UNIVERSIDAD POLITÉCNICA DE MADRID
GRANT: PID2022-137748OB-C31

A1.D3.2 .- "Data integration, interoperability, and communication without contact."

Digitalization As basic Driver for servitization in Industry and Basic Services" (DADIBAS)

Date: 15/03/2026
Doc. Version: 1.1

Document Control Information

Settings	Value
Document Title:	Deliverable A1.D3.2 "Data integration, interoperability, and communication without contact"
Project Title:	Digitalization As basic Driver for servitization in Industry and Basic Services" (DADIBAS)
Document Author:	Joaquín Ordieres-Meré
Project Owner:	AEI
Project Manager:	Joaquín Ordieres-Meré
Doc. Version:	1.1
Sensitivity:	Public
Date:	15/03/2026

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained.
All reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Miguel Ortega-Mier	CoIP	*Review*	20/03/2026

Document history:

The Document Author is authorized to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author.
Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
1.0	01/03/2026	Joaquín Ordieres-Meré	First version of the Deliverable A1.D3.2
1.1	15/03/2026	Joaquín Ordieres-Meré	Improvements in several sections of the document

Configuration Management: Document Location

The latest version of this controlled document is stored in the OneDrive repository of the project.

TABLE OF CONTENTS

1 Introduction

Project Title:	*"Digitalization As basic Driver for servitization in Industry and Basic Services" (DADIBAS)*		
Initiator:	*Joaquín Ordieres-Meré*	**Organisation / Unit:**	*UPM*
Date of Request:	*07/09/2024*	**Target Delivery Date:**	*31/03/2026*
Type of Delivery:	☒ In-house ☐ Outsourced ☐ Mix ☐ Not-known		

The document presents a theoretical framework and a practical implementation for data integration, interoperability, and contactless communication in industrial environments. It highlights the combination of BLE and LoRaWAN technologies for acquiring data from wearable devices, the use of ontologies to enable semantic interoperability, and the importance of MLOps for continuous model monitoring. These concepts are illustrated through the CeDInt building case study, which integrates a heterogeneous sensor infrastructure, an IoT platform, and semantic access to support advanced analytics and decision-making in line with the Industry 5.0 paradigm.

2 Context

The main objective of this Work Package is to enhance process knowledge by incorporating contextual information derived from human operators interacting with industrial environments. In many operational scenarios, human behaviour introduces a significant degree of variability that traditional monitoring systems fail to capture.

Recent advances in wearable technologies, low-power wireless communication protocols, and mobile applications now make it possible to acquire physiological, positional, and behavioural data in a non-intrusive manner. These data sources can complement existing process- and asset-related information and help close current gaps in monitoring and understanding operational conditions.

The purpose of this work is therefore to investigate how digital tools can support the seamless integration of contextual data obtained from wearable devices into digitalized industrial processes. This includes developing interoperable software components capable of collecting data from heterogeneous wearable devices, aggregating information from multiple workers, and merging these data streams with existing operational datasets.

By doing so, the project expands the concept of asset digitalization to explicitly incorporate the human dimension into digital models. This approach aligns directly with the Industry 5.0 paradigm, in which human–machine collaboration plays a central role in process optimization.

The list of relevant tasks is:

T3.1 – Design of mobile applications for data acquisition from wearable devices.

T3.2 – Workflow definition for data ingestion from multiple workers and devices.

T3.3 – KPI extraction and interoperability workflows.

T3.4 – Data integration and distribution through Distributed Ledger Technologies.

T3.5 – Process model enrichment with behavioural data.

T3.6 – Dissemination.

T3.7 – Reporting and configuration management.

The foreseen deliverables are:

D3.1 – Wearables and digital solutions for context enrichment in predictive models [M24]

D3.2 – Data Integration, Interoperability, and Communication without contact [M30]

D3.3 – Dissemination report and KPIs [M48]

The focus of this document is to address the second deliverable (D3.2), which concentrates on the technological mechanisms required to guarantee interoperability, scalable data ingestion, and reliable communication between wearable devices, mobile applications, and the digital infrastructure supporting the project.

Particular attention is devoted to the development of mobile software capable of interacting with heterogeneous wearable devices through Bluetooth Low Energy (BLE). Such an approach enables the creation of flexible applications capable of collecting data from different commercial devices while maintaining a unified data model. In this context, mobile applications such as the HealthyWear platform illustrate the feasibility of building interoperable solutions capable of retrieving physiological and motion-related data from multiple devices through standardized BLE interfaces.

However, in many industrial scenarios operators are not confined to fixed environments and may operate across extended facilities or outdoor infrastructures. In such cases, conventional short-range communication technologies may be insufficient to ensure reliable connectivity. To address this challenge, complementary communication architectures based on Low Power Wide Area Networks (LPWAN), and in particular LoRaWAN technologies, are considered as an enabling solution for monitoring in mobility conditions. These technologies allow long-range, energy-efficient communication suitable for industrial environments where workers move across large areas or where connectivity infrastructure is limited.

Consequently, this document explores the integration of two complementary technological layers. On one hand, BLE-based mobile applications provide flexible and device-agnostic data acquisition from wearable sensors. On the other hand, LoRaWAN-based communication infrastructures enable extended coverage and robust connectivity for mobile operators in large industrial environments. By combining these technologies within a unified data ingestion and integration architecture, the project aims to enable scalable and interoperable context-aware monitoring solutions.

