUST-6G system must provide physical-layer based solutions for trustworthy sensing.	Yes	Trustworthy sensing at the PHY is ensured by evaluating sensing reliability for multiple indicators, including AoA estimation, time of flight and received signal strength.
R1.6	The ROBUST-6G system must provide physical-layer based solutions for generalized anomaly detection.	Yes	Anomaly detection at the PHY is achieved through the detection of jamming and location spoofing attacks, where spoofed locations can be verified against application layer positioning such as GNSS and node messaging, with evaluations performed via vehicular ad hoc network (VANET) simulations.
R1.7	The ROBUST-6G system should include online attack and attacker identification and mitigation solutions with the help of online AI/ML learning mechanisms.	Partially	ROBUST-6G worked on online AoA based authentication that can be used to identify spoofing attacks; online jamming attack detection through monitoring of signal to noise and interference in different spatial locations.
R1.8	The ROBUST-6G system should include ML solutions to learn how to configure the devices and what signals they should transmit to improve the confidentiality of transmissions using wiretap coding.	Partially	ROBUST-6G worked on theoretically achievable secrecy rates developed using finite block length codes. These could be integrated into future machine learning solutions to learn how to configure devices and determine which signals to transmit to enhance transmission confidentiality using wiretap coding.
R1.9	The ROBUST-6G system should integrate confidentiality solutions with authentication and Secret Key Generation (SKG) to optimize the resources (in terms of energy consumption, but also communication overhead).	Partially	Joint resource optimization and confidentiality is investigated through the integration of authentication and key agreement into resource optimization, alongside theoretical analyses for confidentiality.
R1.10	The ROBUST-6G system should include solutions for authentication based on	Yes	Several solutions based on challenge-response authentication have been proposed in the project, including those that

	challenge-response approach operating at the PHY.		use the configuration of an intelligent reflective surface and the position of a moving drone as challenges.
R1.11	The ROBUST-6G system should include solutions for the SKG techniques robust against eavesdropping, injection (man-in-the-middle), spoofing, and jamming.	Yes	Robust SKG techniques are addressed through a secret key generation solution that can handle worst-case eavesdropping scenarios (e.g., adversaries located at one wavelength distance), injection attacks through pilot randomization and jamming via adaptive privacy amplification (hashing rate).
R1.12	The ROBUST-6G system should be robust against adversarial attacks against ML models used for PHY security.	Partially	ROBUST-6G proposed novel attacks based on adversarial ML to evaluate the robustness of Integrated sensing and communication (ISAC) systems, considering different scenarios and attacker capabilities. ROBUST-6G has also proposed a technique to analyze the spectrum of cellular networks to detect jamming attacks. A Carlini-Wagner (C-W) attack has been investigated, together with an adversarial training technique to enhance the detector's robustness.
R1.13	The ROBUST-6G system should include privacy-preserving solutions while following principles such as privacy by design, local processing, and confidential computing, as well as anonymization, pseudonymization, obfuscation, and perturbation.	Partially	ROBUST-6G proposed an obfuscation-based solution for localization privacy preserving and latency optimization leveraging differential privacy. Obfuscation-based solutions have been developed to further protect location information while optimizing latency, leveraging differential privacy techniques.
R1.14	The ROBUST-6G system should include solutions for positioning privacy at the PHY using channel charting.	Yes	ROBUST-6G developed approaches based on channel charting and CSI pre-distortion to achieve differential privacy.
R1.15	The ROBUST-6G system should include anomaly detection and restoration techniques inspired to image forensics based on an image of the environment obtained from both in-band and opportunity signals.	Yes	ROBUST-6G proposed unsupervised learning frameworks, including a solution based on generative adversarial networks for Cloud radio access networks, to detect unseen contention anomalies by analyzing cross-layer key performance indicators. Furthermore, the vulnerability of Integrated Sensing and Communication (ISAC) systems to adversarial machine learning attacks that trick receivers into misidentifying target locations has been evaluated. Finally, **image-based approaches were applied to the detection jamming attacks.**
R1.16	The ROBUST-6G system should include generalized cross-layer anomaly detection techniques using continuous learning and unsupervised learning.	Partially	Continuous learning approaches for AoA authentication on a real outdoor dataset have been built to guarantee general cross-layer anomaly detection techniques.

