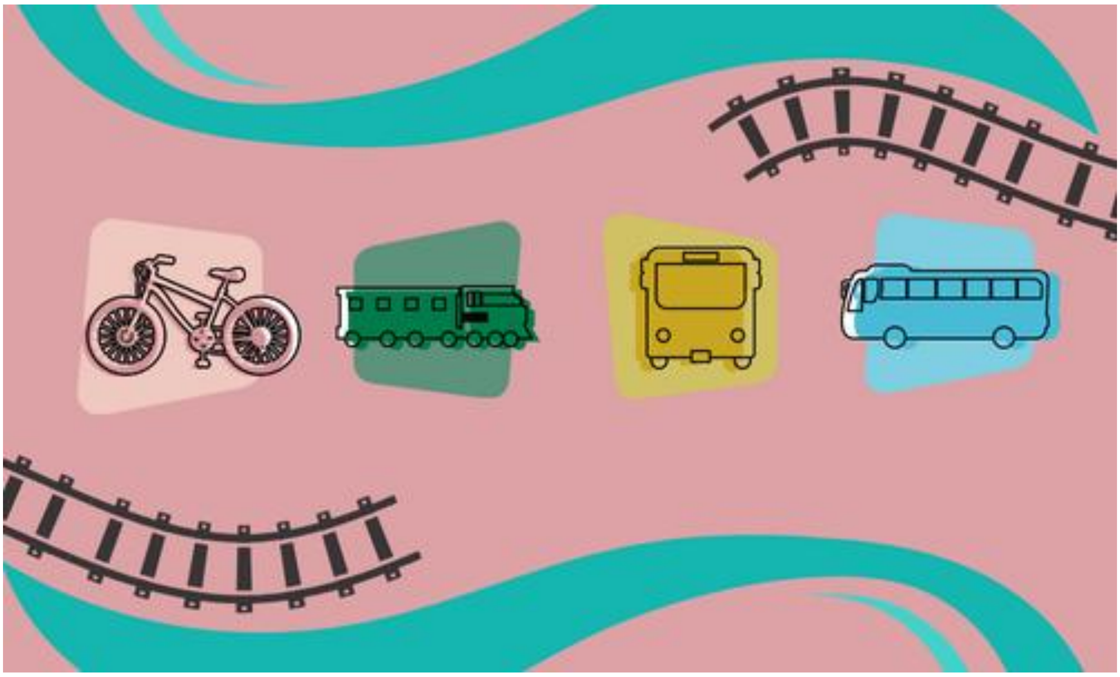
Ref. Ares(2026)2170844 - 26/02/2026

D1.3 - Technical specifications including system functional architecture



Multimodal Optimisation leveraging Data Acquisition from Local Stakeholders towards a Holistic Improvement of Freight and People Transport

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 CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Project Information

Grant Agreement Number	101203040	**Acronym**	MODALSHIFT
Full Title	Multimodal Optimisation leveraging Data Acquisition from Local Stakeholders towards a Holistic Improvement of Freight and people Transport		
Call and topic identifier	HORIZON-CL5-2024-D6-01-06		
Duration	42 months		
Start Date	1 June 2025	**End date**	30 November 2028
Website	https://modalshift.eu/		
Coordinator	Instituto Tecnológico de Aragón (ITA)		

Document Information

Beneficiary responsible	GeneGIS GI (GGI)
Author/-s	Ivo Fumagalli (GGI)
Co-author(s)	David Escuín (ITA) Tiago Dias (ATOBE) AIT Veronique Rottier (ID4M) KEDGE Lynn D'eer (IMEC) Vincent Leblond (TELLAE) Cristiana Caponio (NGS) Olivier Regnier-Coudert (OCTO) Yoann Katchourine (OCTO)
Work Package	WP1 - Contextual analysis, use case study definition and requirements
Deliverable	D1.3 - Technical specifications including system functional architecture

Nature of the deliverable		
R	Document, report (excluding the periodic and final reports)	X
DEM	Demonstrator, pilot, prototype, plan designs	
DEC	Websites, patents filing, press & media actions, videos, etc.	
DATA	Data sets, microdata, etc.	
DMP	Data management plan	
ETHICS	Deliverables related to ethics issues	
SECURITY	Deliverables related to security issues	
OTHER	Software, technical diagram, algorithms, models, etc.	

Dissemination level		
PU	Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)	X
SEN	Sensitive, limited under the conditions of the Grant Agreement	

Quality procedure			
Date	**Version**	**Reviewer**	**Action**
10/02/2026	V1		First draft
13/02/2026	V2	IMEC	Reviewer
19/02/2026	V3	Najat Bara (KEDGE) & S. Ehsan Hashemi-Petroodi (KEDGE)	Reviewer
26/02/2026	V4		Final version

DISCLAIMER

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. This document may contain material, which is the copyright of certain MODALSHIFT consortium parties, and may not be reproduced or copied without permission.

Neither the MODALSHIFT consortium, nor a certain party of the MODALSHIFT consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk and does not accept any liability for loss or damage suffered by any person using this information.

Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authorities can be held responsible for them.

ACKNOWLEDGEMENT

This document is a deliverable of MODALSHIFT project. This project has received funding from the European Union's Horizon Europe Programme under grant agreement No 101203040.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 PROJECT ABSTRACT

The vision of **MODALSHIFT** lies in the creation of a transport network and traffic management optimisation framework, trusted and valuable for local stakeholders, that bridges the data from infrastructures, logistics and mobility operators.

New IoT devices such as a smart box enabling Capacity-as-a-Service, and an e-subscription device for public transport access for vulnerable people, will increase the sources for data collection. A mobility data space, associated to novel geolocation data anonymisation, will be set up in the 3 Case Studies (Bulgaria, Italy, Spain) to ensure trusted and secure data exchange between data providers and users. This multisource data will enhance traffic state forecasting and increase the detection rate of events by 15%. On this basis, predictive and prescriptive analytics, along with synchromodality-based scenarios, will be tested in digital twins and early pilots. These analyses will help identify optimal actions for transport stakeholders to adjust their operations. The objective is to achieve a 25% reduction in interconnection or transhipment delays.

Agent-based modelling will identify how a modal shift towards low-carbon, active and shared mobility services can be acceptable by end-users and support a reshape of the public transport services and the use of urban space. A multimodal traffic management platform will orchestrate, upon the data space, the cooperation of stakeholders at network and multimodal hub scales. It enables the connection of dynamic optimisation algorithms to operational drivers' tool for mobility operators, and of static models to visual interface for transport planners. The determination of governance models, values for each stakeholder, dynamic pricing and business models, will steer the participation of 8 stakeholders for each Case Study in the multimodal traffic management system.

With this approach, MODALSHIFT stimulates new uses of the transport network to reduce traffic congestion for low-carbon and inclusive mobility, avoiding pernicious rebound effects.

2 EXECUTIVE SUMMARY

2.1 PURPOSE

D1.3 - Technical specifications including system functional architecture aims to establish the project's technological baseline by gathering the capabilities of technical partners and outlining the needs and business requirements of the case studies. This allows identifying potential gaps, between technical capabilities and the business requirements, early in the project, ensuring seamless integration of the project's innovations. These gaps shall correspond to the new features and requirements to be implemented within WPs 3 and 4. Some gaps might be already identified from the project proposal phase whereas others may be only identified now.

The document begins by examining the three case studies to identify and articulate both current and future business requirements and use-case scenarios. Next a thorough assessment of the consortium's partners is given, dividing them into different categories and detailing existing capabilities within their technical offering for the project. With the technical strengths and core business needs defined, a gap analysis was conducted to compare what is offered against what is needed, highlighting any shortfalls at an early stage.

This deliverable covers the following points:

- The primary drivers and business requirements for the three case studies.
- An overview of each technological partner, their main offerings, and their role in the project.
- A gap analysis that aligns technical capabilities with identified business needs.
- A technical description that includes:
  - The high-level integrated system architecture and collaboration principles among partners.
  - The core modules slated for implementation.
  - Detailed specifications for the various digital services.

- Detailed specifications for the hardware solutions.

- The data privacy guidelines that each partner must respect to develop their products.

Since several use cases are still being defined and new stakeholders with additional requirements are expected over the next 6–12 months, the technical specifications will be treated as a living document. This agile, iterative approach ensures the teams remain adaptable to evolving project dynamics.

2.2 INTENDED AUDIENCE

This document targets the people involved in the development of the MODALSHIFT's technical features in various levels:

System architects and designers examine the high-level structure, integration points, and trade-offs laid out in the document. The objective is to ensure that the proposed solution fits within the broader ecosystem, scales appropriately, and aligns with the project goals.

Developers turn to the technical specifications as a blueprint. It spells out the exact functional behaviour, the APIs that will be called, data structures that will be manipulated, and any constraints such as performance limits or security requirements. With that information in hand, developers can translate the abstract requirements into concrete code or hardware configurations.

Quality assurance professionals use the specification to derive acceptance criteria and test cases. By mapping each requirement to a measurable outcome, verification that the implementation behaves as intended under both normal and edge-case conditions can be performed.

2.3 DESCRIPTION OF THE MAIN ACTIVITIES

The activity starts with a review of the project proposal and the collection of all available and relevant information. This is followed by discussions with the relevant stakeholders to identify and clarify the specific needs to be addressed. Based on these inputs, the technological tools, including both software and hardware components, are defined to tackle the identified challenges. The system's functional architecture is then designed, describing the main building blocks and their interactions. Finally, data protection and privacy guidelines are defined to ensure compliance across all developed components.

2.4 KEY RESULTS

- **System functional architecture:** defines the main system building blocks and their interconnections.
- **Technical specifications:** technical activities that the participating organizations will carry out within the technical work packages, detailing how the services defined in a previous deliverable will be designed and developed in order to address the identified challenges in the case study definition.
- **Data privacy guidelines:** rules that must be complied with by all software components developed within the project.

2.5 RESEARCH AND IMPLICATIONS

This document contains highly technical content that is specifically tailored to the MODALSHIFT's requirements. It is primarily intended for an internal audience and is designed to support the technical partners in the development of their respective solutions, as well as to facilitate effective collaboration among them.

The technical specifications and system architecture described in this document are defined with the actual project organisation and the selected case studies in mind. They reflect the specific operational contexts and are carefully adapted to the scenarios identified within the project, ensuring that the proposed solutions are aligned with real needs and constraints.

2.6 CONCLUSION

This deliverable establishes the project’s technological baseline by consolidating the current capabilities of the technical partners and the business requirements of the three case studies. By systematically comparing what is needed with what is available, it enables the early identification of gaps between technical offerings and business expectations. These gaps directly inform the new features and requirements to be addressed within following WPs, ensuring a coherent and well-aligned implementation of the project’s innovations.

Through an analysis of use-case scenarios, partner capabilities, and a structured gap assessment, the deliverable provides a shared understanding of the project’s technical direction. It defines the foundations for system architecture, core modules, digital and hardware solutions, and data privacy obligations, creating a common reference point for all partners.

3 ABBREVIATIONS

Abbreviation	Definition
ABM	Activity-based modelling
ADF	Adriafer s.r.l.
AI	Artificial intelligence
AIT	Austrian institute of technology
AMQP	Advanced message queuing protocol
ANSSI	French national agency for the security of information systems
API	Application programming interface
ATOBE	A-to-Be mobility technology sa
AVANZA	Avanza Spain s.l.
CaaS	Capacity as a service
CBV	Core business vocabulary
CCAM	Cooperative connected autonomous mobility
CCPA	California consumer privacy act
CESIT	Centre of excellence in supply chain innovation and transportation
C-ITS	Cooperative intelligent transport systems and services
CIL	City intelligence lab
CJEU	Court of justice of the European Union
CNIL	Commission nationale de l'informatique et des libertés: is the French national commission on informatics and liberty
CO_2	Carbon dioxide
CRUD	Create, read, update and delete
CS	Case study
CSV	Comma separated values
DNS	Domain name system
DS	Data space
DSC	Data space connector
DT	Digital twin
EDPB	European data protection board
EDPS	European data protection supervisor
eFTI	Electronic freight transport information
EPCIS	Electronic product code information services

EU	European Union
GDPR	General data protection regulation
GGI	GeneGIS GI
GIS	Geographic information system
GPS	Global positioning system
GTFS	General transit feed specification
H2020	Horizon 2020
HIPAA	Health insurance portability and accountability act
Horeca	Hotel, restaurant and café
HTTPS	Hypertext transfer protocol secure
ICT	Information and communications technology
ID	Identifier
ID4M	ID4Mobility
IM	Instant messaging
IMEC	Interuniversitair micro-electronica centrum (interuniversity microelectronics centre)
IoT	Internet of things
ITA	Instituto tecnológico de Aragón (technological institute of Aragon)
JSON	JavaScript object notation
KEDGE	KEDGE business school
KPI	Key performance indicator
LEZ	Low emission zone
LiDAR	Light detection and ranging
ML	Machine learning
MTGNN	Multivariate time series forecasting with graph neural networks
MTM	Multimodal traffic management
MTMS	Multimodal traffic management system
NGS	New generation sensors
nLDP	See nFADP
nFADP	New Swiss federal act on data protection
OBU	On board unit
OCTO	Octopize mimethik data - big data santé
PCS	Port community system
PIA	Privacy impact assessment

PDF	Portable document format
PNAEAS	Port network authority of the Eastern Adriatic Sea
PNRR	Piano Nazionale di Ripresa e Resilienza: is the Italy's recovery and resilience plan
PTV	Planung Transport Verkehr GmbH
QGIS	Quantum GIS
R&D	Research and development
REST	Representational state transfer
RFI	Rete Ferroviaria Italiana: is the Italian railway infrastructure manager
RTK	Real-time kinematic positioning
RTO	Research and technology organization
SaaS	Software as a service
SECO	State secretariat for economic affairs
SME	Small medium enterprise
SP-off-RP	Stated-preference-off–revealed-preference
SRB	Single resolution board
SSN	Social security number
STG	Spatial-temporal graph
SUMO	Simulation of urban mobility
SUMP	Sustainable Urban Logistics Planning
T&T&M	Track & Trace & Monitoring
TEN-T	Trans-European transport network
TRL	Technology readiness level
UC	Use case
V2X	Vehicle to everything
VRU	Vulnerable road user
WP	Work package
WP29	Working party 29

Table 1: List of abbreviations

4 INTRODUCTION

Deliverable 1.3 “Technical specifications including system functional architecture” aims to define the technological starting point joining the business requirements of the three Case Studies with the technical expertise of the technological partners.