There are numerous potential application domains where such approaches could be applied, ranging from industrial manufacturing and infrastructure maintenance to logistics or emergency response scenarios. Each domain presents its own constraints in terms of connectivity, data frequency, privacy, and integration with operational data systems. Therefore, this document focuses on defining a technological framework that supports heterogeneous wearable devices, multiple communication protocols, and interoperable data workflows.

By addressing these challenges, the deliverable contributes to the broader objective of integrating behavioural information into digital process models. The resulting data ecosystem enables richer predictive models, improved situational awareness, and enhanced decision-making processes that account not only for asset status but also for the contextual conditions under which human operators interact with industrial systems.

3 Theoretical Framework for Data Integration, Interoperability, and Contactless Communication

3.1 Introduction

The topic of this chapter is the design of a digital infrastructure able to support data integration, interoperability, and communication without contact in environments where multiple assets, software services, sensors, and human operators coexist. In DIGEST, this challenge is not peripheral; it is central to A1.WP3, which explicitly addresses wearable-enabled data collection, multi-user data ingestion, KPI extraction, interoperability, data integration through digital technologies, and the enrichment of process models with human-related information. In particular, task A1-T3.1 focuses on mobile applications for collecting data from heterogeneous wearable devices, A1-T3.2 addresses ingestion from multiple workers and brokers, A1-T3.3 focuses on KPI extraction and interoperability, and A1-T3.4 targets data integration and data availability through DLT-oriented mechanisms. Deliverable A1.D3.2 is therefore the natural place to articulate the theoretical foundations that justify these technological choices.

From a theoretical perspective, the problem can be stated as follows: how can heterogeneous and continuously evolving data streams be converted into reliable, interoperable, and operationally useful information without forcing direct coupling among all system components? This question is increasingly relevant in Industry 5.0 settings, where value creation depends not only on the state of physical assets but also on the behaviour of operators, the flexibility of digital services, and the capacity of the infrastructure to sustain trustworthy distributed decision support [1], [2], [21]–[23]. The answer cannot be reduced to communication protocols alone. It requires a full-stack view in which data capture, semantic transformation, distributed exchange, model lifecycle control, and governance mechanisms are all treated as parts of the same architecture [2], [18]–[24].

A second reason why this topic deserves a dedicated theoretical chapter is that wearable monitoring produces data that are simultaneously rich and difficult to exploit. Wearable devices offer unprecedented opportunities for longitudinal observation of human activity in real contexts, but raw time-series streams are not directly useful for most industrial or health-related decisions. Recent work on semantic gait monitoring makes this point clearly: interpretability requires a transformation from raw inertial streams into structured, time-bounded events that support traceability, longitudinal reasoning, and interoperability with higher-level systems [3]. This separation between raw persistence and semantic representation is essential for auditability and reuse, and the same principle is directly transferable to DIGEST.

Recent industrial AI literature adds a further requirement. Data integration is insufficient if the analytical models consuming these data are not monitored after deployment. Data-driven maintenance and prescriptive analytics require mechanisms for self-monitoring and controlled model updating when conditions change [1], [24]–[28]. In DIGEST, where wearable, process, and asset-related streams are expected to evolve over time, model observability becomes part of the interoperability problem rather than an optional downstream enhancement.

3.2 From raw sensing to integrated digital knowledge

A foundational principle of the chapter is that data integration is not merely a transport problem. In distributed sensing environments, information passes through different stages of abstraction: acquisition, ingestion, persistence, semantic interpretation, modelling, and decision support. The quality of the final knowledge depends on how consistently these transitions are defined [3], [4], [6], [8], [9].

The literature on wearable sensing has repeatedly shown that raw inertial, physiological, or positional streams are high-volume, high-frequency, and difficult to interpret directly [4], [6]–[10]. Classical approaches to gait and activity analysis often relied either on laboratory instrumentation or on signal-processing pipelines tightly coupled to specific datasets. While such approaches can achieve strong local performance, they frequently struggle with longitudinal interpretability, semantic reuse, and deployment across heterogeneous real-world settings [4], [5], [8]–[10].

The semantic digital health pipeline used as a contextual input to this chapter formulates a generalizable principle: raw signals should be preserved as immutable traces, while higher-level semantic events should be derived into a separate, structured representation [3]. This separation enables transparency, reproducibility, and future reuse by downstream analytical services. In DIGEST terms, data acquired through smart insoles, position sensors, heart-rate wearables, or mobile phones should enter an ingestion layer that preserves raw observations, while a semantic layer transforms them into interpretable events such as movement episodes, stress-related signal excursions, operator displacement episodes, or task-aligned behavioural segments.

Such a view also resonates with broader Industry 5.0 literature. Flexible architectures that integrate process, product, and human dimensions without losing transparency are increasingly proposed as the basis for higher-value digital services [2], [21], [23]. DIGEST does not treat wearable information as an isolated subsystem; it treats it as a contextual enrichment layer for process knowledge. Consequently, the challenge of A1.D3.2 is not only to transmit wearable data, but to place those data in a framework where they can be linked to operational states, business KPIs, and decision models.

The architecture described here can be conceptualized as a set of functional layers that separate data acquisition, semantic transformation, and downstream analytical services. This multilayer perspective is summarized in Figure 1, which illustrates the proposed stack that ensures a coherent transition from raw signals to actionable digital knowledge.

