



DELPHI

FeDerated nEtwork of pLatforms for Passenger and freight Intermodality

Grant Agreement Number: 101104263

D.3.2: Multimodal Data Management and Orchestration System

Document Identification			
Status	Draft	Due Date	Friday, 31 October 2025
Version	0.1	Submission Date	
Related WP	WP3	Document Reference	D.3.2
Related Deliverable(s)	D2.1, D2.4, D2.5, D3.1	Dissemination Level	PU
Lead Participant	IFAG	Document Type:	Other
Contributors	ICCS, WINGS, eBOS, NTTD	Lead Author	Kay Bierzynski, IFAG
		Reviewers	NTTD Kyriaki Psara, eBOS



DELPHI project has received funding under grant agreement No 101104263. It is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Document Information

Author(s) – in alphabetical order		
First Name	Last Name	Partner
Sokratis	Barpounakis	WINGS
Kay	Bierzynski	IFAG
Xianyuang	Peng	IFAG
Kyriaki	Psara	eBOS
Georgel	Saveanu	NTTD

Document History			
Version	Date	Modified by	Modification reason
0.1	14/07/2025	Kay Bierzynski	Initialize Documents
0.2	20/08/2025	Georgel Saveanu	Added NTTD contribution
0.3	24/09/2025	Xiangyuan Peng, Kyriaki Psara	Added IFAG and eBOS contribution to section 2
0.4	15/10/2025	Sokratis Barpounakis	Added WINGS contributions
0.5	11/11/2025	Kay Bierzynski	Introduction, conclusion and chapter 3 added
0.6	12/11/2025	Kay Bierzynski	Formatting
0.7	13/11/2025	Kyriaki Psara, Kay Bierzynski	Fix section titles, improve images and contributions
1.0			Final version to be submitted

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Kay Bierzynski	dd/mm/yyyy
Quality manager		
Project Coordinator		

Executive Summary

Deliverable 3.2, “Multimodal Data Management and Orchestration System” documents the results of task 3.2 “Data generation, exposure, acquisition, ingestion and processing for diverse multi-modal passenger and freight transport systems” of WP3. Specifically, it provides an overview of the exploration, design and implementation work done by the partners to realize the necessary data ingestion, management and processing components for the use case systems as well as the overall Delphi concept.

In detail D3.2 includes the following topics:

- IFAG presents their exploration for extending the Delphi system towards the edge of transportation and logistics applications and describe a demonstrator that shows how the extension looks like
- ICCS
- WINGS show a data ingestion and management system for the Athen pilot use case based on the local DELPHI platform instantiation
- EBOS show the results of their literature-research regarding requirements, standards and general aspects to consider for the implementation of a data management system in logistics and transportation
- NTTD presents data ingestion and management element that they implemented based on Apache Airflow for the Cluj-Napoca pilot use case

Further, this deliverable discusses trends and challenges that data ingestion, orchestration and processing layers of transportation and logistic systems need to consider today and in the future. Some of these trends are for example autonomous vehicles, humanoid robots, AI agents, adversarial attacks and city-wide sensing infrastructures generating high amounts of data. Hence, besides the project intern developers, this deliverable addresses researchers, businesses and engineers that want to realize the next generation of traffic, transportation and logistics systems.

Table of Contents

Executive Summary.....	3
1. Introduction	7
1.1 Purpose of the document	7
1.2 Intended readership.....	7
1.3 Document Structure.....	7
2. Overview of the Partner Results.....	8
2.1 Edge-Cloud Data Ingestion and Processing System (IFAG)	8
2.1.1 General Overview of Data Ingestion System.....	8
2.1.2 System Architecture	9
2.1.3 Evaluation Metrics.....	11
2.1.4 Achievements.....	11
2.1.5 Impacts	12
2.2 ICCS	12
2.3 Athen Pilot Data Ingestion and Processing System (WINGS).....	13
2.3.1 General overview	13
2.3.2 System architecture	13
2.3.2.1 Data sources integrated	13
2.3.2.2 Cloud-based ingestion and processing.....	15
2.3.2.3 From data to decision support	15
2.3.3 Evaluation metrics.....	16
2.3.4 Achievements.....	16
2.3.5 Impact.....	17
2.4 Unified Data Framework and Technical Specifications (eBOS).....	18
2.5 Cluj-Napoca Pilot Data Ingestion and Processing (NTTD)	23
2.5.1 Workflow Orchestration and Automation	23
2.5.2 Support for Multiple Data Formats	24
2.5.3 Centralized Storage and Access.....	24
3. Consideration for an generalized Delphi Data Ingestion and Orchestration System	25
3.1 Trends and Challenges	25



3.2 Realisation of a generalized solution for Data Ingestion, Orchestration and Processing Layers of DELPHI System..... 27

4. Conclusions 28

References..... 29

List of Tables

No table of figures entries found.