A chapter is dedicated to the data privacy guidelines that define the way data is handled in this project to foster trust among stakeholders.

4.1 CASE STUDIES

In an era where efficient, sustainable, and interconnected mobility is paramount, cities and transport hubs across Europe are leveraging digital innovation to address longstanding challenges. MODALSHIFT explores three case studies aimed at transforming logistics and urban mobility through cutting-edge technologies and collaborative platforms.

The Port of Trieste, a vital Mediterranean hub, is tackling inefficiencies caused by limited interoperability and aging infrastructure. By developing a collaborative digital platform powered by AI and digital twins, the port seeks to optimize shunting operations, reduce delays, and create a scalable model for other European ports.

In Varna, Bulgaria, seasonal tourism and traffic congestion strain urban mobility. The city’s response includes a "pay-per-use" e-cargo bike system and enhanced train-bus connectivity, supported by IoT sensors and AI-driven digital twins. This initiative aims to reduce emissions, alleviate congestion, and provide a replicable framework for cities facing similar challenges.

Madrid’s Estación Sur, a bustling multimodal hub serving over 20 million passengers annually, is addressing congestion and fragmented communication through a digital twin and smart locker systems. These innovations, integrated into Madrid’s Sustainable Urban Mobility Plan, promise to improve punctuality and optimize resource allocation.

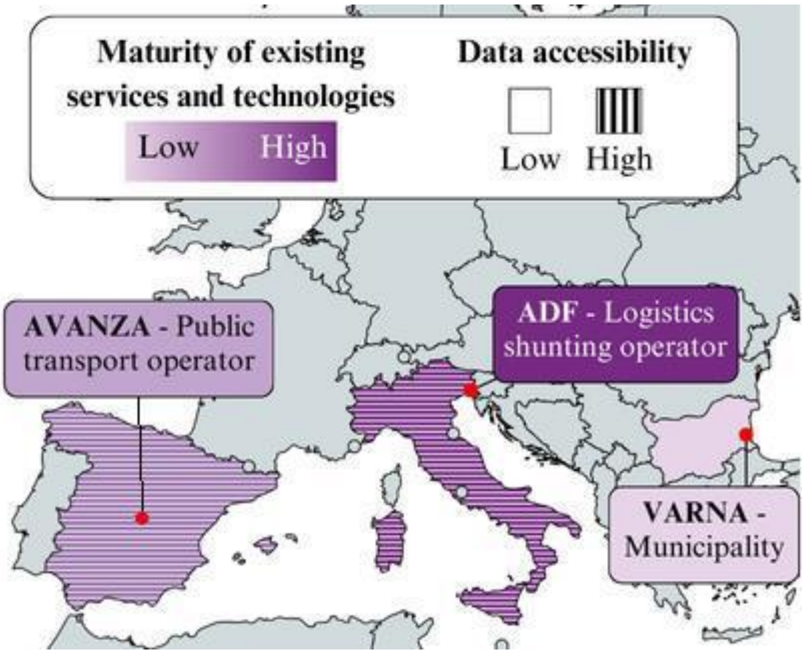


Figure 1 Case studies location

4.1.1 PORT OF TRIESTE

4.1.1.1 Overview

Trieste is the capital city of the Friuli Venezia Giulia region in Italy, it has about 200'000 inhabitants and it is located at the north of the Adriatic Sea, close to the Slovenian border. The Port of Trieste, an essential hub in the regional logistics chain, saw railway traffic surge dramatically in 2023, processing more than 12'000 trains in that year. Trieste is Italy's leading cargo port and the Mediterranean's largest oil terminal. It serves as a pivotal connector within the European TEN-T Mediterranean and Adriatic-Baltic corridors, linking Austria and Slovenia with major cities across the North Adriatic and Eastern Europe. Its rail infrastructure, fully integrated with both national and international networks, enables every dock to be reached directly by track, with the option to shunt or assemble trains right at the individual terminals. The last-mile rail operations, critical to the region's rail efficiency and competitiveness, are overseen by ADF, a subsidiary of PNAEAS (the Port Network Authority of the Eastern Adriatic Sea).

Figure 2 Port of Trieste

4.1.1.2 Use case

The primary objective is to boost interoperability among waterways, rail, and last-mile transport so that goods move smoothly from one mode to the next. To do this, MODALSHIFT aims to improve real-time data exchange between the national railway infrastructure (RFI) and ADF regarding disruptions and delays, synchronize loading and unloading timetables, allocate last-mile shunting slots, coordinate train departures and arrivals and assessing how passenger-trains delays affect freight operations.

These goals will be achieved by building a collaborative platform that pulls together information from port authorities, rail operators, and logistics firms. Using PCS Sinfomar, the platform will aggregate data from maritime, customs, terminal, shunting, and IM activities. It will further incorporate digital twins and AI-driven simulations to forecast traffic patterns, enabling dynamic decision-making and more efficient resource management.

4.1.1.3 Impact and exploitation

The outcome will be best practices to be shared with other shunting managers and railway companies in the region. Also, ADF promotes initiatives to improve the railway sector highlighting the environmental benefit it brings.

4.1.2 MUNICIPALITY OF VARNA

4.1.2.1 Overview

Varna is a Bulgarian city located on the Black Sea; with one million people, it is the second largest city in the country after the capital Sofia. Its municipality already offers several essential infrastructures and services, such as a pedestrian zone in the downtown area and a robust public-transport system that includes trains, buses (among them 60 electric ones) and trolleybuses. While the Sustainable Urban Mobility Plan (SUMP) is well advanced in its design phase, its implementation has fallen behind schedule. The city experiences strong seasonal pressures, with mobility demand surging in summer when the population effectively doubles. During this peak period, a maritime shuttle operates, particularly when the bridge linking Varna's northern and southern areas is out of service. Traffic-light management is overseen by a municipal company, yet urban planning and modelling capabilities remain modest. The initiative will build on a range of past and ongoing innovation projects in Varna, such as the H2020 CityChanger CargoBike (2018-2022) program that introduced six locally produced e-cargobikes for free trials by logistics operators in the city centre, alongside the efforts to electrify public transport.

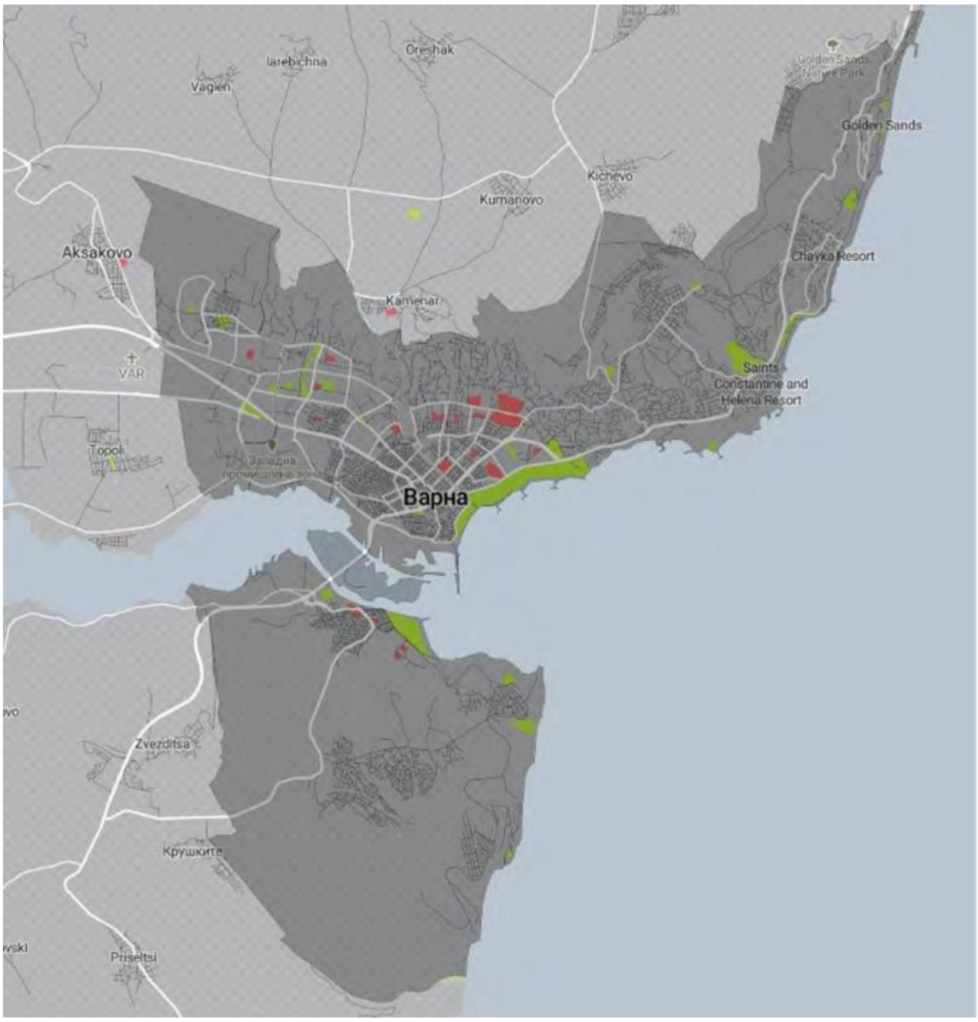


Figure 3 Varna

4.1.2.2 Use cases

The first use case (UC) aims to boost low-carbon urban freight by tapping into the municipality's existing cargo-bike fleet, which should ease traffic in the city centre. Sensors on the bikes will be installed and a specialised "pay-per-use" bike-sharing platform for small businesses and Horeca will be developed. This tool will let the users check bike availability, make reservations, and gather valuable data for traffic-flow analysis and forecasting.

The second UC targets broader improvements to public transport and active mobility across two fronts: first enhancing intermodal connections between trains and buses at the railway station, and second redesigning public-transport and active-mobility schedules to better match seasonal demand and upcoming bike-lane networks. Because data is currently limited, the project also seeks to raise data literacy and refine collection methods by establishing a data space and a digital twin. Development will focus on ensuring that the new collaborative digital platform can integrate smoothly with legacy traffic-infrastructure systems. These upgrades will run in parallel with the work planned for Estación Sur, as described in the Spanish CS.

4.1.2.3 Impact and exploitation

MODALSHIFT will support the city in easing the traffic and reducing pollutant emissions, promoting a synergy between local business and a sustainable transportation solution. ID4M and AIT will help with guidelines for the completion and implementation of the SUMP providing a baseline for other cities to implement green logistics in long-term planning.

4.1.3 MOBILITY/FREIGHT IN BUSES (MADRID)

4.1.3.1 Overview

Madrid spans 605.77 km^2 and is home to roughly 3.3 million people; its wider metropolitan region stretches over 8.000 km^2 and accommodates about 6.7 million residents. AVANZA, the leader private carrier within the Madrid Transport Consortium, contracts the core commuter routes that link Madrid's periphery. Within the capital, AVANZA's operations concentrate on the southern districts, while also serving many towns throughout the Community of Madrid.

Located in the city's south, Estación Sur functions as a strategic transport hub where several travel modes for both passenger and freight, intersect. This key node underpins AVANZA's extensive suburban and inter-urban network, moving millions of commuters each day. By strengthening regional connectivity and sustainability, it improves access to education, health care, and commerce, thereby raising the overall quality of life for residents.

Figure 4 Estación Sur in Madrid

4.1.3.2 Use cases

In the first use case a digital twin will be built to virtually represent the entire transport network of Madrid, drawing on open mobility data, integrating additional data sources, and producing upgraded public-transport services throughout the Community of Madrid. From both planning and operational perspectives, inter-city bus routes and schedules will be redesigned, resource utilization will be optimized, and on-demand mobility options will be examined to improve punctuality, reliability, user satisfaction, and overall adoption.

The second use case is at the multimodal hub of Estación Sur (South Station) where the communication system among all stakeholders will be unified to facilitate real-time data exchange, including up-to-date traffic information, enabling dynamic adjustments to operations such as routing, scheduling, and shared-infrastructure allocation. A focus will be placed on developing synchromodality: coordinated passenger transport services that reduce connection delays, together with a Capacity-as-a-Service model that makes available public-transport capacity usable for freight.