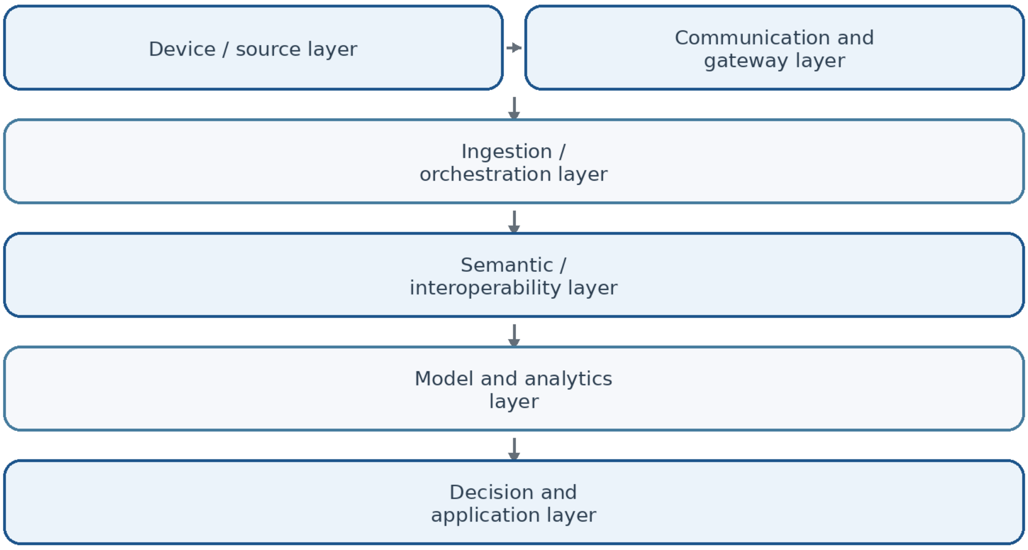

Figure 1. Layered architecture for data integration, interoperability, and communication without contact.

3.3 Interoperability as a multi-layer requirement

Interoperability is frequently invoked as a technical goal, but the literature shows that it comprises at least three distinct layers: technical interoperability, syntactic interoperability, and semantic interoperability [2], [16]–[20], [22]. Each layer must be addressed explicitly if the resulting architecture is expected to survive real deployment conditions.

Technical interoperability concerns whether devices and systems can exchange data at all. In DIGEST, this includes compatibility among wearable devices, mobile gateways, backend services, message brokers, databases, and stakeholder-facing applications. At this level, Bluetooth Low Energy is highly relevant because it is the dominant interface for low-power wearable devices and is widely supported by smartphones and cross-platform application frameworks [11], [12]. BLE offers low energy consumption, ecosystem maturity, and adequate bandwidth for many physiological and motion signals [11]. These advantages explain why app-based collection strategies such as HealthyWear are useful reference points for DIGEST.

However, BLE alone is insufficient in scenarios involving spatial mobility, large facilities, intermittent gateway availability, or the need to reach remote infrastructure. For such cases, LPWAN technologies—especially LoRaWAN—provide an alternative trade-off: lower throughput but longer range and lower transmission energy, making them suitable for wide-area or mobility-sensitive monitoring [13]–[15]. In practical terms, DIGEST should not frame protocol choice as BLE versus LoRaWAN, but as BLE plus LPWAN in a multi-tier architecture where short-range body-area capture and wide-area infrastructure communication serve complementary roles [11], [13], [14].

Syntactic interoperability concerns shared data formats and communication conventions. Even when devices connect correctly, poor harmonization of timestamps, identifiers, session boundaries, metadata fields, or event schemas creates brittle pipelines. The semantic pipeline literature again offers a useful pattern: incoming files contain a metadata header and standardized time-stamped samples, which are ingested into a time-series repository before later semantic processing [3]. For DIGEST, this implies that ingestion services should normalize metadata across devices and users early in the pipeline while preserving enough provenance to trace each semantic event back to its original source.

Semantic interoperability is the strongest and most consequential layer. It addresses whether two components assign the same meaning to exchanged data. Ontology-based approaches have become central here because they enable explicit modelling of domain entities, temporal events, and relationships across heterogeneous systems [16]–[20]. In healthcare, semantic modelling has already been proposed as a way to structure complex smart-health data [16], [17]. In digital-twin research, ontology-driven methods and OBDA frameworks are increasingly seen as key enablers of interoperability, reasoning, and cross-system querying [18]–[20], [22]. For DIGEST, semantic interoperability is especially important because behavioural data must be combined with process data and possibly with asset-centric digital twins. If wearable-derived signals remain trapped in app-specific formats, they cannot enrich process models in a robust way.

These three dimensions—technical, syntactic, and semantic interoperability—form complementary levels within a unified conceptual framework. Figure 2 synthesizes this interoperability stack and highlights how each layer contributes to the robustness and scalability of the architecture defined in A1.D3.2.

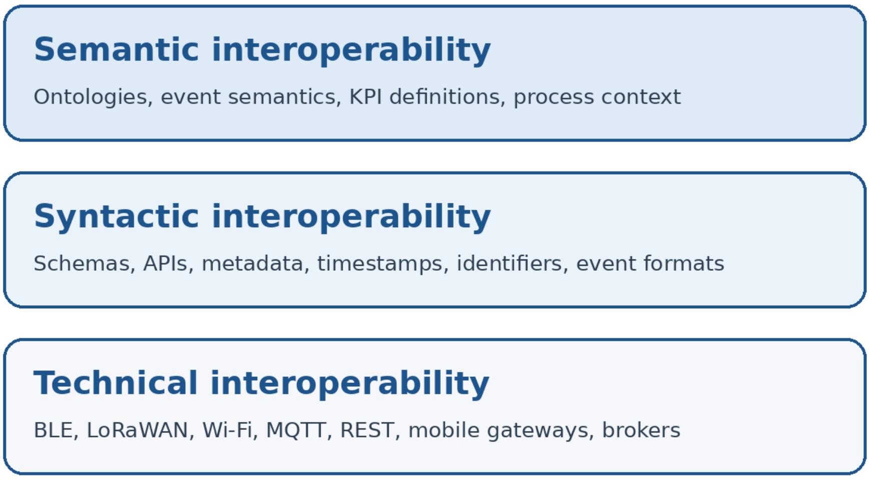

Figure 2. Interoperability stack used in A1.D3.2.