List of Figures

Figure 1 The main structure of data ingestion system..... 9

Figure 2 WINGS wi.MOVE platform that acts as the main platform of the Athens pilot..... 13

Figure 3 Primary data sources..... **Error! Bookmark not defined.**

Figure 4 Data Handling Structure 23

Figure 5 Transfer Execution using DAG 24

Abbreviations & Acronyms

Abbreviation / acronym	Description
AI	Artificial Intelligence
API	Application Programming Interface
D3.2	Deliverable number 2 belonging to WP 3
DAG	Directed Acyclic Graph
DCAT	Data Catalog Vocabulary
eBOS	EBOS TECHNOLOGIES LIMITED
EC	European Commission
ML	Machine Learning
IoU	Intersection over Union
EC	European Commission
ERD	Entity Relationship Diagram
FIBO	Financial Industry Business Ontology

Abbreviation / acronym	Description
GDPR	General Data Protection Regulation
GTFS	General Transit Feed Specification
GUI	Graphical User Interface
HDFS	Hadoop Distributed File System
ICCS	EREVNITIKO PANEPISTIMIAKO INSTITOUTO SYSTIMATON EPIKOINONION KAI YPOLGISTON-EM
IFAG	INFINEON TECHNOLOGIES AG
MCU	Microcontroller
MOS	Motion Object Segmentation
MQTT	Message Queuing Telemetry Transport
NoP	Network of Platforms
NTTD	NTT DATA ROMANIA SA
ODRL	Open Digital Rights Language
SSN/SOSA	Semantic Sensor Network Ontology / Sensor, Observation, Sample, and Actuator
TFLM	TensorFlow Lite for Microcontrollers
TLS	Transport Layer Security
TR	Technical Requirements
UAV	Unmanned Aerial Vehicle
VDS	Vehicle Detection System
VSSo	Vehicle Signal and State Ontology
WINGS	WINGS ICT SOLUTIONS INFORMATION & COMMUNICATION TECHNOLOGIES IKE
WP	Work Package

1. Introduction

1.1 Purpose of the document

The purpose of Deliverable D3.2, “Multimodal Data Management and Orchestration System”, is to document the results of task 3.2 of Work Package (WP) 3. The goal of task 3.2 “Data generation, exposure, acquisition, ingestion and processing for diverse multi-modal passenger and freight transport systems” is to explore, develop and deliver solutions for an integrated end-to-end, big data ingestion and processing system, capable of receiving, storing, as well processing in real-time large amounts of data from heterogeneous sources and systems across transportation and logistics network for the Delphi use cases. This deliverable contributes to building the basis for work of WP4 and WP 5 by providing an overview of the data ingestion and processing exploration and implemented components of the partners IFAG, ICCS, WINGS, EBOS and NTTD. Furthermore, it discusses trends and challenges that need to be considered for data ingestion and processing for transportation and logistic systems.

1.2 Intended readership

The intended audience for D3.2 are the technical teams of WP4 and WP5, who are responsible for implementing the use cases and AI modules of the Delphi system. Understanding the data ingestion layers and the overarching architecture of the processing system is crucial for the partners to implement the input layers of their AI correctly and to have cohesive structure for the use case systems. Furthermore, external audiences, including researchers, policymakers, and businesses in the fields of transportation and logistics will benefit from reading this deliverable as well to learn from the lessons learned and the challenges that data ingestion and processing systems face today and the future.

1.3 Document Structure

The following structure provides an overview of the partner specific work and general discussion of challenges in the field of data ingestion and processing for transportation and logistics systems:

The document starts with Section 1 which introduces the purpose, audience, and structure of the document.

Following, Section 2 separates itself into multiple sub sections for the results of IFAG, EBOS, ICCS, WINGS and NTTD.

Section 3 discusses the trends and challenges that make implementation of a general solution for data ingestion and processing systems for transportation and logistics difficult.

Section 4 concludes the document, summarising the key contributions of D3.2.

2. Overview of the Partner Results

2.1 Edge-Cloud Data Ingestion and Processing System (IFAG)

A fully operational edge–cloud data ingestion system designed for multi-modal transport environments was developed by IFAG to explore the topic of edge-cloud continuum data ingestion and processing. The system supports local preprocessing, edge inference, gateway-level fusion, and cloud monitoring of heterogeneous data streams. By distributing computation between devices and the cloud, it provides timely responsiveness, reduces communication demands, and enables scalable cloud-based monitoring. This result of task 3.2 supports advanced transport monitoring, fleet management, and intelligent mobility services.

2.1.1 General Overview of Data Ingestion System

Data ingestion represents the foundational process of systematically collecting, importing, and preparing data for downstream processing, while data pipelining ensures that the data flows continuously through various transformation and integration stages to support timely analysis [3].

With the exponential growth of data volume, velocity, and variety, large-scale distributed platforms such as Hadoop have played a pivotal role by providing fault-tolerant storage through Hadoop Distributed File System (HDFS) and parallel computation via MapReduce [4]. Complementary technologies such as Apache Kafka and Apache Flume [5] have become central in enabling both batch and stream ingestion. Despite these advancements, challenges remain, particularly with the increasing demand for low-latency processing, scalability, and the ability to handle heterogeneous multi-modal data [6]. Traditional data integration tools, which are typically on-premises and GUI-driven, remain optimized for batch-oriented workflows, with limited support for real-time ingestion and unstructured data. In contrast, stream ingestion platforms are specifically designed for continuous processing, supporting structured, semi-structured, and unstructured formats while enabling real-time data pipelining [7]. At the same time, cloud-native ingestion and processing frameworks provide elastic scalability and hybrid deployment options across both batch and stream workloads, though their extensibility often remains constrained by the predefined platform features.

Recent developments in the field place increasing emphasis on cloud-native architectures, edge intelligence, and AI-enhanced pipelines. Cloud services deliver on-demand scalability and resource elasticity, while edge computing reduces communication overhead and enables responsive decision-making close to data sources. Furthermore, machine learning and AI are being integrated into ingestion workflows, enabling adaptive pipeline orchestration, anomaly detection, and

predictive analytics, thereby improving adaptability and operational intelligence [8]. Both research and industrial solutions highlight the growing importance of hybrid edge–cloud paradigms [9], where preprocessing, lightweight inference, and local fusion at the edge are tightly integrated with cloud-based storage, analytics, and monitoring systems.