4.1.3.3 Impact and exploitation

The use case will be replicable in all AVANZA's companies. The project result will be designed for high replicability, enabling other public transport companies and municipalities to easily implement them.

4.2 TECHNOLOGICAL EXPERTISE

In this section the technical partners are introduced that are involved in MODALSHIFT with a presentation of their capabilities and their contribution to the project.

4.2.1 ACADEMY

4.2.1.1 KEDGE Business School - KEDGE

KEDGE Business School (KEDGE) is a vibrant business school with 4 campuses in France (Paris, Bordeaux, Marseille, and Toulon), 3 abroad with 2 in China (Shanghai and Suzhou), 1 in Africa (Dakar) and 3 associated campuses in France (Avignon, Bastia, Bayonne). KEDGE is ranked 31st amongst European Business Schools, KEDGE Business School is AACSB, EQUIS and AMBA accredited, and a member of the French Conférence des Grandes Écoles.

The team involved in MODALSHIFT are attached to the Centre of Excellence in Supply Chain Innovation and Transportation (CESIT), one of the research centres at KEDGE that aims at meeting the economic, environmental and societal challenges raised in the Logistics and Supply Chain networks. Their actions impact academic research, the corporate world, society and education. CESIT experts are very active in several domains, such as Operations Research, Combinatorial Optimisation, Simulation applied in Supply Chain Management, Urban and Maritime Logistics, and Manufacturing System Design.

Colleagues from KEDGE recently addressed critical challenges in multimodal city logistics by exploring innovations such as dynamic delivery systems and crowdsourced transportation, aligning them with the principles of the Physical Internet and Hyperconnected City Logistics. Moreover, KEDGE studied several research on urban synchromodality for highlighting synergies between passenger and freight mobility, leveraging public transport networks for on-demand urban delivery services, with a complex, time-constrained scheduling model optimizing multi-mode synchronization and capacity-constrained resources, aiming for high service levels with minimal urban footprint. Within these studies, KEDGE researchers have been developed several simulation models, mathematical programming models, and optimisation algorithms that can efficiently solve any of those certain logistic problems targeted.

MODALSHIFT aims to enhance multimodal transportation by integrating traffic control and emission optimization/energy consumption through advanced techniques that foster the improved quality and availability of data.

As a result, match-making the abilities and knowledge at KEDGE and the need of MODALSHIFT, KEDGE can develop new optimisation techniques as well as adapt several of the techniques that have been developed in the literature to the use cases existing in MODALSHIFT. Several mathematical programming models will be developed that would make some bases of the decision-making technique adapted for the need of each use case. Notably the combination of AI/machine learning (ML) techniques with robust and stochastic optimization and dynamic programming. With this strategy, the dynamic decision-making algorithms that can enhance several KPIs considered in the project concerning any use case in the MODALSHIFT. Merging optimisation techniques that will be developed by KEDGE with AI and learning techniques provided by IMEC will help to rely on more robust and accurate results for the project.

The contribution from KEDGE will be focused on developing optimisation algorithms to enhance the decision-making process. The base modelling approach for optimisation is to build mathematical programming models for each use case. Precisely, mathematical programming and optimisation algorithms are systematic methods used to support and improve decision-making by translating real-world problems into formal mathematical models, where decisions are represented as variables, objectives define what should be maximised or minimised (such as profit, cost, or efficiency), and constraints capture the practical limitations of the system. Note that all the data/parameters given from the system either in deterministic or uncertain forms are fed to these techniques to finally provide the best/optimal decisions. These models are then solved using specialised algorithms, such as linear, mixed-integer, or nonlinear optimisation techniques that intelligently explore the space of feasible solutions, using mathematical properties (e.g., convexity, duality, and branching strategies) to efficiently converge toward optimal or near-optimal decisions. In practice, this allows complex trade-offs between competing objectives and constraints to be evaluated rigorously and consistently, providing decision-makers with solutions that are not only optimal in theory but also implementable and robust in real operational environments. In addition, considering uncertainties in the systems, there are several robust optimisation algorithms to develop which can provide the most robust solutions optimising for the worst-case uncertainty possible to the system. As mentioned, merging these optimisation techniques with learning ones which provide more accurate estimation of the future uncertain data, would lead to not only robust but more accurate results.

For the moment the main focus is on two use cases in Trieste and Madrid and a corresponding mathematical model will be developed for the case in Varna. As example, here are some possible ideas of optimisation techniques for two of use cases:

- **Madrid:** In the context of collaboration between multiple stakeholders and with the objective of reducing congestion inside the city, KEDGE aims to design a mathematical model integrating freight and passenger transport within the urban area to improve overall traffic conditions for the use case of Madrid. Following the perspective of hyperconnected city logistics, the city will be represented as a multi-echelon system. Under a collaborative framework involving public and private actors, each tier will include decisions related to the allocation of freight demand to intermodal facilities, enabling public transport vehicles to carry freight alongside passengers. Once public transport enters the city centre, the model will determine how city freighters deliver or pick up the parcels. The initial focus will be on small parcels. To achieve this, the problem will be modelled as a combination of a **service network design problem** and a **vehicle routing problem** across multiple tiers. In the middle tier (public transport), the model can be extended to incorporate on-demand passenger subscriptions. Vulnerable users will be represented as a specific commodity with origins inside the public network and destinations in the last tier. The mathematical model that will be used, will reduce the transportation and externalities costs while respecting some constraints of capacity about the vehicles and facilities. To solve this model, a decomposition method can be developed solving stochastic model like a Benders method which is a known optimisation algorithm.
- For **Trieste** use case, this will be an extension of the **Madrid** model. In addition to optimise the transportation and externalities costs, KEDGE will also try to optimise the shunting operations, to reduce the delays by considering the different capacity constraints. The method of resolution can be a kind of decomposition algorithm similar to the case of Madrid, or a robust optimisation algorithm considering uncertainty in the network.

4.2.1.2 Interuniversitair Micro-Electronica Centrum – IMEC

IMEC (Interuniversity Microelectronics Centre) is a world-leading research institute based in Leuven, Belgium, focused on nanoelectronics and digital technologies. It brings together industry leaders, academia, and governments to advance innovations in semiconductors, artificial intelligence, healthcare, energy, and sustainable technologies.

IMEC has developed expertise on road traffic prediction models capable of improving the accuracy of traffic state forecasting through additional components like road features, time features and/or weather information. These models are applicable on both public transport and logistics traffic as well as prediction on road network, particularly on segments where induction loops, cameras, and/or map-matched vehicle trajectories are available. Moreover, these models can forecast the freight demand (e.g., volume of incoming cargo) and capacity (e.g., workforce capacity by men hours) at logistics hubs which is useful for the optimization of transport routing.

MODALSHIFT will improve IMEC's prediction models abilities by researching forecasting model architectures, transfer learning techniques, and training strategies such training with much shorter duration of training data.

Below, the techniques used in MODALSHIFT are detailed.

Road and Rail Traffic Forecasting

Traffic state forecasting refers to the prediction of future traffic conditions, such as speed, flow, density, or congestion levels, across a transportation network which can be multimodal, and enables proactive traffic management, route optimization, and urban planning.

The temporal horizon of such prediction can vary. Short-term forecasting predicts the traffic states in minutes to hours ahead and is used for dynamic traffic control. Longer-term predictions where hours, days or even weeks to months ahead are predicted, are used for logistics and planning, infrastructure development and policymaking.

Traffic state forecasting models often use historical traffic data to detect patterns over time. Moreover, (real-time) sensor data coming from loop detectors and cameras, and geolocated data can further advance predictions. In addition, road features, time features and weather information may improve the accuracy of the forecasts. State-of-the-art forecasting models will be used as benchmarks for the forecasting research including Graph WaveNet (Zonghan W. et al. 2019), MTGNN (Zonghan W. et al. 2020), STG-Mamba (Lincan L. et al.) and PatchSTG (Yuchen et al.).

A significant challenge in traffic state predictions is the lack of data, especially in a multimodal setting such as the case studies of MODALSHIFT, where multi-source data are present with varying access and quality. IMEC will improve the models' multimodality by combining different data sources that MODALSHIFT provides. This includes rail traffic data, possible anonymized data, and so on.

Additionally, transfer learning techniques will be used, i.e., a machine learning technique where a model trained on one task is reused to improve performance on a related task. This approach is especially useful when there is limited data available for the new task or case study, as it allows the model to leverage knowledge gained from the original task or case study. Possible research directions for this include parameter efficient transfer learning, domain adaptation from cities with sufficient data to cities without that, and synthetic data generation using generative diffusion models.

The specific contribution to the case studies is as follows:

- Port of Trieste: IMEC will study road and rail forecasting models to further develop the current network state forecasting and extend it to rail network application to provide more effective and seamless transfer of goods, and to enable dynamic decision making and more efficient resource management.
- Estación Sur Madrid: the forecasting models IMEC will study aim to improve traffic and freight demand forecasting across road and rail networks, even in data-scare areas, specifically to advance passenger mobility. In addition, it enhances a better understanding of Madrid's transport network for better urban planning decisions, including optimized traffic management for passenger and last-mile transport.

4.2.1.3 Austrian Institute of Technology – AIT

The AIT Austrian Institute of Technology is Austria's largest Research and Technology Organization (RTO), symbolizing the country's strong commitment to research and innovation. While deeply rooted in Austria, AIT plays a significant role on the international stage.

In line with its mission, AIT is strategically positioned as a key player within the Austrian and European innovation landscape by conducting applied research and fostering the market adoption of innovative infrastructure-related solutions. Acting as a bridge between research and technology commercialization, AIT facilitates the development of ground-breaking technologies and drives economic growth.

The "Digital Resilient Cities" team, part of the AIT Centre for Energy, develops cutting-edge urban management and planning solutions for creating sustainable, smart, liveable, and resilient urban environments. The interdisciplinary team consists of experienced experts, including urban planners, architects, transport and energy specialists, spatial planners, statisticians, geoscientists, and IT professionals, who work collaboratively to devise innovative, practice-oriented solutions. A particular area of expertise lies in measuring, analysing, simulating, and managing traffic flows.

The City Intelligence Lab (CIL), a core initiative within "Digital Resilient Cities," serves as a pioneering platform for collaborative urban planning research and development. Utilizing advanced technologies such as artificial intelligence, generative design, and immersive visualizations, the CIL introduces ground-breaking approaches to shaping urban environments. The lab provides a unique space where cities, planners, and researchers co-create innovative solutions to complex urban challenges, ranging from traffic optimization to energy-efficient building designs and resilient urban development.

A particular area of AIT's expertise lies in measuring, analysing, and managing complex traffic flows through a sophisticated multimodal mobility simulation framework. This asset integrates traditional 4-step models with high-fidelity agent-based modelling in MATSim, allowing researchers to investigate the impacts of specific interventions on the multimodal system. AIT has also provided open-access implementations such as the MATSim model of Vienna.

Complementing this effort, AIT has developed an activity-based modelling approach which is a Python-based implementation designed for rapid assessment in data-scarce settings. Unlike conventional trip-based models, this framework provides a more detailed depiction of individual daily trip patterns and improves on trip/tour generation, destination choice, and mode choice. It enables rapid assessments due to low computational runtimes and is designed for integration with external assignment and routing systems such as PTV Visum or SUMO. The model supports scalable, adaptable simulations of mobility scenarios, making it ideal for urban planning in rapidly growing or under-resourced contexts.

AIT further conducts advanced mobility simulations across multiple spatial and modal scales. These include also pedestrian flow simulations using AIT's agent-based simulation framework SIMULATE. These tools allow for microscopic analysis of interactions between different transport modes and spatial configurations, supporting evidence-based planning of transport infrastructure and public spaces.

In addition to its core modelling capabilities, AIT has developed and applied a range of complementary tools and methods for empirical mobility analysis and simulation. This includes the MyTrips survey tool for the collection of stated-preference-off–revealed-preference (SP-off-RP) data based on respondents' actual mobility behaviour. MyTrips (Rudloff et al.) enables detailed insight into individual travel decisions and supports behaviourally informed transport modelling and policy evaluation.

4.2.1.4 Technological Institute of Aragon – ITA

The Technological Institute of Aragon (ITA) is a leading Spanish Research and Technology Organisation with a strong focus on applied research, technology transfer, and innovation in mobility, logistics, and digital systems. Deeply embedded in the regional innovation ecosystem of Aragón, ITA plays an active role at the European level and has extensive experience participating in and coordinating EU-funded research and innovation projects.

ITA acts as a key enabler of digital transformation for transport and logistics systems, bridging advanced research with real-world deployment. ITA supports public authorities and private stakeholders in the design, implementation, and adoption of innovative solutions, contributing to more efficient, sustainable, and resilient mobility and freight transport networks.

A core area of ITA's expertise lies in the development and application of digital twins as virtual models for the simulation, analysis, and optimisation of complex transport systems. ITA has developed multi-agent digital twin frameworks that enable scenario-based evaluation of multimodal and synchromodal transport strategies, capacity allocation, and traffic management measures. These digital twins allow stakeholders to test alternative policies and operational strategies in a risk-free virtual environment before real-world implementation and have been successfully applied in previous European projects such as AMIGOS, ICONET and DECARBOMILE.