3.4 Contactless communication architectures

The phrase communication without contact in A1.D3.2 should be interpreted broadly. It refers not only to wireless data transmission, but to an architectural principle whereby devices, services, and stakeholders exchange information without requiring direct physical or organizational coupling. At the device level, BLE supports local transfer from wearables to a mobile gateway [11], [12]. At the site or territorial level, LoRaWAN supports wide-area uplink from distributed devices, including worker-safety and remote monitoring settings [13]–[15].

At the service level, contactless communication implies decoupling through APIs, brokers, event streams, and workflow orchestration. Instead of building monolithic point-to-point integrations, the architecture should expose reusable services for ingestion, event extraction, KPI generation, identity management, and model monitoring. This is consistent with the microservice and reusable-service perspective proposed for Industry 5.0 integration architectures [2].

At the stakeholder level, contactless communication implies that different actors can consume selected information without full access to all underlying systems. This is where provenance, digital identity, controlled views, and ledger-supported traceability become important. DIGEST does not need to force every actor into the same database; it needs to guarantee that the right semantic events and derived KPIs can be distributed securely and meaningfully. The use of stream-based orchestration, digital identities, and DLT-inspired distribution mechanisms is therefore not an accessory to interoperability; it is part of its trust model

3.5 Why MLOps belongs in a chapter on interoperability

At first glance, MLOps may seem like a separate issue from interoperability. In reality, they are closely linked. Once multiple heterogeneous data sources are integrated, analytical models become the mechanisms through which this integrated data produces value. If those models degrade silently, the interoperability infrastructure may remain technically functional while the decision-support layer becomes unreliable [1], [24]–[28].

The MLOps literature consistently defines operational machine learning through principles such as automation, orchestration, modularity, versioning, reproducibility, and monitoring [27]. Post-deployment monitoring is not optional but necessary to manage production risks [24]. In predictive maintenance and other continuous monitoring contexts, concept drift detection is relevant because data distributions shift over time as systems, users, and environments change [25], [26]. This matters directly for DIGEST. If A1.WP3 creates workflows that merge wearable and process data, then feature distributions seen by downstream models may evolve because of device replacement, firmware changes, worker heterogeneity, context changes, seasonality, or organizational adjustments in process execution.

Therefore, DIGEST should not treat model monitoring as a downstream afterthought. It should embed it into the architecture from the outset. Models using operator-related contextual data should expose data quality monitoring, drift monitoring, performance monitoring, version tracking, and retraining triggers [1], [24]–[28]. In this sense, MLOps is not separate from interoperability; it is the operational discipline that ensures integrated data remain analytically trustworthy.

The operational lifecycle of machine-learning models must be treated as an integral part of an interoperable architecture, particularly in environments where data evolve over time. Figure 3 illustrates the MLOps loop adapted to behavioural–process models, emphasizing the importance of tracking data drift, model performance, and data quality to maintain analytical reliability after deployment.

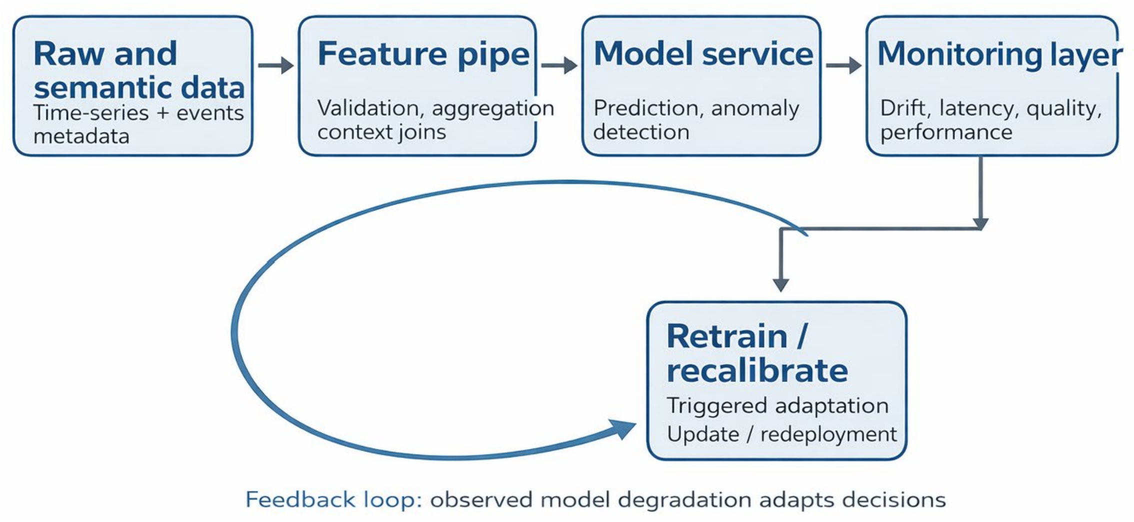

Figure 3. MLOps loop for integrated behavioural-process models

3.6 Dimensions that A1.D3.2 must address

The first required dimension is heterogeneous acquisition. DIGEST must support signals from different wearable devices, sensors, and users while handling asynchronous transmissions, variable data quality, and different sampling rates [3], [11], [12]. The second dimension is multi-user ingestion and provenance. Session identifiers, anonymized subject identifiers, timestamps, and source metadata become first-class architectural elements rather than implementation details.