In summary, modern data ingestion systems are characterized by the convergence of distributed processing frameworks, real-time ingestion platforms, cloud-native scalability, and AI-driven intelligence. These technologies are reshaping the way organizations manage and exploit big data, enabling greater agility, resilience, and operational efficiency across domains such as transportation, healthcare, and industrial IoT.

2.1.2 System Architecture

The system follows a sensor–edge–gateway–cloud hierarchy, as shown in Figure 1:

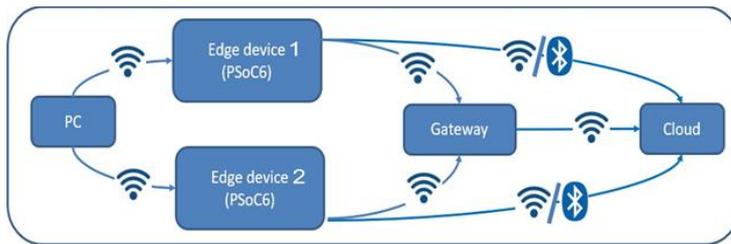


Figure 1 The main structure of data ingestion system.

The data flow follows the structure shown in Figure 2. Radar point clouds after sampling are first analyzed on edge devices and then fused with camera images in the gateway for final segmentation results. Both the output from the edge devices and the gateway will be updated and stored in the cloud server for global monitoring.

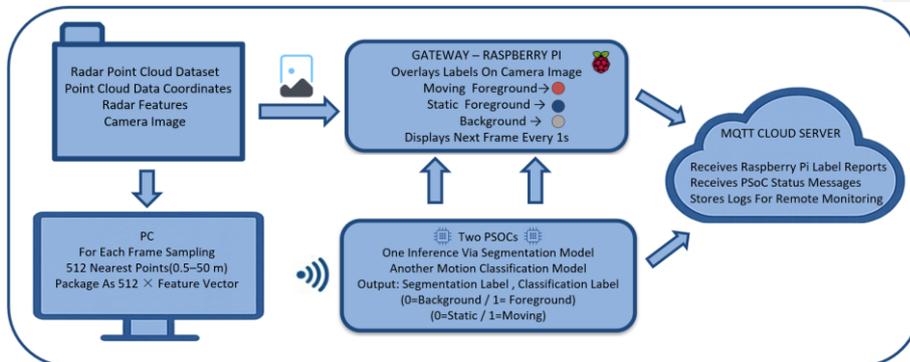


Figure 2 The data flow of the designed ingestion system.

The detailed components of our ingestion system are:

1. Data Preprocessing on PC serving as a simulated data source or sensor
 - a. Radar point clouds and camera images are preprocessed and downsampled on the PC. The preprocessing pipeline is implemented in Python, making use of libraries including Nand camera images are preprocessed and downsampled operations, data resizing, and noise reduction.
 - b. This preprocessing step reduces the data volume, lowers bandwidth requirements, and ensures compatibility with downstream edge devices before Wi-Fi transmission via socket-based communication.
2. Edge Inference on Embedded Devices
 - a. Two PSoC6 microcontrollers serve as edge devices, each deployed with lightweight machine learning models compiled into C/C++ code. Inference is executed using TensorFlow Lite for Microcontrollers (TFLM), which provides optimized kernels for devices with strict RAM and Flash constraints.
 - b. Edge device 1 runs a Motion Object Segmentation (MOS) model, distinguishing moving radar points from static ones.
 - c. Edge device 2 performs foreground-background segmentation, isolating relevant elements in the radar data.
 - d. The models are quantized to 8-bit integer precision (int8) to minimize memory footprint and are executed through CMSIS-NN kernels for accelerated low-level computation. This modular division allows parallelized inference under limited compute and memory resources.
3. Gateway-Level Fusion and Visualization
 - a. A Raspberry Pi 4 functions as the system gateway. Results from the two PSoC6 devices are transmitted via Wi-Fi and aggregated at the gateway.
 - b. Data fusion logic, implemented in Python, merges two segmentation results with camera imagery. For real-time visualization, Matplotlib and OpenCV are applied to overlay radar-derived object segmentation on live video feeds.
4. Cloud Ingestion and Monitoring
 - a. The edge device status messages (e.g., “receiving,” “sending”) and the gateway fusion results are published to the cloud layer over Wi-Fi under the MQTT protocol (Message Queuing Telemetry Transport), facilitated by the paho-mqtt Python library.
 - b. In the cloud environment, a Python-based MQTT broker (Eclipse Mosquitto) handles message routing.

Commented [KB1]: Consider creating a separated image for data flow/processing/ingestion steps.

Commented [PA2R1]: Add data flow chart (Figure 2)

Commented [KB3]: Should be marked in figure 1

Commented [PA4R3]: marked

Commented [KB5]: Should be marked in figure 1

Commented [PA6R5]: marked

- c. Persistent data storage is supported by Influx DB while data security is maintained through TLS certificates and device authentication at the MQTT layer.

2.1.3 Evaluation Metrics

To assess the effectiveness and robustness of the edge–cloud data ingestion system, a set of evaluation metrics is defined across different system layers:

- Ingestion performance: End-to-end latency (p50/p95), bandwidth usage, packet loss rate.
- Edge inference: Inference time per frame, peak RAM usage, model accuracy under downsampling.
- Fusion quality: Stability of moving/static object detection, IoU/F1 for segmentation performance.
- Cloud observability: MQTT connection uptime, reconnection count, security certificate validity.