Complementing its digital twin capabilities, ITA has strong expertise in data spaces for mobility and logistics, focusing on interoperability, data governance, and secure data sharing across multiple actors. ITA has designed and deployed mobility data space architectures based on FIWARE technologies, including the implementation of FIWARE-based Data Space Connectors (DSC) to enable trusted, sovereign, and standardised data exchange between public and private stakeholders. These developments are openly documented and maintained through ITA's data space repositories.

Through the integration of digital twins and data spaces, ITA enables end-to-end decision-support ecosystems that combine real-time and historical data ingestion, advanced analytics, and simulation-based optimisation. This approach supports system-wide traffic forecasting, impact assessment, and evidence-based policy design in complex multimodal environments.

Within MODALSHIFT, ITA acts as project coordinator and leads the development of the multimodal transport network digital twin as well as the design and structuring of trusted mobility data spaces. ITA ensures technical coherence across work packages, alignment with European standards and initiatives, and effective collaboration among consortium partners. By combining its strengths in digital twins, data spaces, and system integration, ITA provides the technological backbone required to achieve MODALSHIFT's objectives and to enable scalable, replicable solutions across European transport networks.

4.2.2 HARDWARE

4.2.2.1 New Generation Sensors – NGS

New Generation Sensors (NGS) is an Italian SME, founded in 2015 as a spin-off of the Scuola Superiore Sant'Anna of Pisa, specializing in Internet of Things (IoT) technologies for logistics, environmental monitoring, and smart process control. The company designs and develops low-power hardware devices, custom firmware, and integrated cloud platforms for real-time data collection, processing, and sharing.

NGS aims to enable interconnected, transparent, and sustainable ecosystems, in full alignment with the Physical Internet vision, where goods, vehicles, and infrastructures communicate through standardized and interoperable technologies.

To advance this mission, the company is actively involved in several national research initiatives within the Italian PNRR and European Horizon projects, contributing its expertise in IoT systems, interoperability, and data standardization.

TrackOne solution

TrackOne is NGS's proprietary IoT data management and analytics platform, designed to ensure visibility, reliability, and interoperability across logistics and transport networks. Built on open standards and a cloud-native architecture, it collects and harmonizes data from multiple devices and stakeholders, providing a unified and secure view of the supply chain.

The platform enables real-time awareness and data-driven collaboration, supporting decision-making and improving efficiency across multimodal transport operations.

Compliant with the GS1 EPCIS standard and integrated with blockchain, TrackOne guarantees traceability, transparency, and secure data exchange. Through its modular and interoperable design, it transforms traditional logistics into connected, intelligent, and collaborative networks, aligned with the Physical Internet vision.

In its full configuration, the platform combines:

- IoT devices for sensing and tracking,
- a cloud application and dashboard for data processing and visualization,
- and a blockchain ledger connection to certify the authenticity and integrity of shared information.

NGS in the MODALSHIFT Project

Within the project, NGS provides the IoT hardware infrastructure forming the physical layer for data acquisition across the three Case Studies

- Trieste – Port Intermodality

  In Trieste, NGS contributes to the improvement of data collection and interoperability within the port and rail ecosystem. Discussions are ongoing between NGS, GGI, and ADF to define the possible modalities of NGS's contribution, including the potential deployment of dedicated IoT devices and data integration services within the port and rail premises. Within this framework, NGS supports the definition and progressive implementation of IoT-based monitoring solutions for locomotives and operational assets, enabling the acquisition of high-frequency positional and operational data. When deployed, these data are standardised by NGS in compliance with GS1 EPCIS 2.0 specifications and integrated into the Dynamic Rolling Stocks Management System developed by GGI.

  Such an infrastructure, when implemented, is expected to support advanced data acquisition aligned with European interoperability frameworks and to enable the integration of real-time operational data with network reference information, traffic conditions, and event-based data. This strengthens coordination between port, rail, and hinterland logistics operations and supports predictive analysis and dynamic resource allocation. Through this contribution, NGS aims to enable improved visibility of rolling stock status, utilisation patterns, and operational constraints, supporting more efficient shunting planning, reduction of delays and disruptions, and enhanced multimodal coordination. By integrating heterogeneous data sources into a unified operational framework, a solution will be jointly developed to overcome the limitations of mode-specific management tools and to contribute to a more resilient and data-driven approach to multimodal freight operations.

- Varna

  In the Varna Case Study, NGS contributes to the implementation of an IoT-based infrastructure supporting innovative urban logistics and mobility services developed in cooperation with the municipality and local stakeholders.

  NGS designs and provides connected IoT devices to be installed on cargo bikes used by small local businesses for last-mile goods transport within the city.

  These devices enable continuous monitoring of vehicle position, usage patterns, battery status, and operational parameters. The collected data are automatically converted into the GS1 EPCIS 2.0 standard and transmitted to the project's shared digital environment, where they are integrated with municipal and open data sources.

This integrated data flow enhances data visibility, supports predictive maintenance and operational analysis, and enables evidence-based planning for sustainable urban logistics. By combining real-time monitoring with flexible reservation mechanisms, the service enables professional pay-per-use access to municipal cargo-bike fleets and facilitates the inclusion of shared mobility assets into city-wide transport planning. NGS's contribution in Varna strengthens the technical foundations for interoperable data exchange between local operators, public authorities, and platform providers, fostering scalable and replicable smart city logistics solutions.

- Madrid – Estación Sur

In the Madrid Case Study at Estación Sur, NGS contributes to the implementation of complementary digital infrastructures supporting Capacity-as-a-Service models and connected multimodal mobility services.

Smart Box and Smart Locker Infrastructure

NGS designs and deploys a Smart Box and Smart Locker system integrated with a cloud-based management platform, enabling the dynamic allocation and management of available space on buses and terminal facilities for goods transport in combination with passenger services.

The Smart Box and Locker solutions support real-time tracking of parcels and containers, automated access control, environmental and operational monitoring, and secure data transmission. This infrastructure enables continuous supervision of capacity usage and service performance, contributing to improved reliability, transparency, and optimisation of multimodal transport services.

To go beyond the state of the art, NGS aims to develop Smart Boxes and Smart lockers that introduce the integration of freight flows into public transport systems through a Capacity-as-a-Service model. They enable automated parcel exchanges, reduces congestion and emissions, and support small local businesses. Standardised interoperability and traceability mechanisms ensure seamless data exchange across stakeholders and platforms.

All data generated by the Smart Box and Locker systems are made available through the project Data Space and analysed within the Multimodal Transport Management Tool, supporting synchromodal optimisation and coordinated operational planning.

From a technological perspective, the Smart Locker is designed to reach TRL5/6 within the project, focusing on the construction and technical validation of functional prototypes. Different transfer management modes are tested to identify the most efficient and reliable configurations for public transport environments. The Smart Box, which reaches TRL7, represents a high-maturity, standardised logistics unit equipped with an advanced Track&Trace&Monitoring system. It enables real-time data collection and automated characterisation of logistics units, acting as the core technological enabler of the Capacity-as-a-Service model.

From a technical perspective, the service relies on routing and timetable data from both couriers and public transport operators to coordinate loading and unloading events. These data are complemented by real-time vehicle status, congestion levels, and timing updates provided by multimodal traffic management systems, enabling continuous adaptation of logistics operations to evolve traffic conditions.

On-Board Unit and e-Subscription Device

This part focuses on the design, implementation, and validation of an integrated IoT ecosystem composed of an On-Board Unit (OBU)/Gateway, proximity transceivers, and personal Bluetooth beacons, developed within the Madrid use case. The primary objective of this service is to enhance the safety, accessibility, and wellbeing of Vulnerable Road Users (VRUs).

The implementation of this service builds on the availability of technical interfaces on board vehicles and on access to operational and scheduling data provided by public transport operators, including routes, stops, and timetables. Municipal authorities support the recruitment of pilot participants and the definition of user registration data, enabling the deployment of the service in real operational conditions.

The system combines historical and real-time bus data, including vehicle trajectories, stop-level information, timetable deviations, connectivity status, and device interaction events. In parallel, user-related data generated by personal trackers are collected and formatted in compliance with GS1 EPCIS standards. This integrated data flow enables accurate detection of boarding and alighting events, proactive safety alerts, and continuous monitoring of service performance.

The architecture is composed by the following elements:

Bluetooth Beacon (Personal Device)

Each vulnerable user is equipped with a compact Bluetooth beacon with a credit-card-like form factor. The device uniquely identifies the user and continuously communicates with the onboard infrastructure, enabling presence detection, flow tracking, and occupancy measurement.

Proximity Transceiver

Proximity transceivers are the central processing and communication hub of the system. Its main functions include:

- acquisition of precise geographical positioning through GPS/RTK, supporting accurate bus stop identification and V2X communication,
- monitoring of onboard environmental parameters, such as temperature and humidity,
- processing and validation of data collected from beacons and transceivers,
- management of data transmission to cloud platforms and external V2X systems.

This component ensures the reliable aggregation, validation, and secure exchange of operational and user-related data.

The aforementioned architecture provides three complementary digital services supporting vulnerable users and transport operators.

A. Automated Presence Monitoring (Be-In / Be-Out Proof of Concept)

This functionality enables the automated detection of passenger presence without requiring any manual interaction. The system logs the date, time, and specific bus stop associated with each user's boarding and alighting events by analysing beacon signals.

This proof-of-concept demonstrates the feasibility of reliable, real-time tracking of vulnerable user flows and vehicle occupancy. The collected data support improved punctuality, accessibility, and operational planning.

B. Proactive V2X Safety for Vulnerable Road Users

Proactive V2X safety mechanisms represent the core protection feature of this service. When a user approaches a vehicle exit, the proximity transceiver triggers the Gateway, which broadcasts a V2X alert to nearby connected vehicles.

This alert informs surrounding traffic participants that a vulnerable passenger is about to alight, reducing the risk of accidents and improving situational awareness. The system is validated at TRL6 on the transmitter side, ensuring readiness for integration within future smart city and connected mobility ecosystems.

C. On-Board Wellbeing and Environmental Monitoring

The OBU/Gateway integrates environmental sensors to monitor onboard comfort conditions, which are particularly relevant for fragile users.

The system automatically generates alerts when predefined thresholds, such as excessive temperature or humidity levels, are exceeded. These notifications are transmitted to cloud platforms, enabling fleet managers to take timely corrective actions.

This functionality contributes to public health protection and enhances the overall quality of transport services

All data generated by beacons, transceivers, and the OBU/Gateway are processed, validated, and transmitted to the project's digital platforms. Through standardised interfaces, these data streams support:

- real-time monitoring of passenger flows and safety conditions,
- integration with multimodal planning and operational systems,
- performance analysis and reporting,
- decision-support processes for transport operators.

The architecture enables secure, scalable, and interoperable data exchange, in line with the project's Trust-by-Design and multimodal integration principles

Cross-cutting contribution - Multimodal, Trust-by-design & Secure Data Space

Across all three Case Studies, NGS contributes to the development of the Multimodal, Trust-by-Design & Secure Data Space by providing a GS1 EPCIS-compliant repository and a blockchain-enabled cloud gateway, ensuring secure, traceable, and standardised logistics data exchange across all pilots — the Port of Trieste, Varna Municipality, and Estación Sur in Madrid — supporting future progress in multimodal data integration and regulation.

4.2.3 SOFTWARE

4.2.3.1 A-to-Be - MOBILITY TECHNOLOGY SA – ATOBE

ATOBE is a leading infrastructure technology provider, part of the Brisa Group, offering solutions for motorways, bridges, tunnels and urban mobility, including tolling, traffic management, and smart mobility. With a strong presence in the United States and Portuguese markets and growing in Europe, ATOBE has deployed extensive roadside, central management infrastructure and end-user facing tools such as back-office solutions and mobility Apps, validated in interoperability scenarios.

ATOBE has a significative commitment with innovation and has been investing strategically in several key areas related to the future of the mobility sector. With more than 15 years of experience in Cooperative Connected Autonomous Mobility (CCAM) and C-ITS, and participation in pivotal funded projects such as C-ROADS and 5G-MOBIX, ATOBE has progressed from prototypes and pilots to production-ready CCAM solutions. At the same time, the company continues to anticipate the growing role of autonomous vehicles and recognises the importance of convenient payments and service activation in connected vehicles. With a strong track record of more than 1.500 km of motorways, tunnels and bridges managed with its traffic management solution (Atlas), ATOBE has been exploring, namely in the funded TANGENT project, how to scale from vertical segmented traffic management solutions into holistic cooperative traffic management, allowing agencies, operators and other key stakeholders to work together in strategic, tactic and operational levels. Finally, ATOBE has been actively investing in innovation to develop next-generation tolling solutions, leveraging both its extensive experience and continuous R&D efforts. Through previous and ongoing projects, ATOBE is integrating advanced sensing technologies — including computer vision, 3D LiDAR, and other smart detection systems — to enhance its existing solutions and expand its portfolio with novel, high-performance tolling capabilities. With more than two decades of expertise in integrating computer vision and intelligent sensing, ATOBE has continuously evolved its products, introducing cutting-edge solutions such as automated vehicle classification and in-vehicle passenger detection, reinforcing its position as a leader in innovative mobility technologies.