A third dimension is semantic abstraction and KPI extraction. The literature suggests that KPI extraction should not occur directly on raw streams whenever interpretability and reuse are important. Event-based semantic layers provide a more robust intermediate representation [3], [16]–[18]. A fourth dimension is distributed data availability. Different actors may require shared evidence about events without receiving full access to all raw data. Distributed ledgers, digital identities, and stream-based publication mechanisms can support this selective transparency when attached to well-defined event models and provenance metadata [2], [15].

A fifth dimension is model observability. Once semantic and operational data are joined, analytical models become a living part of the infrastructure. A1.D3.2 should therefore define what is monitored, where it is logged, who can inspect it, and when corrective actions are triggered [1], [24]–[28]. The sixth dimension is privacy, anonymization, and governance. Because operator-related data may fall under sensitive data-protection regimes, interoperability must be privacy-aware by design and not privacy-added later. Identity, data minimization, event granularity, and stakeholder-specific visibility rules must be considered at the architectural level.

3.7 How DIGEST plans to address these dimensions

DIGEST is well positioned to address these dimensions because its task structure already matches a layered architecture. First, A1-T3.1 provides the acquisition layer by focusing on mobile applications compatible with heterogeneous wearable devices. A BLE-compatible gateway app, inspired by architectures such as HealthyWear, can unify body-area acquisition while abstracting away vendor-specific device details. The semantic pipeline discussed in the attached manuscript demonstrates the practicality of this pattern by using a cross-platform mobile application as a gateway between wearable inertial sensors and backend infrastructure, with raw data preserved in a time-series database and semantic events stored separately in a relational repository [3].

Second, A1-T3.2 and A1-T3.3 together define the ingestion and semanticization core. Their theoretical role is to create a pipeline where multi-user streams are normalized, transformed into semantically meaningful events, and turned into KPIs that remain interpretable and reusable across work packages. This is precisely where ontology-inspired data models and event-based semantics become most useful [16]–[20]. Third, A1-T3.4 extends the architecture into distributed availability. The proposal's mention of NiFi, Airflow, digital identity, and stream-oriented distribution mechanisms suggests that DIGEST is not aiming for a single monolithic database but for a controlled, orchestrated data ecosystem.

Fourth, the DIGEST use of advanced models in other work packages creates a natural need for MLOps-compatible observability. The industrial AI literature already argues for self-monitoring and controlled model updating in maintenance-related settings [1]. DIGEST can extend that logic so that models using operator-related contextual data are continuously checked for drift, degradation, and relevance. In summary, DIGEST plans to address A1.D3.2 not as a narrow communication problem, but as a full-stack interoperability problem involving acquisition, semantics, distribution, trust, and model lifecycle management.

Figure 4 provides an integrated view showing how the tasks defined in WP3 map onto the architectural functions previously described, clarifying the alignment between the project structure and the proposed interoperability framework.

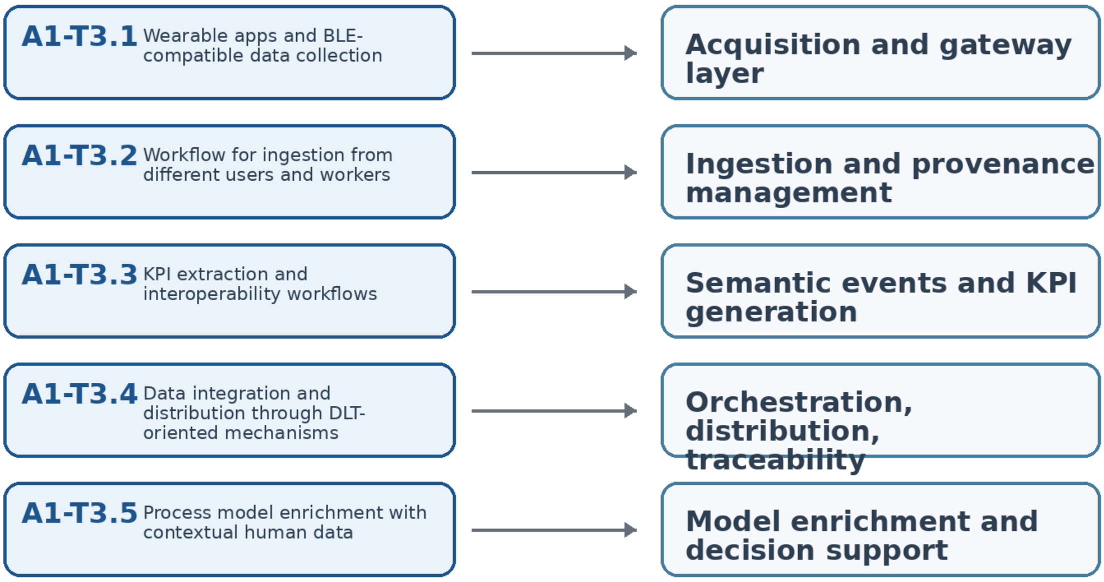

Figure 4. Mapping of DIGEST A1.WP3 tasks to architectural functions.

3.8 Concluding synthesis

From the theoretical perspective developed above, data integration, interoperability, and communication without contact should be understood in DIGEST as the coordinated design of a distributed knowledge infrastructure. The literature shows that isolated data transport is insufficient: value emerges only when heterogeneous streams can be semantically structured, securely distributed, and consumed by models whose operational behaviour is itself monitored [1]–[3], [18]–[28].