2.1.4 Achievements

The data ingestion system demonstrates how distributed processing enhances efficiency, responsiveness, and scalability in transport-related applications. Key results include:

- End-to-End Data Ingestion Pipeline: Established a seamless flow from raw data to cloud storage, with standardized data interfaces.
- Real-Time Edge Intelligence: Deployed two lightweight ML-models on PSoC6 microcontrollers (MCUs), achieving low-latency processing under constrained resources.
- Decision-Level Fusion: Demonstrated integration of dual edge results at the Raspberry Pi for high-level information fusion.
- Visualization and Monitoring: Provided fused outputs aligned with camera imagery and cloud-based dashboards for remote oversight.
- Edge and Cloud Ingestion Cooperation:
 - 1) Edge ingestion: Inference on edge devices enables low-latency decisions, reduced bandwidth, and operation of independence. But it restricts the memory and computing capability of microcontrollers.
 - 2) Cloud ingestion: Cloud server offers centralized monitoring and long-term data storage, but requires network availability and higher cloud resources.
 - 3) Our system uses edge ingestion for real-time data segmentation, while cloud ingestion keeps lightweight logs of only compact output, providing both fast local intelligence and scalable global visibility.

Commented [KB7]: Here or in 2.1.5 some sentences about differences between cloud and edge ingestion would be good.

Commented [PA8R7]: add "Edge and Cloud Ingestion Cooperation" in achievement

2.1.5 Impacts

This system can enhance modern delivery and transportation operations by combining the responsiveness of edge computing with the strategic oversight of cloud analytics. It supports real-time adaptation to changing conditions while enabling long-term planning and coordination. The integrated approach improves efficiency, scalability, and reliability across data operations, contributing to smoother and more resilient mobility services.

2.2 ICCS

Text

ToDo: Describe the results of your work of Task 3.2

2.3 Athen Pilot Data Ingestion and Processing System (WINGS)

2.3.1 General overview

WINGS implements a data ingestion and integration framework based on the local DELPHI platform instantiation, leveraging its wi.MOVE platform (see Figure 3Figure 1). This contribution is directly applied in the Athens pilot use case, which focuses on improving multimodal transport efficiency by combining road traffic data with urban rail data. The objective is to enable better management of congestion on Attiki Odos and support decisions on freight offloading to metro services when appropriate. To achieve this, WINGS instantiates the DELPHI architecture leveraging wi.MOVE's capabilities in interoperability, data collection from heterogeneous sources, AI/analytics, and operational dashboards.

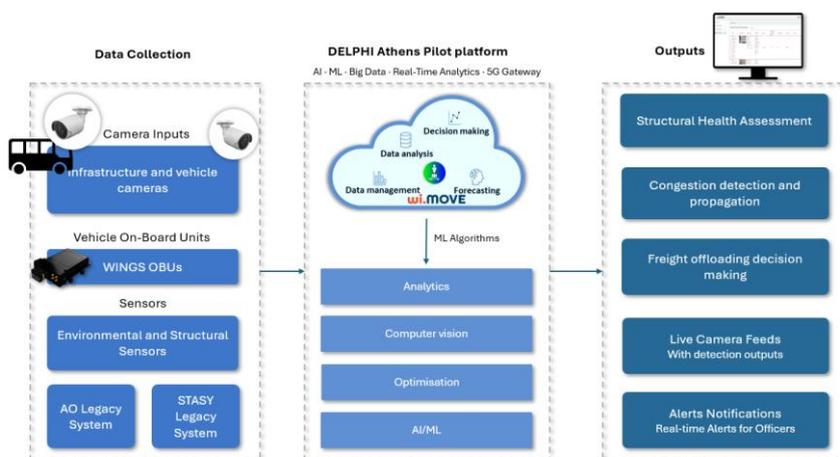


Figure 3 WINGS wi.MOVE platform that acts as the main platform of the Athens pilot

The Athens pilot architecture reflects the varying nature of the data sources handled by WINGS: on the one hand, sensor-based inputs (structural, positional, environmental) and on the other, legacy systems feeds (Attiki Odos traffic management data, STASY schedules and ticketing data, and third-party service information). These diverse streams are channeled through wi.MOVE into analytics, computer vision, and AI/ML modules, producing insights that drive real-time decision support for multiple actors.

2.3.2 System architecture

2.3.2.1 Data sources integrated

WINGS developed a wi.MOVE-based ingestion layer that connects legacy traffic-management and public-transport systems to the DELPHI Network of Platforms.

The Athens pilot combines physical sensor data and legacy systems.

- On the sensor side, wi.MOVE can ingest structural sensors (e.g., infrastructure condition), position sensors (e.g., location and movement of vehicles or cargo), and environmental sensors (e.g., air quality, weather conditions). These enrich the transport picture with situational awareness signals.
- On the legacy systems side, the focus is on Attiki Odos traffic management feeds (loop-detector/ Vehicle Detection System (VDS) data that capture flow, speed, occupancy) and STASY metro data (timetables, schedules, and aggregated ticketing counts per station). Additional inputs from third-party service providers may complement these core datasets, extending the multimodal view.

The Athens pilot showcases a representative mix of sensor data and legacy system interfaces, both of which must be ingested and aligned before they can support decision-making.

- Sensor data.
 - *Structural sensors* provide insights into the state of transport infrastructure (e.g., bridges, road sections), which can influence routing and safety considerations.
 - *Position sensors* track the movement of vehicles and goods, adding a layer of spatial awareness to the system.
 - *Environmental sensors* capture external conditions such as air quality, weather, or temperature. While not always directly linked to congestion, these indicators are valuable for understanding the wider context of freight and passenger operations.
- Legacy systems.
 - *Attiki Odos traffic management system*: The main source of road traffic data, delivered through loop detectors/VDS. These provide metrics such as traffic flow, speed, and occupancy, which are critical for congestion detection and prediction.
 - *STASY metro system*: Data from the Athens metro network, including timetables/schedules (often GTFS-based) and aggregated ticketing information at the station level. This allows the system to estimate capacity and demand on metro lines and identify opportunities for transferring freight from road to rail.