In the MODALSHIFT project ATOBE will:

- Provide and configure a web-based platform for Tactical and Operation Management of Transportation, offering monitoring, cooperation and incident management features
- Develop a Cooperative Multimodal Traffic Management System (MTMS) that supports the following Use Cases:
    - Trieste – (section 4.1.1.2) Provide a unified solution where the live port situation can be easily understood and acted upon and where forecasts can help prepare for the near future.
    - Varna – (2nd Use Case – section 4.1.2.2) Provide a unified solution where the situation of the entire city can be easily understood and acted upon and where forecasts can help prepare for the near future, offering the possibility to focus on specific areas and to manage planned and unplanned situations.
    - Madrid – (2nd Use Case – section 4.1.3.2) Provide a unified solution where the context of the Estación Sur (South Station) can be evaluated with relevant data and where forecasts can help prepare for the near future, offering the possibility to manage planned and unplanned situations.
- Integrate with the technical partners from WP3 to provide a user interface for any Tactical and Operational Management features, extending the available information with forecasts and other added value services like optimisation. Optimisation can be employed in Incident Management within the platform developed by ATOBE.
- Integrate with the other technical partners from WP4 to integrate live data from other services that is relevant for the cooperative MTMS.
- Provide training and support to Case Study users.

ATOBE brings to the MODALSHIFT project the features of the funded TANGENT project Dashboard allied with the more operational features of ATOBE's Atlas Traffic Management product.

The TANGENT Dashboard is focused on the intuitive visualisation of transport data, offering a fully customizable multi-tenant dashboard where allowed users can easily configure visualisations for existing and new data sources.

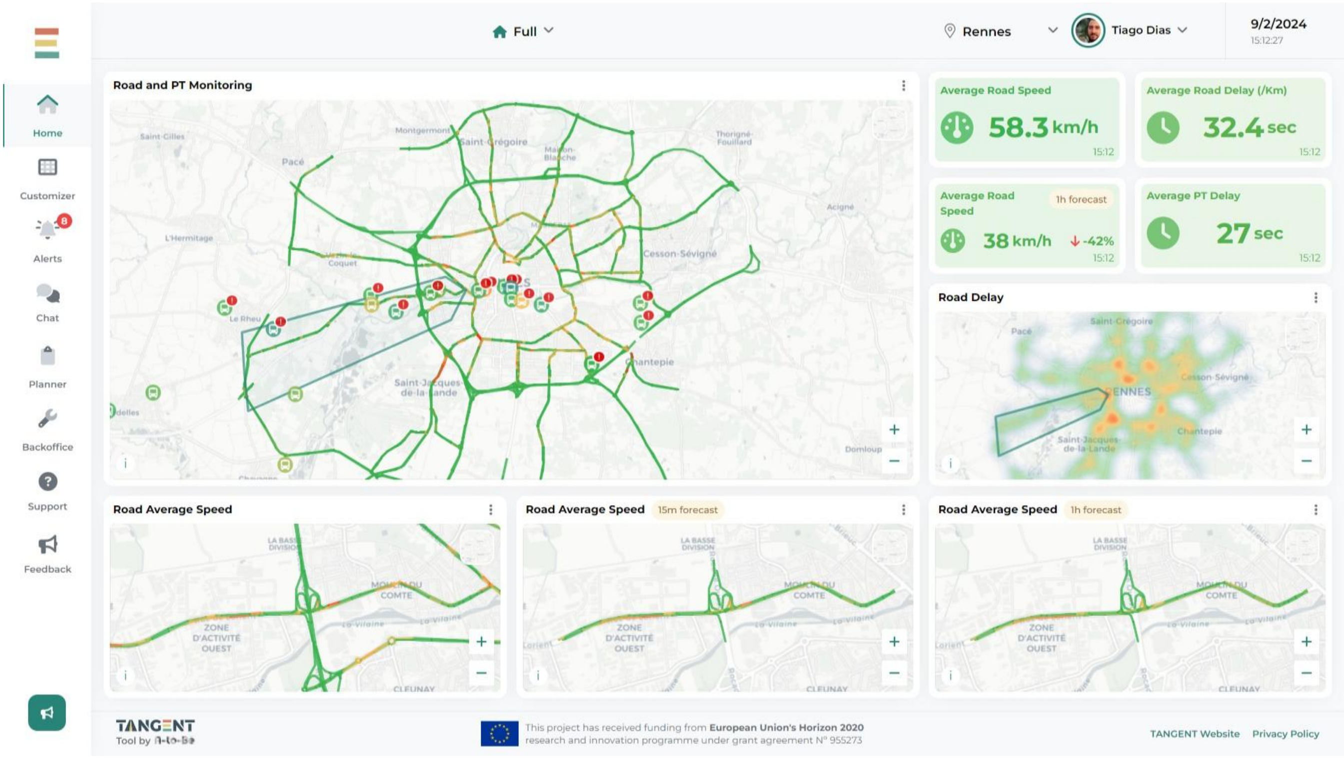


Figure 5 TANGENT Dashboard screenshot with the city of Rennes (France)

The TANGENT Dashboard shows live and forecast data in either Map or Key Performance Indicator (KPI) widgets. Map widgets can combine several different data sources into a single display while KPI widgets focus on a single aggregated measurement. Colours and symbols are widely used to intuitively depict the network status. The generic nature of the technology behind the TANGENT Dashboard is not limited to transportation data and can easily scale to other kinds of data from other domains. Even in transportation it is frequently necessary to display related data from other domains.

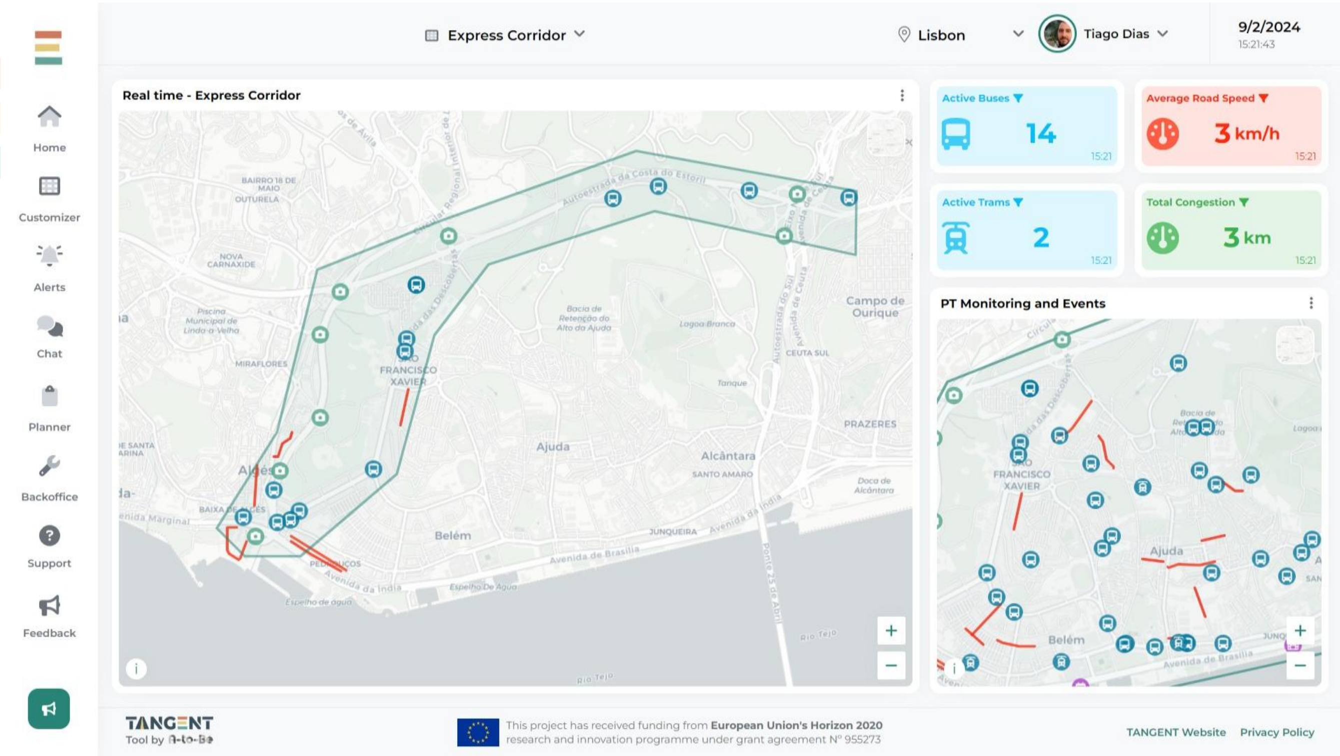


Figure 6 Express Corridor, Lisbon (Portugal)

In TANGENT it is possible to filter data and focus widgets in specific geographic areas or corridors, both for maps and KPIs, allowing for the easy creation of more specialized and focused dashboards, tailored, namely, for Incident Management.

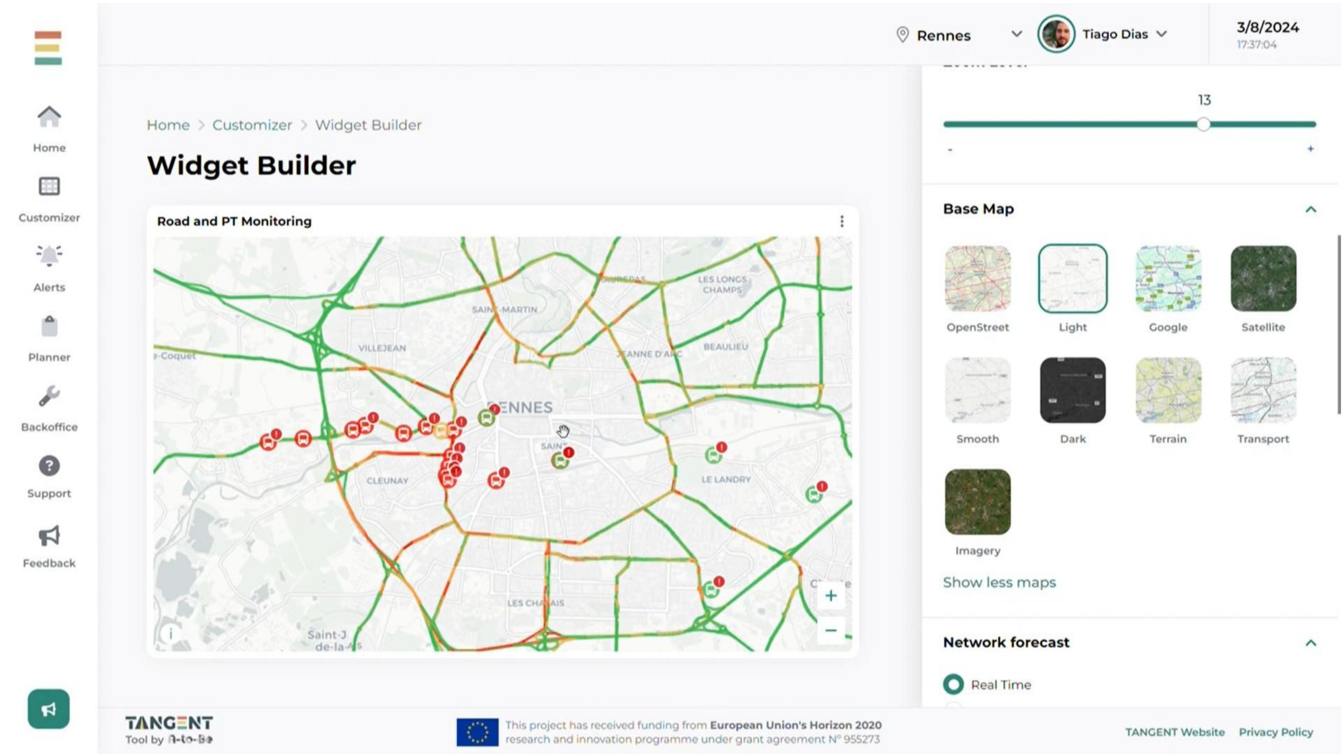


Figure 7 Map widget customization

By being able to easily add new data sources and visually configure them, the TANGENT dashboard fully empowers agencies, operators and emergency services.

A key feature in the TANGENT Dashboard is the possibility to prepare Operational and Tactical Response Plans, for example to cope with a football match or with a flood situation. These Response Plans use Simulation and/or AI technology to evaluate the key situation and provide an optimized action plan that can be triggered. Actions can include changing scheduling or capacity of transport network elements, changing traffic light control plans or other relevant actions over elements in the transport network.

These plans can be created and managed cooperatively by multiple stakeholders. Cooperation is also supported by built-in chat features in TANGENT.

These Response Plans can then be monitored automatically using KPIs defined from the available data (for e.g. the current travel time for a route or measured water levels) and are able to trigger an Incident once the Response Plan needs to be activated. This initiates Incident Management which allows the activation of the action plan and uses the TANGENT Dashboard visualisation features to allow operators to monitor the resolution of the situation, either for an entire city or focused on a specific area, depending on the incident.