Accordingly, the contribution of DIGEST in A1.D3.2 is not limited to connecting devices. Its real contribution lies in articulating a framework where wearable apps, BLE-compatible acquisition, LoRaWAN-enabled mobility support, semantic event models, orchestration services, DLT-oriented availability, and MLOps-based model observability are treated as mutually reinforcing layers of a single interoperable architecture. This is what allows behavioural data to become trustworthy process knowledge rather than merely additional telemetry.

4 Application of the Framework to the CeDInt Use Case.

4.1 Overview of the CeDInt Living Lab

The theoretical framework presented in Chapter 3 highlights the importance of combining heterogeneous sensing infrastructures, semantic interoperability mechanisms, and operational data pipelines to support advanced analytics and decision-making. To illustrate how these principles can be implemented in practice, this chapter presents the CeDInt (Centro de Domótica Integral) building at Universidad Politécnica de Madrid as a representative use case.

CeDInt acts as a living laboratory for cyber–physical building systems, where energy systems, environmental monitoring devices, and digital infrastructures coexist and generate continuous streams of telemetry data. The building is instrumented with a heterogeneous IoT sensor network that captures electrical consumption, environmental conditions, indoor air quality, lighting state, and occupancy-related signals. These sensing streams are integrated with semantic building models derived from Building Information Modeling (BIM), enabling the creation of a semantically enriched digital twin of the facility.

One of the main motivations for this infrastructure is the need to overcome the structural mismatch between static building models and dynamic IoT telemetry streams. BIM models represent the spatial structure, topology, and functional decomposition of the building, while IoT platforms generate high-frequency time-series data describing the operational state of sensors and devices. Integrating these two worlds requires mechanisms capable of linking dynamic telemetry data with the structural knowledge embedded in BIM models.

The CeDInt architecture addresses this challenge through an Ontology-Based Data Access (OBDA) approach, which enables semantic queries over relational time-series data stored in an IoT platform. This architecture forms the foundation for advanced analytical applications such as energy forecasting, anomaly detection, and decision-support tools.

4.2 Physical sensing infrastructure

The CeDInt building is treated as a measurable cyber–physical system composed of floors, zones, and electrical subsystems. Sensor deployment follows a zone-oriented monitoring philosophy, where devices are installed according to the functional role of each space and the physical assets to be observed. This strategy ensures that the collected data preserve the physical meaning of measurements and can later be associated with specific spatial contexts.

The sensing infrastructure combines multiple families of IoT devices designed to capture complementary aspects of building operation.

4.2.1 Electrical energy monitoring

Electrical monitoring forms the backbone of the sensing infrastructure. The core devices are BM321 multi-phase energy meters, which are installed in electrical panels and used to monitor lighting circuits, HVAC systems, and other electrical loads at the circuit level. Each BM321 device captures multiple electrical metrics across several phases and circuit lines, including active power, instantaneous power, active energy, reactive energy, and current.

This multi-channel design enables fine-grained analysis of electrical behavior across different building subsystems while preserving the semantic structure of phases and circuits.

Additional devices complement the BM321 meters to provide monitoring at different levels of abstraction. For example, BMT01 meters monitor simpler electrical contexts, while BMP02 devices capture aggregated system-level parameters such as phase currents and voltages.

4.2.2 Environmental monitoring

Indoor environmental quality is monitored using BSens devices installed within occupied spaces. These sensors measure variables such as temperature, relative humidity, atmospheric pressure, and ambient luminosity. Such measurements are essential for understanding comfort conditions and for correlating environmental parameters with energy demand.

Air quality monitoring is performed through BSA02 sensors, which measure equivalent CO_2 levels, volatile organic compounds (VOCs), and derived indoor air quality indices. These signals provide valuable context for analysing the interaction between occupancy patterns, environmental conditions, and energy consumption.

4.2.3 Lighting monitoring and actuation

Lighting systems are instrumented using BSL01 smart lighting nodes, which monitor luminance levels, dimming states, and operational status of lighting fixtures. These devices also support actuation capabilities, enabling the development of closed-loop energy optimization strategies.

Overall, the heterogeneous but systematically organized sensing infrastructure provides the raw telemetry necessary to support advanced analytics within the CeDInt digital ecosystem.

4.3 Data ingestion and IoT platform architecture

The CeDInt monitoring infrastructure follows a modular layered architecture designed to decouple data acquisition, storage, and consumption processes. This architecture aligns closely with the layered digital ecosystem model discussed in Chapter 3.

The architecture can be decomposed into four main layers:

- Physical sensing layer
- Telemetry integration layer
- IoT platform and storage layer
- Semantic access layer

4.3.1 Telemetry acquisition and normalization

Sensor data are initially retrieved through a telemetry acquisition mechanism based on structured naming conventions reflecting the building hierarchy (Building → Zone → Device). This structured naming strategy simplifies the identification of devices and facilitates the integration of heterogeneous sensor streams.

A custom middleware layer normalizes the incoming telemetry data before injecting them into the IoT platform. This middleware ensures consistency in data formats and timestamps, thereby improving the robustness of the ingestion pipeline.

4.3.2 IoT platform and storage

The central IoT hub of the CeDInt infrastructure is implemented using ThingsBoard, an open-source IoT platform that manages device identities and telemetry streams. Device data are stored in a PostgreSQL database optimized for time-series ingestion.

Each sensor device is associated with a unique identifier, while telemetry observations are indexed by device ID, measurement key, and timestamp. This schema supports efficient storage and retrieval of large volumes of telemetry data generated by the sensor network.