- *Third-party service providers*: Additional feeds (e.g., weather service APIs) may also be incorporated.

2.3.2.2 Cloud-based ingestion and processing

The ingestion of heterogeneous datasets is achieved through wi.MOVE's integration capabilities, which have been designed to handle real-time updates, API-based interoperability, and visual dashboards for monitoring. The platform provides connectors for both structured sensor outputs and legacy data interfaces, ensuring that the information enters the system in a consistent and usable format.

Once ingested, the data undergoes preprocessing, alignment, and quality checks: timestamps are synchronised, location identifiers are harmonised, and basic validation rules are applied to ensure plausibility. This step is essential for enabling cross-domain analytics, where road congestion levels can be meaningfully correlated with metro availability or demand.

From there, three complementary processing approaches are used:

- Analytics modules provide statistics, trend analysis, and historical comparisons. For example, traffic volumes are aggregated per segment, and demand levels are tracked across metro stations.
- Computer vision can be applied where video feeds are available (e.g., from roadside cameras) to enrich traffic information with visual confirmation of congestion, lane usage, or incidents.
- AI/ML models are used to detect patterns, predict congestion propagation, and identify abnormal conditions (such as sudden drops in metro demand or unusual traffic spikes).

The combination of these techniques ensures that both routine monitoring and predictive insights are available in real time.

2.3.2.3 From data to decision support

The ultimate purpose of the ingestion and processing chain is to support decision-making for the key actors. Indicative insights are generated and presented in actionable formats:

- For drivers and freight forwarders, the system will consider recommendations on when to exit the highway earlier to avoid congestion, or whether to transfer freight to the metro network. These decisions consider traffic propagation patterns on Attiki Odos, metro schedules, and station demand levels.

- For infrastructure operators (Attiki Odos), the dashboards will comprise live congestion conditions, predictions of traffic build-up, and the impact of re-routing decisions.
- For freight forwarders and public transport operators (STASY), the system will explore highlighting freight offloading opportunities without disrupting passenger flows, as well as abnormal demand patterns at specific stations.

The decision-support outputs are not limited to end-user recommendations. They also provide a basis for policy and operational planning, enabling stakeholders to explore how multimodal coordination can alleviate congestion in real-world conditions.

2.3.3 Evaluation metrics

To assess the effectiveness of the platform-based ingestion and analytics framework in the Athens pilot, WINGS will apply a set of quantitative and qualitative indicators:

- Data quality and coverage: percentage of records with complete attributes; alignment between traffic counts and metro ticketing data; detection of missing or inconsistent feeds.
- Analytics/AI performance: accuracy of congestion predictions on Attiki Odos (e.g., mean absolute error of travel time forecasts); correlation between predicted vs. observed metro demand levels.
- Decision-support value: number of actionable recommendations produced (e.g., freight offloading suggestions); operator feedback on usability of dashboards; reduction in manual interventions.

2.3.4 Achievements

The Athens pilot has already achieved key milestones thanks to WINGS contribution:

- Seamless ingestion from legacy systems: APIs for Attiki Odos traffic data and STASY schedules/ticketing have been integrated into wi.MOVE, ensuring continuous data availability.
- Sensor data integration: Structural, positional, and environmental sensors have been connected, extending the pilot's situational awareness.
- Operational dashboards: Custom dashboards provide real-time visibility of road congestion, metro capacity, and freight offloading opportunities.
- Analytics and AI pipelines: Initial models for congestion detection and metro demand estimation have been deployed, supporting multimodal optimisation scenarios.

2.3.5 Impact

The WINGS contribution through wi.MOVE is expected to generate both technical and operational impacts:

- Technical impact: Establishes a standardized ingestion layer that can be reused across other DELPHI pilots; demonstrates secure, API-based integration of heterogeneous sources; validates AI-driven analytics on multimodal datasets.
- Operational impact: Enables more efficient freight management through metro offloading; supports congestion mitigation on Attiki Odos; improves situational awareness for operators; and lays the groundwork for replicable multimodal services in other European cities.

2.4 Unified Data Framework and Technical Specifications (eBOS)

T3.1 developed a unified data framework and technical specifications needed to support the DELPHI project goals of bringing together people and freight transport within a unified, secure, and efficient multimodal transport network. This is a critical step, particularly in providing the technical and conceptual building blocks needed to govern and manage large amounts and diverse data in real-time. Key areas of work included:

Harmonised Data Specification

A significant area of work was in reviewing and identifying relevant data standards and ontologies to enable data sharing across multimodal transport and logistics. The project identified important standards such as DATEX II for traffic and travel information, GTFS (General Transit Feed Specification) for public transportation data, and Transmodel (EN 12896) as the European reference data model for public transport. Standards for information security management (ISO/IEC 27001) and privacy (ISO/IEC 29100) were also identified to ensure secure processing and handling of sensitive traffic and personal data. The objective of this review was to make sure that data assets are managed consistently and adhere to user requirements and data governance needs. Key Data Standards identified include:

- DATEX II
- GTFS (General Transit Feed Specification)
- Transmodel (EN 12896)
- ISO/IEC 27001
- ISO/IEC 29100
- IEEE 802.11p
- ISO 20022
- ISO 6709
- ISO 22320

Various ontologies were examined to provide common understanding and formalise relationships between pieces of diverse data. Examples of this work include using SSN/SOSA (Semantic Sensor Network Ontology / Sensor, Observation, Sample and Actuator) to formalise and enable integration of sensor data from traffic-monitoring systems; the Data Privacy Vocabulary to formalise and account for privacy and access control; VSSo (Vehicle Signal and State Ontology) to help agent configure telemetry and real-time states of vehicles; and Transport Disruption Ontology was identified as a candidate for formalising and standardising data on events related to disruption in traffic and other anomalies data. The ability to formalise data in this the semantic way is important for dealing with large amounts of data, and meaningfully engaging and

- TR43: Calibration of the drone data models: Specific to novel data sources like UAVs.