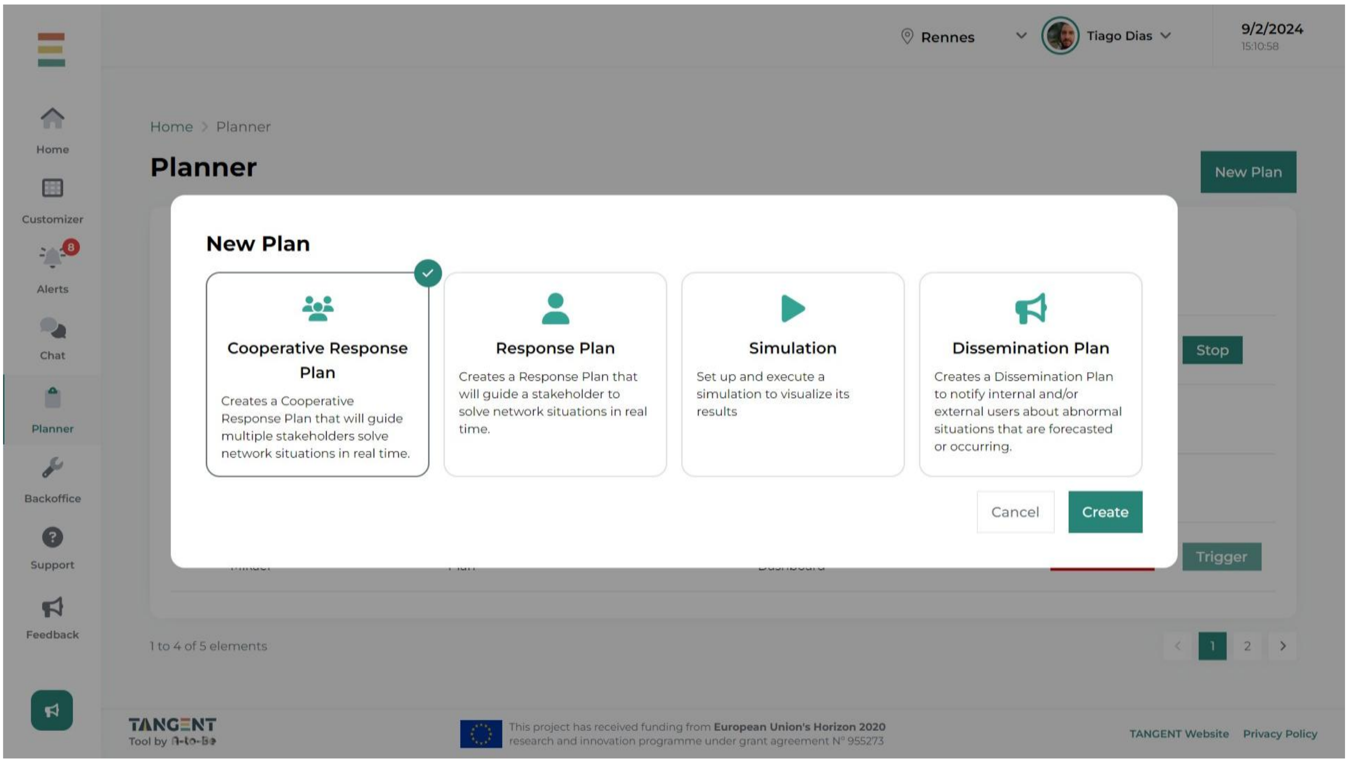


Figure 8 TANGENT supports several types of Plans

The TANGENT Dashboard features will be integrated and extended inside the Atlas product which also adds several other possible features that may become available to the case studies:

- Video camera visualisation and control

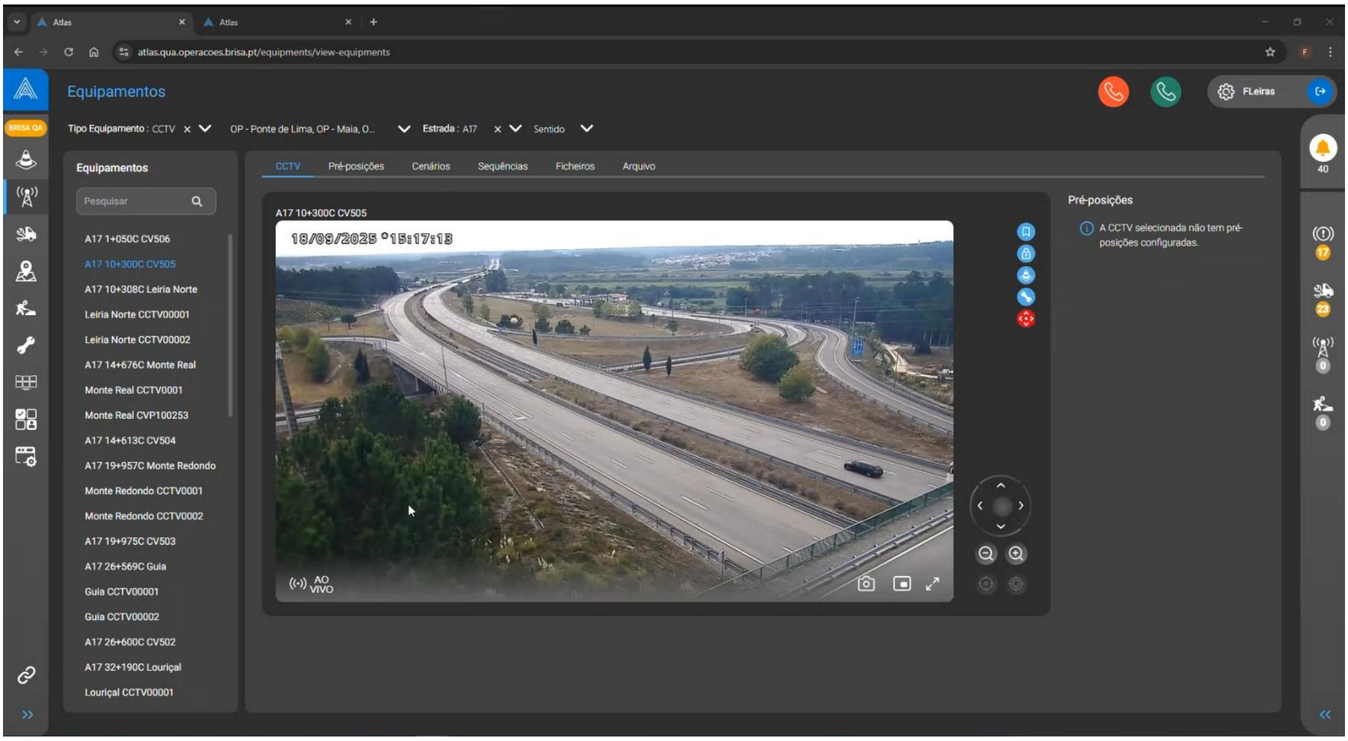


Figure 9 Camera remote view

- Integration with message panels and other operations devices
- Incident Management and Response

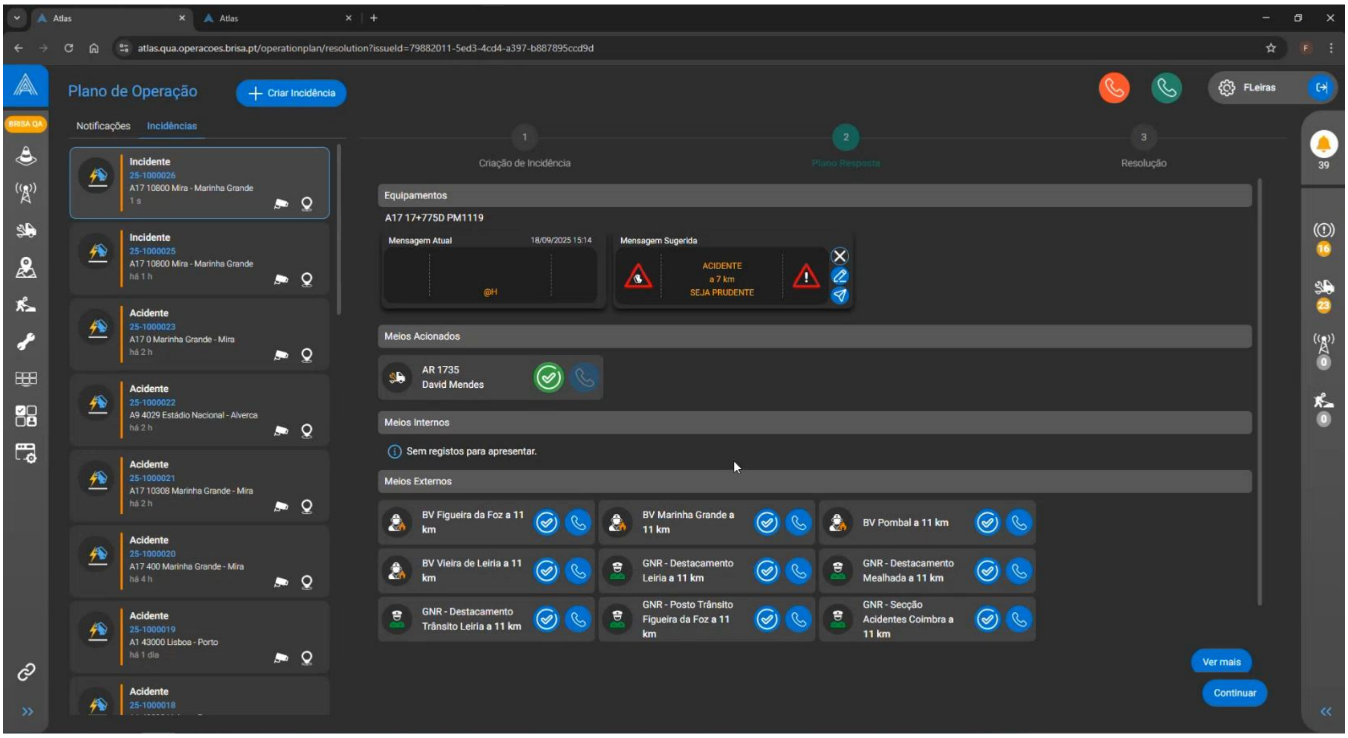


Figure 10 Incident management

4.2.3.2 GeneGIS GI – GGI

GeneGIS GI (GGI) is a Geo-ICT specialist born by merge of different Geo-Spatial-focused companies, the main of which (GESP S.R.L.) was funded in 1977. Since the early 80s, GGI offers a full range of services to follow the entire lifecycle of projects based on geographical data, from acquisition of the raw data, processing and integration, through the development of bespoke software applications.

With more than 100 professionals, 40 software developers, 20 years of experience, services in over 20 countries, GGI is the most complete Geo-ICT solution provider in Italy, with the most significant international experience.

GI stands for Geographical Intelligence and highlights the company's strong focus on an area of technical knowledge based on full understanding of geographical data and involving map-based technologies, area and object monitoring, surveying with drones and satellites, logistic tracking and tracing.

Geographical Intelligence embraces a massive range of requirements: from logistics to infrastructure management, commercial planning to environmental conservation, traffic flow management to socio-economic planning.

GGI works in several areas, included the following:

- **Energy and utilities:** develop complete solutions for network management and implementation, measurement of consumption, operation monitoring.
- **Telecommunications:** manage complex nodal networks, for fibre optics, with contained migration activities and guaranteed territorial continuity.
- **Logistics and Transport:** develop solutions to monitor and improve the efficiency of transport networks by optimising travel and goods handling times.
- **Environment and Natural Resources:** develop solutions to monitor and analyse complex environmental data, for the purposes of planning, control and safety.
- **Public Authority:** provide technological services for management of geographical information, from the Land Register/Urban Plans to traffic flow planning.
- **Cultural Heritage:** apply Geographical Intelligence to projects for the conservation and enhancement of artistic and cultural assets.

In the MODALSHIFT project, GGI will provide three main ICT tools to support the use-cases:

1. In the **Port of Trieste**, will be implemented a system to improve and automatise the communication about delays and service disruptions.
2. In **Varna,** an existing Track&Trace&Monitoring (T&T&M) platform will be expanded and enriched to support vehicles reservation according to the current state of the use-case.
3. In **Madrid** GGI will contribute developing an app to help the user get most of the data computed and elaborated by the algorithms developed in WP3.

More in detail, the T&T&M platform provides comprehensive fleet management, linking each vehicle to its driver, the installed tracking device, and optional probes that measure external parameters such as temperature, humidity, pollutants, battery level, and many others.

Data are captured in real time, so the platform continuously displays up-to-date information. Communication with the tracking devices uses the standard GS1 EPCIS 2.0 format, ensuring maximum compatibility with both proprietary and third-party devices.