However, relational time-series databases typically lack explicit semantic links between telemetry records and the physical entities they represent. As discussed in Chapter 3, this limitation can hinder interoperability and complicate the development of reusable analytics workflows.

4.4 Semantic interoperability through OBDA

To bridge the semantic gap between relational telemetry storage and analytical applications, the CeDInt architecture introduces a semantic access layer based on Ontology-Based Data Access (OBDA). This layer exposes the relational database as a Virtual Knowledge Graph (VKG) that can be queried using semantic concepts rather than database schemas.

The semantic model combines several widely adopted ontologies:

- Building Topology Ontology (BOT), which represents the spatial structure of the building (building, floors, rooms, zones).
- Sensor, Observation, Sample, and Actuator (SOSA) ontology, which models sensing activities and observations.
- QUDT ontology, which provides standardized representations of physical quantities and measurement units.

By linking sensor observations to spatial entities derived from the BIM model, the ontology enables context-aware queries that refer to conceptual entities rather than specific sensors.

For example, applications can request information such as:

- energy consumption of HVAC systems in a given building zone
- environmental conditions within a specific floor
- correlations between occupancy indicators and energy demand

These queries are executed through the Ontop framework, which translates SPARQL queries into optimized SQL queries executed directly on the PostgreSQL database. This approach avoids data duplication while preserving the relational database as the single source of truth.

The OBDA layer therefore enables semantic interoperability between data producers and data consumers, allowing the sensing infrastructure to evolve without breaking analytical workflows.

4.5 Energy forecasting pipeline

One of the primary analytical applications implemented within the CeDInt architecture is an energy forecasting pipeline designed to predict future energy consumption patterns using machine learning and deep learning techniques.

The forecasting system is implemented as a modular Python-based toolkit that integrates semantic data acquisition, data preprocessing, feature engineering, and model training.

4.5.1 Semantic data acquisition

Instead of relying on hard-coded database queries or device identifiers, the forecasting pipeline retrieves data through semantic queries. Engineers can request information using high-level building concepts, such as requesting energy consumption for HVAC systems in a specific building zone.

This semantic abstraction ensures that analytical workflows remain robust even when sensors are added, removed, or reconfigured.

4.5.2 Data preprocessing and feature engineering

Given the irregular sampling characteristics of IoT telemetry streams, the pipeline implements a preprocessing strategy designed to ensure signal integrity. The preprocessing stage includes:

- temporal regularization of time-series signals
- anomaly detection and sensor reset identification
- automatic feature engineering including lagged variables and calendar features

These transformations prepare the data for machine learning algorithms while preserving temporal dependencies and contextual information.

4.5.3 Machine learning and deep learning models

The forecasting toolkit supports multiple modelling approaches, including:

- Classical machine learning models such as Ridge Regression, Random Forest, and XGBoost
- Deep learning architectures such as LSTM networks, convolutional neural networks, and Transformer models

The models are trained using time-series cross-validation techniques to prevent temporal leakage and ensure that the models generalize effectively to future observations.

Experimental results demonstrate that models trained with semantically enriched contextual information outperform models trained only on electrical data. For example, incorporating environmental variables such as temperature and humidity improves prediction accuracy for HVAC energy demand.

4.6 Alignment with the DIGEST framework

The CeDInt use case illustrates how the architectural principles described in Chapter 3 can be applied to real-world monitoring environments. From a DIGEST perspective, several key insights emerge from this case study.

First, the layered architecture implemented at CeDInt demonstrates how heterogeneous sensor infrastructures can be integrated into scalable digital ecosystems through modular data pipelines.

Second, the OBDA semantic layer illustrates how ontology-driven interoperability mechanisms can decouple data producers from analytical services, enabling flexible and reusable data workflows.

Third, the forecasting pipeline highlights the importance of combining semantic data integration with advanced analytical models to support predictive decision-making.

Finally, the CeDInt living lab provides an experimental environment where the interaction between sensing infrastructure, semantic interoperability, and machine learning models can be continuously evaluated and refined.

These characteristics make the CeDInt architecture a valuable reference implementation for the broader DIGEST framework, particularly for tasks related to data integration, interoperability, and context-aware analytics.

5 REFERENCES

[1] J. Ordieres-Meré, M. Hernando, and J. Villalba-Díez, "A Data-Driven Monitoring System for a Prescriptive Maintenance Solution Based on Artificial Intelligence," Applied Sciences, vol. 15, no. 12, p. 6917, 2025. doi: 10.3390/app15126917.

[2] J. Ordieres-Meré, M. Gutiérrez, and J. Villalba-Díez, "Toward the industry 5.0 paradigm: Increasing value creation through the robust integration of humans and machines," Computers in Industry, vol. 150, p. 103947, 2023. doi: 10.1016/j.compind.2023.103947.

[3] A. Farmanesh, J. Ordieres-Meré, R. G. Sanchis, and Y. Aladro-Benito, "A semantic digital health pipeline for interpretable and longitudinal gait monitoring using wearable sensors," Digital Health, accepted manuscript, DOI pending / to be assigned.

[4] A. A. Hulleck, D. Menoth Mohan, N. Abdallah, M. El Rich, and K. Khalaf, "Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies," Frontiers in Medical Technology, vol. 4, p. 901331, 2022. doi: 10.3389/fmedt.2022.901331.

[5] E. Warmerdam et al., "Long-term unsupervised mobility assessment in movement disorders," The Lancet Neurology, vol. 19, no. 5, pp. 462–470, 2020. doi: 10.1016/S1474-4422(19)30397-7.