R1.17	The ROBUST-6G system should include federated solutions operating across several devices (both users and network components) for spatial correlations in detecting federated attacks (e.g., jamming covering an area or distributed attacks).	Partially	ROBUST-6G introduced a solution for authentication based on federated learning, where multiple base stations of a 6G network use local models to determine if the transmitter is transmitting from an authorized area or not, using the estimated CSI.
R1.18	The ROBUST-6G system should create a database of attacks at the physical layer (PHY) and sensing, generated by contributions from all partners participating in the PLS-related WP.	Yes	Demostrated in D5.1 [R6G24-D51].
R1.19	The ROBUST-6G system should develop an AI-enabled library of known RF fingerprints for different identified attacks.		Demostrated in D5.1 [R6G24-D51].
R1.20	The ROBUST-6G system should leverage the Angle of Arrival (AoA) and Channel State Information (CSI) to help ML models estimate the location of threats with less training data.	Yes	Threat localization using AoA and CSI is supported through machine learning approaches for AoA-based node localization in outdoor environments, including both offline supervised learning models and online models that require only 10% of the available dataset, as well as adaptive pre-processing techniques for CSI-based node authentication.
R1.21	The ROBUST-6G system should support the adaptive migration of RF fingerprints among base stations in smart city environments to enable seamless and secure communication for IoT devices across different network nodes.	Partially	ROBUST-6G developed an ADDA-based receiver-invariant RFFI model demonstrating effective cross-receiver fingerprint adaptation with minimal unlabelled data, partially addressing seamless fingerprint migration across heterogeneous network nodes.
R1.22	The ROBUST-6G system should develop predictive models to anticipate changes in RF fingerprints for low-power, infrequently communicating IoT sensors, enabling privacy-preserving and robust sensing.	Yes	ROBUST-6G have worked on CSI pre-processing to address changes over time.

6 Conclusion

The ROBUST-6G project aims to deliver a unified, AI-driven security platform for 6G networks that is autonomous, adaptive, and embedded across every layer of the network. This deliverable presents the final version of the ROBUST-6G architecture and the ROBUST-6G data fabric, consolidating the design work carried out across the project into a complete and integrated security solution.

The final architecture is organized around six interconnected layers that together form a single, end-to-end security platform. The Programmable Pervasive Monitoring Layer continuously collects data from across the network and forwards it to the Data Management Platform, which aggregates, governs, and distributes it to the layers that depend on it: the Zero-Touch Security Management Layer, the Trustworthy and Sustainable AI Services Layer, and the Physical Layer Security Closed Loop. The Zero-Touch Security Management Layer consumes this data to detect and predict threats and autonomously orchestrates mitigation actions across multiple domains through security closed loops, all without human intervention. The Trustworthy and Sustainable AI Services Layer provides the AI models that power these closed loops, ensuring that security decisions are robust, explainable, privacy-preserving, and energy-efficient. The Physical Layer Security Closed Loop applies the same monitor-analyse-actuate principle at the air interface, enforcing rapid mitigation

against threats such as jamming, spoofing, and passive eavesdropping. All these capabilities are made accessible to external consumers, verticals, and developers through the Exposure Framework, which provides standardized, CAMARA-aligned APIs that abstract the underlying complexity of the platform.

The ROBUST-6G data fabric reinforces this integration at the data level. Through semantic technologies, knowledge graph-based data integration, and policy-driven access control, the data fabric ensures that security-relevant data is managed in a governed, interoperable, and trustworthy manner across all participating domains. It serves as the common data foundation on which the intelligence and automation of the platform depend.

The four principal enablers of the platform, namely the Data Management Platform, the Trustworthy AI Services, the Zero-Touch Security Management, and the Physical Layer Security, are each decomposed into their constituent components and validated against the requirements originally defined in D2.2. Together, these enablers demonstrate that the architectural design presented in this document is not only coherent at the system level but also grounded in concrete, implementable, and requirement-compliant building blocks.

The platform's capabilities extend beyond individual components and manifest as concrete, operational services. Transparency in AI-driven security decisions, energy-efficient distributed learning across the edge-cloud continuum, intent-based security orchestration exposed to external consumers through the Exposure Framework, and runtime enforcement of spatial security guarantees at the air interface all emerge from the integrated operation of the platform's layers. It is precisely this interdependence that gives ROBUST-6G its end-to-end security value.

The architecture and data fabric defined in this deliverable serve as the consolidated technical reference for the integration and validation work that follows. The platform will be validated through a structured process that progresses from individual components and functional flows to end-to-end scenario-level execution, with a set of prototypes serving as the primary demonstration across the project's use cases and broader 6G security capabilities.

As 6G networks move closer to full-scale deployment with AI as a native architectural component, the need for security solutions that are not only powerful but coherent, scalable, and truly autonomous becomes increasingly critical. The ROBUST-6G architecture addresses this need by embedding security intelligence across every layer of the network, from the air interface to the application-facing exposure framework, and by ensuring that these layers operate as a single, self-reinforcing system. This is the core contribution of this deliverable and the foundation on which the ROBUST-6G platform stands.