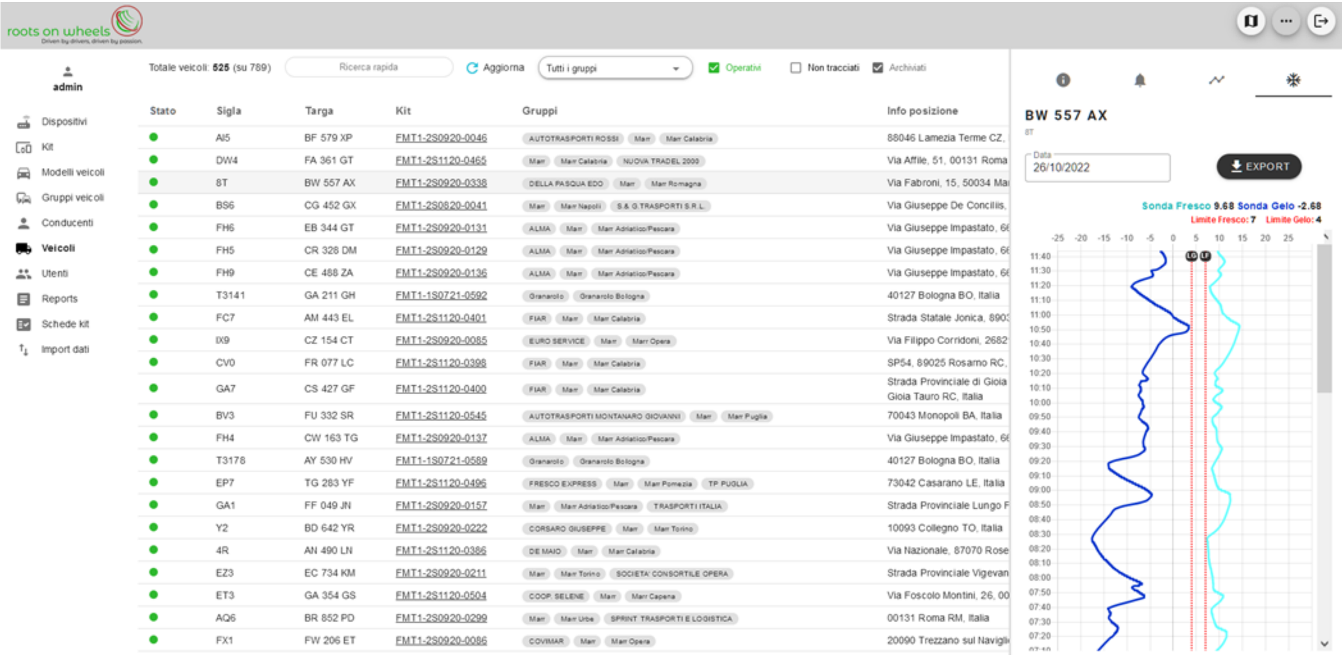


Figure 11 Fleet management

A map complements the management interface, visually displaying each vehicle's location updated in real time, along with its current status and recent events.

From the map you can view the entire fleet at once or zoom in on a single vehicle, tracking its route and activating the "follow-me" function.

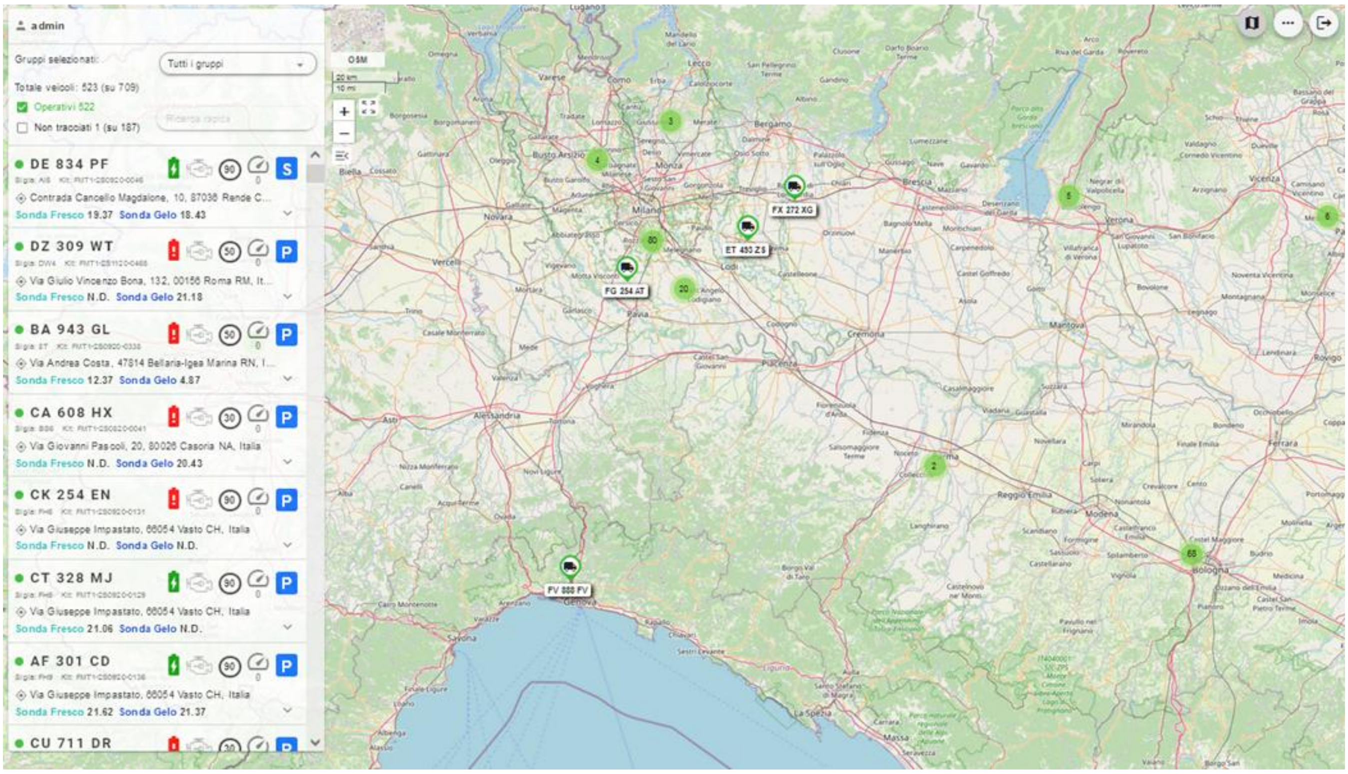


Figure 12 Live tracking

Beyond real-time data, both the administrative dashboard and the map retain the full historical record, allowing you to scroll back in time and review exactly what occurred at any given moment for each monitored vehicle.

Data from probes can also be exported in PDF and CSV format and can be presented as a report or to perform further offline analysis.

4.2.3.3 Tellae

TELLAE is an innovative consulting firm specialized in cities transportation issues. TELLAE provide its clients, both public and private, with solutions to support their strategic and operational decisions.

TELLAE was founded in 2020 with the objective to spread robust scientific approach to transportation decision making. Here is a non-exhaustive list of use cases which are tackled by TELLAE:

- Optimizing investment choices
- Ensuring a reliable development strategy
- Designing and optimizing mobility services
- Simulating and analysing the impact of a new mobility policy
- Conducting a customized study in response to a specific issue
- Improving the robustness of mobility studies

To fulfil its engagements, TELLAE is developing several tools for its internal use and also for its customers.

Starling is an agent-based framework for urban mobility. It enables modelling technicians to easily build a digital twin of a mobility service or of several interconnected mobility services. Focusing on mobility services, it offers functionalities for very detailed simulation, at individuals and vehicles level, enabling strong design evaluation and optimisation of allocated resources.

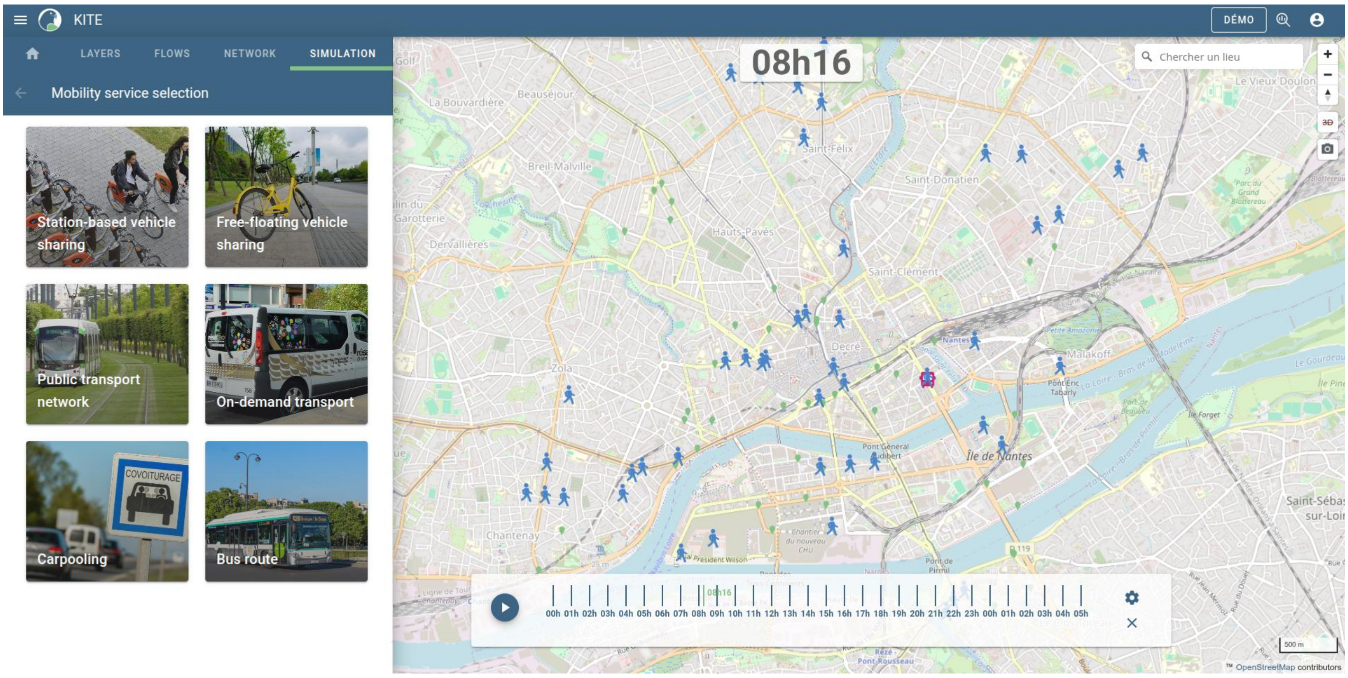


Figure 13 KITE simulation

KITE (https://kite.tellae.fr/) is a SaaS mobility analysis and simulation platform. It is designed for professionals to easily access data territorial with analysis and visualisation functionalities. It also provides simulation functionalities to design and optimize mobility offers and services.

- Layers: easy access to a rich catalogue of data for building maps and analysing an area;
- Flows: advanced functionalities for easily analysing flows;
- Networks: exploration of public transport networks and isochrone analysis;
- Simulation: simulation use case of mobility services.

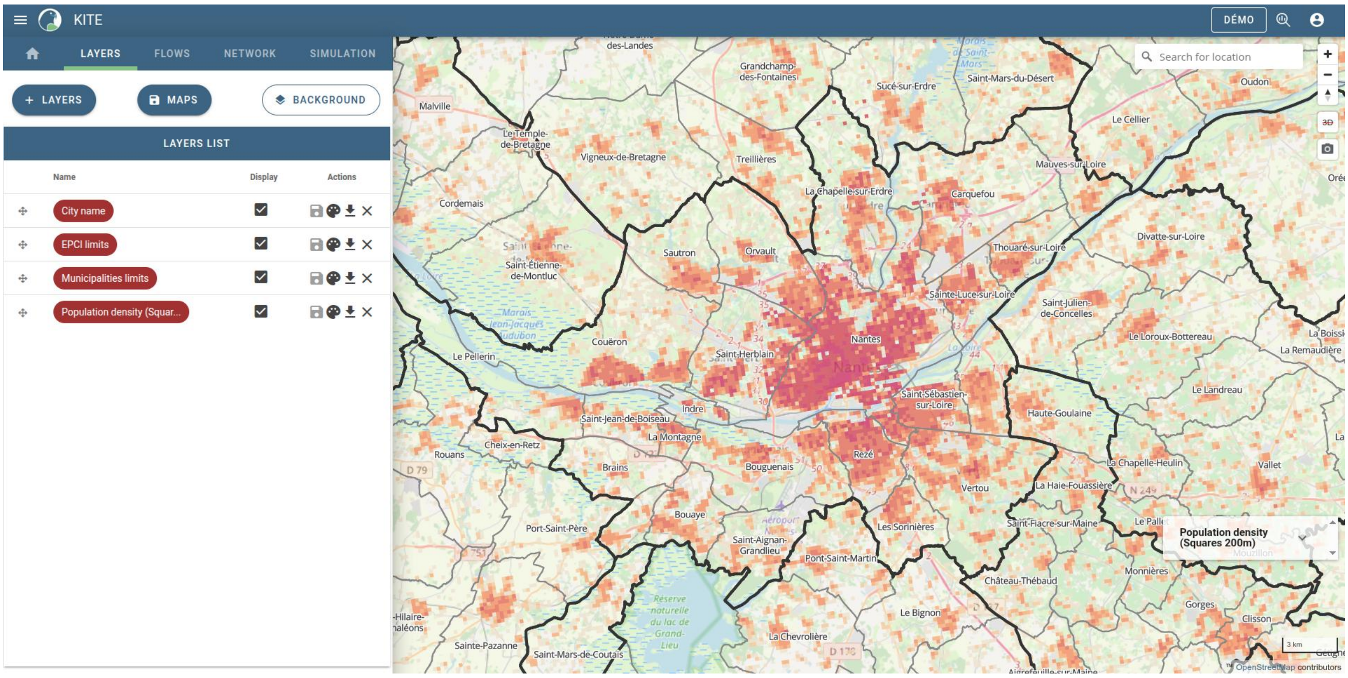


Figure 14 KITE GIS view

A first basic version of a **QGIS plugin** (https://plugins.qgis.org/plugins/tellae/) has recently been introduced in order to give access to KITE functionalities directly from the QGIS Open-Source Software.

In relation with the simulation issues and flows analysis, TELLAE is fully engaged in the contribution of the Open-Source software **EQASIM** (https://eqasim.org/). This software allows the development of synthetic population of trips, using Open Data and mobility surveys, by applying data fusion algorithms.