Methodological Framework for Data Sharing



Figure 6 Objectives of the Methodological Framework

Alongside the harmonised data specification, the project also develops a comprehensive methodological framework allowing secure and efficient data sharing/storage/usage. This framework aims at the DELPHI federated environment having operational data governance incorporating data sharing complexities and control practices learned across both industries and academia. It also aims at ensuring stakeholders have full control over their data during all processing phases.

Key aspects of this framework include:

- Functions to have Data Security incorporate principles of confidentiality, and data protection including GDPR and ISO/IEC 29100. Guidelines on encryption, access controls, secure APIs, and other mechanisms of preventing unauthorised access and data alteration are provisions of data security.
- Functions to have Data Safety align with the Data Governance Act, Data Act, and other legally binding EU privacy instruments through provisions of data validation, and mechanisms of ensuring data integrity, and privacy protection.

D.3.2: Multimodal Data Management and Orchestration System



- Functions to Efficient Data Use align with the principles of optimal data transfer protocols such as HTTP/2 for edge data caching, and other data use standardisation for achieving efficient coordination of data dissemination across traffic participants.
- Functions implementing Neutral Governance Structure develop comprehensive methods of effective data dissemination across traffic participants and ensure operational data governance across all data sharing stakes.
- The framework offers a model for a neutral governance structure with respect to governance function, governance entities, governance legal, and governance procedures for trust, safety, and security for accessing and sharing data.

This framework, as referred to in D3.1 (with D2.5 being a counterpart deliverable), is critical in making certain that the data management and orchestration system of D3.2 aligns with the European values and principles, thus providing the foundational trust for sharing data and the effective implications of abundant data for further innovative applications.

2.5 Cluj-Napoca Pilot Data Ingestion and Processing (NTTD)

As part of task 3.2, the work of NTTD focuses on enabling a flexible and scalable data handling pipeline capable of supporting diverse transport-related data flows in the context of Cluj-Napoca pilot use case, as illustrated in Figure 7.

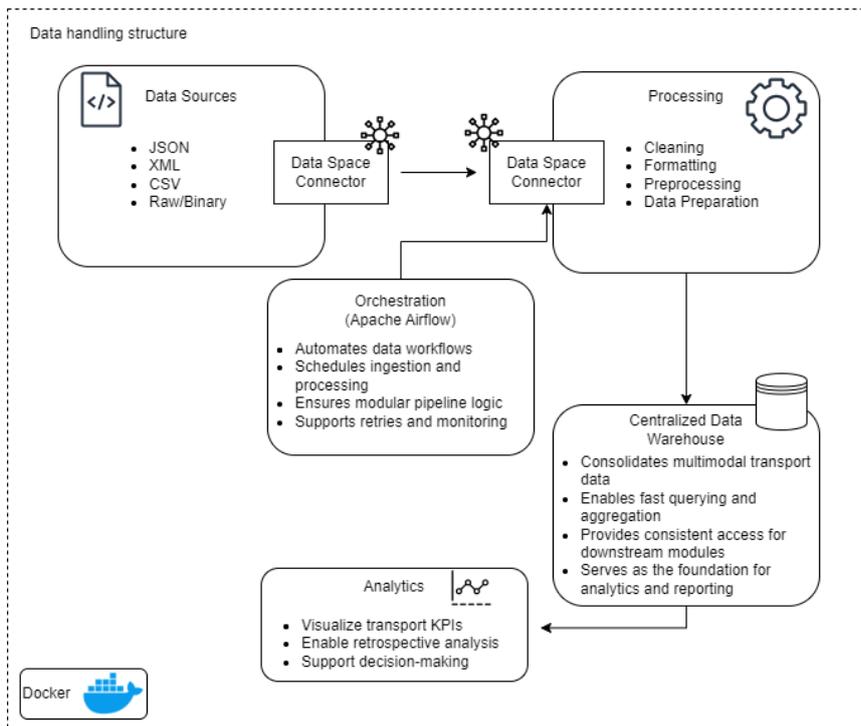


Figure 7 Data Handling Structure

The main components and outcomes are summarized below:

2.5.1 Workflow Orchestration and Automation

An orchestration layer based on Apache Airflow has been implemented to enable the definition and execution of data workflows. This system automates the steps involved in data transfer, processing, and loading, ensuring modularity and reusability of processing logic. The orchestration is designed to support integration with both data space connectors and conventional data sources or sinks.

- Airflow DAGs control the end-to-end flow from data acquisition to processing and storage, as presented in Figure 8.
- Scripts and operators are used to handle transformation and validation tasks within each pipeline.