References

[3GP26a] 3GPP, "CT3 Exposure Framework (CAPIF) APIs". Available: https://www.3gpp.org/technologies/ct3-exp-fr-apis, accessed Apr. 2026.

[3GP26b] 3GPP "Functional architecture and information flows for AIML Enablement Service" Available: https://www.3gpp.org/DynaReport/23482, accessed Apr. 2026.

[3GP26c] 3GPP "Artificial Intelligence/ Machine Learning (AI/ML) management" Available: https://www.3gpp.org/DynaReport/28105, accessed Apr. 2026.

[3GP26d] 3GPP "Study on Artificial Intelligence / Machine Learning (AI/ML) management enhancements" Available: https://www.3gpp.org/DynaReport/28882, accessed Apr. 2026.

[AH11] D. Allemang and J. Hendler, "Semantic Web for the Working Ontologist", Elsevier, 2011. doi: 10.1016/C2010-0-68657-3.

[Apa26a] Apache Jena, "Apache Jena". Available: https://jena.apache.org, accessed Apr. 2026.

[Apa26b] Apache Camel, "Apache Camel". Available: https://camel.apache.org, accessed Apr. 2026.

[Apa26c] Apache APISIX, "Apache APISIX". Available: https://apisix.apache.org/, accessed Apr. 2026.

[BHL01] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities", Scientific American, May 2001.

[CACAO23] CACAO Security Playbooks v2.0 - OASIS Open,Online access on September 23rd, 2025: https://www.oasis-open.org/standard/cacao-security-playbooks-v2-0/

[CAM26a] CAMARA Project, "CAMARA Project". Available: https://camaraproject.org/, accessed Apr. 2026.

[CAM26b] CAMARA Project, "Model as a Service". Available: https://camaraproject.org/model-as-a-service/, accessed Apr. 2026.

[Com26] European Commission, "DCAT application profile for data portals in Europe (DCAT-AP)". Available: https://op.europa.eu/en/web/eu-vocabularies/dcat-ap, accessed Apr. 2026.

[GLP+23] R. García-Castro, M. Lefrançois, M. Poveda-Villalón, and L. Daniele, "The ETSI SAREF Ontology for Smart Applications: A Long Path of Development and Evolution", in Energy Smart Appliances, 1st ed., A. Moreno-Munoz and N. Giacomini, Eds., Wiley, 2023, pp. 183–215. doi: 10.1002/9781119899457.ch7.

[GSC+23] M. Grassi, M. Scrocca, M. Comerio, A. Carenini, and I. Celino, "Composable semantic data transformation pipelines with chimera", in Proc. 4th International Workshop on Knowledge Graph Construction (KGCW'23), CEUR Workshop Proceedings, vol. 3471, 2023, pp. 1–13. Available: https://ceur-ws.org/Vol-3471/paper9.pdf

[GSM26] GSMA, "GSMA Open Gateway initiative". Available: https://www.gsma.com/solutions-and-impact/gsma-open-gateway/, accessed Apr. 2026.

[HEX24-D53] Hexa-X-II Project, "Initial design and validation of technologies and architecture of 6G devices and infrastructure," Deliverable D5.3, Feb. 2024.

[Ide26] K. U. Idehen, “Semantic web layer cake tweak, explained”. Available: https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab, accessed Apr. 2026.

[Igl+23] A. Iglesias-Molina et al., “The RML Ontology: A Community-Driven Modular Redesign After a Decade of Experience in Mapping Heterogeneous Data to RDF”, in The Semantic Web – ISWC 2023, vol. 14266, T. R. Payne et al., Eds., Lecture Notes in Computer Science, Springer Nature Switzerland, 2023, pp. 152–175. doi: 10.1007/978-3-031-47243-5_9.

[Key26] Keycloak, “Keycloak”. Available: https://www.keycloak.org/, accessed Apr. 2026.

[MMG+26] I. Marroquí Penalva, E. T. Martínez Beltrán, M. Gil Pérez, A. Huertas Celdrán, “RepuNet: A Reputation System for Mitigating Malicious Clients in DFL,” Computer Networks, Volume 282, 2026, 112242, ISSN 1389-1286, doi: 10.1016/j.comnet.2026.112242

[Ont26] Ontotext, “GraphDB”. Available: https://graphdb.ontotext.com/documentation/11.0/, accessed Apr. 2026.

[OPA26a] Open Policy Agent, “Open Policy Agent”. Available: https://www.openpolicyagent.org/, accessed Apr. 2026.