[6] A. Muro-de-la-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications," Sensors, vol. 14, no. 2, pp. 3362–3394, 2014. doi: 10.3390/s140203362.

[7] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition Using Wearable Sensors," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192–1209, 2013. doi: 10.1109/SURV.2012.110112.00192.

[8] H. Prasanth et al., "Wearable Sensor-Based Real-Time Gait Detection: A Systematic Review," Sensors, vol. 21, no. 8, p. 2727, 2021. doi: 10.3390/s21082727.

[9] Y. Hutabarat, D. Owaki, and M. Hayashibe, "Recent Advances in Quantitative Gait Analysis Using Wearable Sensors: A Review," IEEE Sensors Journal, vol. 21, no. 23, pp. 26470–26487, 2021. doi: 10.1109/JSEN.2021.3119658.

[10] G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. D'Orazio, "Human Gait Analysis in Neurodegenerative Diseases: A Review," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 1, pp. 229–242, 2022. doi: 10.1109/JBHI.2021.3092875.

[11] J. Tosi, F. Taffoni, M. Santacatterina, R. Sannino, and D. Formica, "Performance Evaluation of Bluetooth Low Energy: A Systematic Review," Sensors, vol. 17, no. 12, p. 2898, 2017. doi: 10.3390/s17122898.

[12] S. J. Curtis et al., “Feasibility of Bluetooth Low Energy wearable tags to quantify healthcare worker proximity networks and patient close contact: A pilot study,” Infection, Disease & Health, vol. 27, no. 2, pp. 66–70, 2022. doi: 10.1016/j.idh.2021.10.004.

[13] M. Luvisotto, F. Tramarin, L. Vangelista, and S. Vitturi, “On the Use of LoRaWAN for Indoor Industrial IoT Applications,” Wireless Communications and Mobile Computing, vol. 2018, Art. 3982646, 2018. doi: 10.1155/2018/3982646.

[14] V. Bonilla, B. Campoverde, and S. G. Yoo, “A Systematic Literature Review of LoRaWAN: Sensors and Applications,” Sensors, vol. 23, no. 20, p. 8440, 2023. doi: 10.3390/s23208440.

[15] I. Froiz-Míguez, P. Fraga-Lamas, J. Varela-Barbeito, and T. M. Fernández-Caramés, “LoRaWAN and Blockchain based Safety and Health Monitoring System for Industry 4.0 Operators,” Proceedings, vol. 42, no. 1, p. 77, 2020. doi: 10.3390/ecsa-6-06577.

[16] R. Reda, F. Piccinini, and A. Carbonaro, “Semantic Modelling of Smart Healthcare Data,” in Intelligent Systems and Applications, 2018, pp. 399–411. doi: 10.1007/978-3-030-01057-7_32.

[17] D. Spoladore, “Ontology-Based Decision Support Systems for Health Data Management to Support Collaboration in Ambient Assisted Living and Work Reintegration,” in Collaboration in a Data-Rich World, 2017, pp. 341–352. doi: 10.1007/978-3-319-65151-4_32.

[18] D. Calvanese et al., “Ontop: Answering SPARQL Queries over Relational Databases,” Semantic Web, vol. 8, no. 3, pp. 471–487, 2017. doi: 10.3233/SW-160217.

[19] E. Karabulut, S. F. Pileggi, P. Groth, and V. Degeler, “Ontologies in digital twins: A systematic literature review,” Future Generation Computer Systems, vol. 153, pp. 1–18, 2024. doi: 10.1016/j.future.2023.12.013.

[20] Y. Huang, S. Dhouib, L. Palacios Medinacelli, and J. Malenfant, “Enabling semantic interoperability of asset administration shells through an ontology-based modeling method,” in MODELS Companion, 2022, pp. 497–502. doi: 10.1145/3550356.3561606.

[21] A. Tóth, L. Nagy, R. Kennedy, B. Bohuš, J. Abonyi, and T. Ruppert, “The human-centric Industry 5.0 collaboration architecture,” MethodsX, vol. 11, p. 102260, 2023. doi: 10.1016/j.mex.2023.102260.

[22] Z. Lv, “Digital Twins in Industry 5.0,” Research, vol. 6, Art. 0071, 2023. doi: 10.34133/research.0071.

[23] J. Barata and I. Kayser, “How will the digital twin shape the future of industry 5.0?,” Technovation, vol. 134, p. 103025, 2024. doi: 10.1016/j.technovation.2024.103025.

[24] T. Schröder and M. Schulz, “Monitoring machine learning models: a categorization of challenges and methods,” Data Science and Management, vol. 5, no. 3, pp. 105–116, 2022. doi: 10.1016/j.dsm.2022.07.004.

[25] J. Zenisek, F. Holzinger, and M. Affenzeller, “Machine learning based concept drift detection for predictive maintenance,” Computers & Industrial Engineering, vol. 137, p. 106031, 2019. doi: 10.1016/j.cie.2019.106031.

[26] S. Arora, R. Rani, and N. Saxena, “A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques,” WIREs Data Mining and Knowledge Discovery, 2024. doi: 10.1002/widm.1536.

[27] D. Kreuzberger, N. Kühl, and S. Hirschl, “Machine Learning Operations (MLOps): Overview, Definition, and Architecture,” IEEE Access, vol. 11, pp. 31866–31879, 2023. doi: 10.1109/ACCESS.2023.3262138.

[28] M. I. Zarour, H. Alzabut, and K. Alsarayrah, “MLOps best practices, challenges and maturity models: A systematic literature review,” Information and Software Technology, vol. 183, p. 107733, 2025. doi: 10.1016/j.infsof.2025.107733.