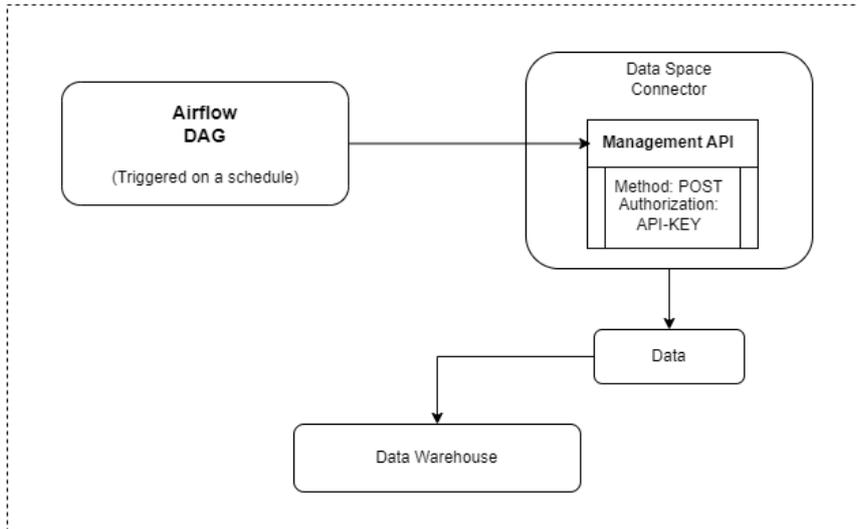


Figure 8 Transfer Execution using DAG

2.5.2 [Support for Multiple Data Formats](#)

The developed solution is agnostic to data types and formats, with the ability to ingest and process structured and semi-structured data such as:

- JSON
- XML
- CSV
- Raw files and binary formats

This allows integration of heterogeneous data sources across passenger and freight transport systems, supporting future extensibility without architectural changes.

2.5.3 [Centralized Storage and Access](#)

Ingested and processed data is stored in a centralized data warehouse designed to support analytics and reporting use cases. This layer:

- Consolidates multimodal transport data in a structured manner
- Enables querying, aggregation, and visualization downstream
- Acts as a foundation for future WP5 work related to use case demonstration and dashboarding

3. Consideration for an generalized Delphi Data Ingestion and Orchestration System

In this section 3 trends and challenges and their impact on data ingestion and processing are discussed. Afterwards, a suggestion is provided on how provide a flexible solution that could be offered to wide variety of customers.

3.1 Trends and Challenges

In the following we list trends and challenges that have or are expected to have significant impact on data ingestion, orchestration and processing of transportation and logistic systems.

Legacy system, fractured data and varying degree of digitalisation: One of the biggest challenges for data ingestion and processing is that systems of public institution for traffic and transportation and logistic companies differ greatly in interfaces, data formats, security standards, communication protocols and data quality. Additionally, it may very well be that parts of traffic and warehouse infrastructure are not even digitalised making it impossible to consider them in the planning of the overall control system.

Growing sensing capabilities of the infrastructure: Due to trends such as smart cities, smart containers or autonomous vehicles, more and more sensors are installed all over the traffic, transportation and logistic networks. This makes it possible to have real-time information about the situations of all system nodes. However, at the same time the data streams transferred are increasing proportionally leading to increase of cost and complexity of data ingestion and processing layers.

Integration of AI everywhere: To deal with the previously mentioned increase of data streams as well as to enable the realization of many different applications like automatization of traffic control AI is implemented on all levels of the traffic, transportation and logistic systems. However, AI differs in many aspects from traditional software. For example, it does not have the same level of explainability as traditional software which makes it more difficult to certify, trust and secure. Furthermore, AI opens new attack vectors like adversarial attacks where input data is manipulated in specific ways to generate malfunctions of AI models. These facts and other aspects of AI need to be carefully considered in the design of the architecture and security of data ingestion and processing layers. Especially when it is intend to integrate AI into these layers.

The growth of edge computing and increase of decentralization: The ever-increasing amount of data generated by sensors, devices and nodes in transportation, traffic and logistics systems are bringing the network infrastructure closer and closer to its limits. This and requirements of real-time data of many applications like autonomous vehicles

lead to the necessity of moving more and more data processing closer to the data sources and with it to the edges of the transportation, traffic and logistic networks. These edge computation nodes are limited in their resources, safety and security capabilities compared to the cloud. This makes them more untrustworthy which needs to be considered in the specification of the data ingestion and orchestration. Furthermore, they change the character of data transferred in the system from raw data to more precise information.

Autonomous and Digitalized Vehicles and Vehicle to Everything: The automatization of cars, busses, trucks and other vehicles lead to new applications and requirements for data ingestion and processing of traffic, transportation and logistic systems especially for the input and output prioritization. For example, would it be possible for the core of a traffic system to send command to vehicles for the autonomous creation of rescue lanes. The other way is also possible where electrical vehicle would send warning to the overall traffic system when its battery fails and starts to burn.

Another aspect of this trend is the vehicle to everything where vehicles communicate with each other and other local nodes such as traffic lights of the traffic, transportation and logistic networks. This is important to consider when designing data ingestion, orchestration and communication, since it enables the possibility of having specific data gathering and command sending nodes instead of communicating with every vehicle separately as well as to cluster and control multiple vehicles as a swarm.

Blockchain and digital product passes: For financial and insurance applications as well as product transparency, blockchain and similar technologies became an avenue that is explored to realize and digitalize these application. This also impacts transportation and logistics in form of insurance and the concept of digital product passes that for example need to include information about the cooling chain. Hence, when data ingestion and processing layers are implemented, they should be made compatible with the blockchain.

Post-Quantum-Cryptography (PQC) refers to algorithms that are secure against classical and quantum computers. This means that their encryption methods cannot be broken by quantum computation like it is the case for many state-of-the-art algorithms. Hence, PQC algorithms should be considered for data ingestion and processing layers to secure them for the future.