[OPA26b] Open Policy Agent, “Policy Language”. Available: https://www.openpolicyagent.org/docs/policy-language, accessed Apr. 2026.

[Ope26] OpenLDAP, “OpenLDAP Software”. Available: https://www.openldap.org/, accessed Apr. 2026.

[PFF+22] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, and R. García-Castro, “LOT: An industrial oriented ontology engineering framework”, Engineering Applications of Artificial Intelligence, vol. 111, p. 104755, May 2022. doi: 10.1016/j.engappai.2022.104755.

[R6G24-D22] ROBUST-6G, “D2.2: Use Cases, Requirements, ROBUST-6G Initial Architecture and Initial ROBUST-6G Dataspace,” Horizon Europe Project, 2024. Available: https://robust-6g.eu/wp-content/uploads/2025/01/ROBUST-6G-D2.2_v1.0-1.pdf

[R6G24-D41] ROBUST-6G, “D4.1: Security Automation for 6G,” Horizon Europe Project, 2024, Grant Agreement No. 101139068. Available: https://robust-6g.eu/wp-content/uploads/2025/01/ROBUST-6G-D4.1_Security-Automation-for-6G_v1.0.pdf

[R6G24-D51] ROBUST-6G, “D5.1: Library of Known PHYs Attacks and PLS Dataset,” Horizon Europe Project, 2024, Grant Agreement No. 101139068. Available: https://robust-6g.eu/wp-content/uploads/2025/01/ROBUST-6G-D5.1_v1.0.pdf

[R6G25-D32] ROBUST-6G, “D3.2: Initial Report on 6G Trustworthy and Sustainable AI Architecture and Requirements for Integrating Selected XAI Measures”, Horizon Europe Project, 2025, Grant Agreement No. 101139068. Available: https://robust-6g.eu/wp-content/uploads/2025/07/ROBUST-6G-D3_2_v1_0.pdf

[R6G25-D43] ROBUST-6G, “D4.3: ROBUST-6G AI/ML Driven Zero-Touch Security Management Platform Consolidated Design,” Horizon Europe Project, 2025, Grant Agreement No. 101139068. Available: https://robust-6g.eu/wp-content/uploads/2026/04/ROBUST-6G-D4.3_v1.0.pdf

[R6G26-D33] ROBUST-6G, “D3.3: Trustworthy and Sustainable AI Prototype,” Horizon Europe Project, 2026, Grant Agreement No. 101139068. Available: https://robust-6g.eu/wp-content/uploads/2026/04/ROBUST-6G-D3.3_v2.pdf

[R6G26-D44] ROBUST-6G, "D4.4 ROBUST-6G AI/ML Driven Zero-Touch Security Management Platform Final Prototype" Horizon Europe Project, 2026, Grant Agreement No. 101139068. Available: https://robust-6g.eu/wp-content/uploads/2026/04/ROBUST-6G_D4.4_v1.0.pdf

[RDF26] RDFLib, "RDFLib: a python library for working with RDF". Available: https://github.com/RDFLib/rdflib, accessed Apr. 2026.

[SCG+24] M. Scrocca, A. Carenini, M. Grassi, M. Comerio, and I. Celino, "Not everybody speaks RDF: Knowledge conversion between different data representations", in Proc. 5th International Workshop on Knowledge Graph Construction (KGCW 2024), CEUR Workshop Proceedings, vol. 3718, Crete, Greece, May 2024.

[SSW+24] B. Siniarski, C. Sandeepa, S. Wang et al., "Robust-6G: Smart, automated, and reliable security service platform for 6G," Proc. 15th Int. Conf. on Ubiquitous and Future Networks (ICUFN), Budapest, Hungary, July 2024, pp. 384–389.

[VER25-D22] VERGE Project, "Final report on VERGE Edge4AI design," Deliverable D2.2, Apr. 2025.

[VDH+23] D. Van Assche, T. Delva, G. Haesendonck, P. Heyvaert, B. De Meester, and A. Dimou, "Declarative RDF graph generation from heterogeneous (semi-)structured data: A systematic literature review", Journal of Web Semantics, vol. 75, p. 100753, Jan. 2023. doi: 10.1016/j.websem.2022.100753.

[Wil+16] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship", Scientific Data, vol. 3, no. 1, p. 160018, Mar. 2016. doi: 10.1038/sdata.2016.18.

[ZSM9-1] ETSI, "Zero-touch network and Service Management (ZSM); Closed-Loop Automation; Part 1: Enablers," ETSI GS ZSM 009-1 V1.1.1, Jun. 2021. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/00901/01.01.01_60/gs_ZSM00901v010101p.pdf