AI Agents: The capabilities of large language model and other generative AI methods are steadily increasing every month. Hence, it can be assumed that in near future AI agents based on these methods will become capable of executing a wide range of tasks. This entails that data ingestion and processing layers will interact more and more with AI agents in place of traditional systems or humans. As AI agents will behave quite differently for example in regards to response time the layers need to be adapted accordingly. Furthermore, integration of agents into the data ingestion and processing

layers themselves should be considered to deal with data management tasks more efficiently.

(Humanoid) Robots and Drones: In the overall automatization of transportation and logistic systems robots already playing an important role in warehouses and drones are closely investigated to solve the last mile problem of package delivery. Hence, data ingestion, orchestration and processing need to be able to deal with incoming data from them as well as outgoing commands to the drones or robots. Additionally, it is very likely that in one to two decades humanoid robots will be deployed widely in society. It can be assumed that these robots will then use transportation networks in a similar way that humans do today to do different kinds of tasks like visiting a house for electrician work. However, these robots will be interacting differently with transportation and traffic systems than humans do, since for example they can communicate directly without additional steps like humans need to take via smartphones. This should be considered when creating data ingestion, orchestration and processing layers to make them ready for these developments.

The list of topics in this section should not be seen as complete, since more points can be added to this list of considerations that impact data ingestion, orchestration and processing. However, this list provides a good starting point.

3.2 Realisation of a generalized solution for Data Ingestion, Orchestration and Processing Layers of DELPHI System

As the list of topics to consider for the design of data ingestion, orchestration and processing layers in section 3.1 shows, it is very difficult and costly to develop a solution that fits all traffic, transportation and logistic systems and applications. Hence, the partners focused in their development on the specific contexts of the pilot use cases and edge data ingestion in case of IFAG.

The partners suggest that for the development of a more generalized solution generative AI and AI agent solutions should be explored to add self-configuration and self-adaptation capabilities to the data ingestion, orchestration and processing layers. In addition to this work, libraries of modules for data ingestion, orchestration and processing should be developed as a base for the (automated) creation of custom solutions. This should also include the realization of a European foundation model for traffic, transportation and logistics.

4. Conclusions

Deliverable 3.2, “Multimodal Data Management and Orchestration System” documents the results of task 3.2 “Data generation, exposure, acquisition, ingestion and processing for diverse multi-modal passenger and freight transport systems” of WP3. Specifically, it provides an overview of the exploration, design and implementation work done by the partners to realize the necessary data ingestion, management and processing components for the pilot use case systems as well as the overall Delphi concept. From the overview the following key point can be extracted:

- EBOS presented their exploration of standardized data specifications and ontologies as well as an overview of the technical requirements, frameworks and model architectures, developed in DELPHI, that are relevant for data ingestion and processing. This work provides the starting points for the DELPHI data ingestion and processing work as well as provides direction for leveraging Delphi results to contribute to standards and regulations after the project.
- IFAG showed how the extension of the DELPHI system data ingestion and processing towards the edge of transportation and logistic system could look like. This work points out the important topic of edge-cloud continuum that should be explored after the project more in depth to prepare the DELPHI concept for future trends.
- WINGS and NTTD provided technical documentation of the data ingestion, orchestration and processing components that they developed for their corresponding use cases. These documentations provide good examples for how cloud-based data ingestion and processing based on DELPHI concepts can look like.

Besides the partner specific work, this deliverable discusses trends and challenges that data ingestion, orchestration and processing layers of traffic, transportation and logistic systems need to consider today and in the future. Some of these trends are for example autonomous vehicles, humanoid robots, AI agents, adversarial attacks and city-wide sensing infrastructures generating high amounts of data. Additionally, a way for a more generalized solution for the DELPHI data ingestion and processing system is suggested. This part of the deliverable indicates that DELPHI is and can only be a starting point for the digitalization and federation of traffic, transportation and logistic systems.

References

- [1] Papadopoulos, M. (2022). *FeDerated nEtnetwork of pLatforms for Passenger and freigHt Intermodality* [online] NewYorkNews. Available at: <https://www.vehicle.com> [Accessed 05 September 2022].
- [2] Papadopoulos, S., Li, R., Crome, B. and Graham, J. (2022). *FeDerated nEtnetwork of pLatforms for Passenger and freigHt Intermodality*, 15, pp.362-369.
- [3] S. Murarka, A. Jain and L. Singh, "Advanced Techniques in Data Ingestion and Pipelining for Scalable Big Data Platforms: A Comprehensive Review," 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India, 2024, pp. 1-6, doi: 10.1109/ICTBIG64922.2024.10911053.
- [4] Ghazi, Mohd Rehan, and Durgaprasad Gangodkar. "Hadoop, MapReduce and HDFS: a developers perspective." *Procedia Computer Science* 48 (2015): 45-50.
- [5] T. Hlupić and J. Puniš, "An Overview of Current Trends in Data Ingestion and Integration," 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2021, pp. 1265-1270, doi: 10.23919/MIPRO52101.2021.9597149.
- [6] Meehan, John, et al. "Data Ingestion for the Connected World." *Cidr*. Vol. 17. 2017.
- [7] Redyuk, Sergey, et al. "Automating Data Quality Validation for Dynamic Data Ingestion." *EDBT*. 2021.
- [8] Mishra, Vaishali, Ram Chandra Sachan, and Rishit Lakhani. "Navigating the Evolution and Management of AI-Driven Data Platforms: An In-Depth Review and Expert Analysis." *International Conference on Advances in Communication Technology and Computer Engineering*. Cham: Springer Nature Switzerland, 2024.
- [9] Kiourtis, Athanasios, Argyro Mavrogiorgou, and Dimosthenis Kyriazis. "A computer vision-based IoT data ingestion architecture supporting data prioritization." *Health and Technology* 13.3 (2023): 391-411.ⁱ