4.2.3.4 Octopize Mimethik Data - Big Data Santé – OCTO

OCTO specializes in data privacy and data augmentation. Through its avatar solution, OCTO provides anonymization solutions to enable secondary usage of data in full compliance with regulatory frameworks such as GDPR and AI Act. The expertise provided by OCTO spans beyond anonymization and includes privacy assessment, data augmentation and data desensitization.

In parallel to anonymization algorithms, OCTO develops a set of privacy metrics that can be used to evaluate the anonymous character of transformed data. While originally aimed at evaluating the output of the avatar solution, the privacy metrics are compatible with other methods and so privacy audits can be carried out. The set of metrics used for this purpose is directly related to the 3 GDPR criteria: singling-out, linkability and inference.

The core Avatar algorithm represents individuals to anonymize in a space of numerical coordinates and generates new coordinates for each individual. Those coordinates are generated by mixing characteristics of neighbouring individuals. While this process is designed with privacy in mind, it globally generates new individuals that share many similarities with originals, a property compatible with the need of data augmentation. OCTO offers several data augmentation functionalities via its software:

- Pure data augmentation where the aim is to produce a dataset containing more individuals that the original data. This augmentation can offer if required some privacy guarantees (anonymous augmented data).
- Class balancing where under-represented classes can be balanced to reduce biases and increase quality of subsequent models.

As detailed in the privacy guidelines section of the present document (Chapter 6), GDPR only applies to personal data. However, other barriers may exist that prevent data sharing or data usage. Industrial or commercial sensitivity is one of them and involves dataset that could provide competitors an advantage if shared by revealing some sensitive information. To be shared, this data needs to be transformed and the sensitive aspects removed or reduced to an acceptable level. Criteria from GDPR can be used as indicators but tailored metrics can be required. The mechanisms used for anonymization can be used for data desensitization.

In the MODALSHIFT context, data desensitization represents a lever for some industrial data providers to make data available.

The avatar solution is available to all MODALSHIFT partners via its SaaS architecture. Users can either setup anonymization jobs using a web or a python client. By having the two options, different user profiles can gain access to anonymization. The web interface is designed with the aim to democratize anonymization for all and ensure non-technical profiles can benefit from the solution. The web interface is available in all languages represented by the MODALSHIFT consortium.

Figure 15 Avatar landing page

The overall process of setting up and running an anonymization job goes as follow.

1. Context setup. Some context is given about the future use of the data. This step enables the automatic reports to be tailored to the need of the user and in particular the Privacy Impact Assessment (PIA) report.
2. Data loading and setup. The data that can consists in one or many tables is loaded. Variables are detected and some actions can be performed on them such as freezing or excluding some (with some privacy risks).
3. Data relationship setup. When more than one table is provided, the data must represent a relational database. Links between tables need to be specified.
4. Anonymization setup. The anonymization process can be parameterized to explore different options and to define the level of privacy to expect. Augmentation can also be setup in this step. Note that a default parametrization is provided based on the data provided.

5. Configuration overview and export. In this stage the configuration can be reviewed prior to a run and can be exported to be run from the python client. This step process useful to move from an exploration to an industrialization phase that would typically run without the web interface.
6. Results analysis. Once complete, an overview of the generated data is provided with privacy and utility metrics. Meta-metrics are provided for a quick understanding of the job results and more detailed metrics available if required for a fine-grained analysis of the anonymization. From this page, the avatars can be downloaded along with the technical and PIA reports.

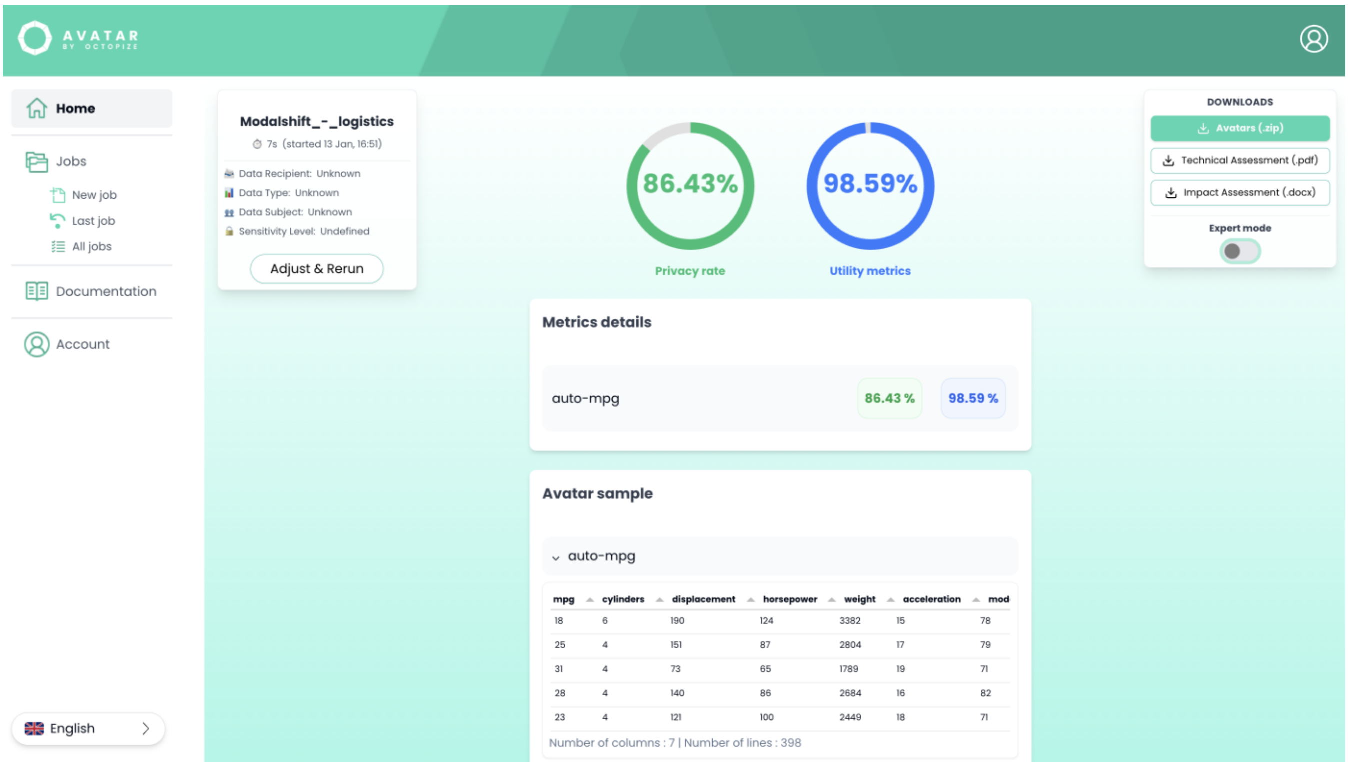


Figure 16 Avatar result page

In Chapter 6 there is a more detailed description of anonymization, the principles behind Avatar and how to apply it in MODALSHIFT.

4.3 GAPS

During the definition of the case studies and the involvement of stakeholders, it emerged that the issues affecting the project’s case-study cities are aligned with those faced by many cities across Europe and currently being addressed by the European Commission. The increasing complexity of freight and people transport, combined with sustainability, resilience and competitiveness objectives, highlights the need for a truly integrated multimodal logistics ecosystem. While significant progress has been made at modal level, important gaps remain that hamper multimodal transport from reaching its full potential. These gaps are particularly evident in data sharing and integration between players, infrastructure evolution to offer current level of services and be future-ready, and the emerging role of railways in both people and freight transport.

At present, data within the multimodal logistics chain is highly fragmented and poorly shared. Each actor, including infrastructure managers, rail operators, terminal operators, logistics service providers and shippers, typically manages its own data in isolated systems. Data exchange, where it exists, is often limited to bilateral arrangements, relies on heterogeneous formats, and is frequently manual or semi-automated. This fragmentation limits visibility across the transport chain and hinders coordinated planning and decision-making.

The target state for a European multimodal logistics system is one in which data can be shared seamlessly and securely across all relevant actors, enabling end-to-end visibility and more efficient operations. Achieving this requires not only technical interoperability but also trust. In this context, the absence of a neutral, fair and trusted third-party entity represents a major gap. Such an entity could act as a data governance and orchestration layer, ensuring transparency, data sovereignty, and compliance with EU regulatory and competition requirements, while encouraging participation from all stakeholders.

Without a common data governance framework and trusted data-sharing mechanisms, stakeholders remain reluctant to share information, leading to suboptimal use of infrastructure, reduced resilience to disruptions, and limited ability to innovate or scale multimodal services.

Multimodal transport relies on the close interaction between physical infrastructure and digital services. However, current infrastructure, both physical and digital, has largely evolved in a mode-specific and fragmented manner. Many services are tightly coupled to existing infrastructure and legacy systems, making them difficult to adapt as infrastructure evolves or as new operational requirements emerge.

The desired future state is one where infrastructure and services evolve together. Digital services should be modular and scalable, capable of supporting increasing volumes, new actors, additional transport corridors, and emerging technologies. At the same time, infrastructure upgrades should be designed with interoperability and digital readiness in mind, enabling seamless integration across modes and borders.

The gap lies in the misalignment between infrastructure development and service design. Insufficient emphasis has been placed on scalable digital architectures that can grow incrementally, rather than requiring costly and disruptive system replacements. This limits the ability of multimodal solutions to respond to changing market conditions, policy objectives, and user expectations, ultimately reducing their competitiveness compared to more flexible, single-mode alternatives.

Railways play a crucial role in achieving the European Union's objectives for sustainable and efficient freight transport. As a high-capacity, low-emission mode, rail is a cornerstone of multimodal transport chains. Despite this, rail freight often remains insufficiently integrated into multimodal logistics systems, particularly in terms of digital visibility and traceability.

Today, rail freight tracking and tracing capabilities are uneven across Europe and across operators. Information on train position, wagon status, or cargo condition is frequently delayed, incomplete, or not easily shared with other modes. This lack of reliable, real-time traceability undermines confidence in rail-based multimodal solutions and makes it difficult to coordinate rail operations with road, maritime or inland waterway transport.

The target state is the full end-to-end traceability of rail freight as an integral part of multimodal visibility platforms. This requires the wider deployment of digital tracking technologies, harmonised data standards, and better integration between rail operators' systems and multimodal data platforms. The current gap in rail freight traceability represents a critical barrier to improving reliability, resilience and customer confidence in multimodal transport solutions.

Within this context, the project assessed partners' technical capacities to deliver and further develop the tools required for deploying use cases that address the challenges identified in the case studies.

A comparison of baseline challenges, business needs, and potential use cases against existing technical capacities was conducted to determine the need for complementary solutions and their alignment with the project scope. This analysis was informed by targeted meetings with technical partners and potential stakeholders in the three CSs.

To address all the requirements emerging from the case studies effectively and efficiently, eleven services have been identified in Task 1.3, labelled from A to K. These services have been defined to ensure comprehensive coverage of the functional and operational needs highlighted throughout the analysis.

Within the scope of MODALSHIFT, a service is understood as a functional unit that is directly accessible to the end user and designed to perform a specific task. By contrast, a subservice is an independent building block that supports one or more services. Subservices are not necessarily exposed to the end user, but instead operate as modular components that enable, enhance, or extend the functionality of the main services.

This structured organisation ensures that the identified use cases are addressed precisely and systematically. By leveraging the modular nature of the consortium's contributions, the architecture promotes reusability, flexibility, and seamless integration across all developed components. Such an approach not only facilitates efficient development and deployment but also supports scalability and future extensibility of the system.

A more detailed description of each service and subservices is provided in chapter 6 of MODALSHIFT Deliverable 1.2 *Use Cases and strategies for social optimum*.

The high-level outcomes of the gap analysis are summarised in the table below.

	Service	Trieste	Varna	Madrid
A	E-subscription device for public transport			✓
B	T&T&M cargo-bike reservation tool		✓	
C	Dynamic fleet management and drivers' traffic optimisation tool		✓	✓
D	Road & Rail traffic forecasting tool	✓		✓
E	Smart lockers & boxes Capacity-as-a-Service			✓
F	Multimodal, trust-by-design & secure data space	✓	✓	✓
G	Agent-based transport modelling tool	✓	✓	✓
H	Transport planning, what-if & visualisation tool	✓	✓	✓
I	Cooperative Multimodal Traffic Management System	✓	✓	✓
J	Shunting operations	✓		
K	Anonymisation	✓	✓	✓

Table 2: Link services to case-studies

5 TECHNICAL DESCRIPTION

This chapter introduces a high-level integrated system functional architecture laying out the collaboration principles among partners, the core modules planned for implementation and detailed specifications for the various digital services and hardware solutions.

5.1 ARCHITECTURE

This section describes the basis of the architecture which will be adopted in this project and aims to propose a reference architecture for projects managing multimodal transport and shifting drivers and passengers to more optimized modes.

The projects' functional architecture is broken down into four integrating layers. The Data Layer collects data and distributes it within the project partners; it also acts as the integration mechanisms between the different modules of the project. The Optimisation Layer uses data to produce forecasts, analysis, hypothesis studies and recommendations. The integration of data and the functionalities of the Optimisation Layer into the Applications Layer allows the project to implement a set of use cases depicted in the Use Cases Layer.

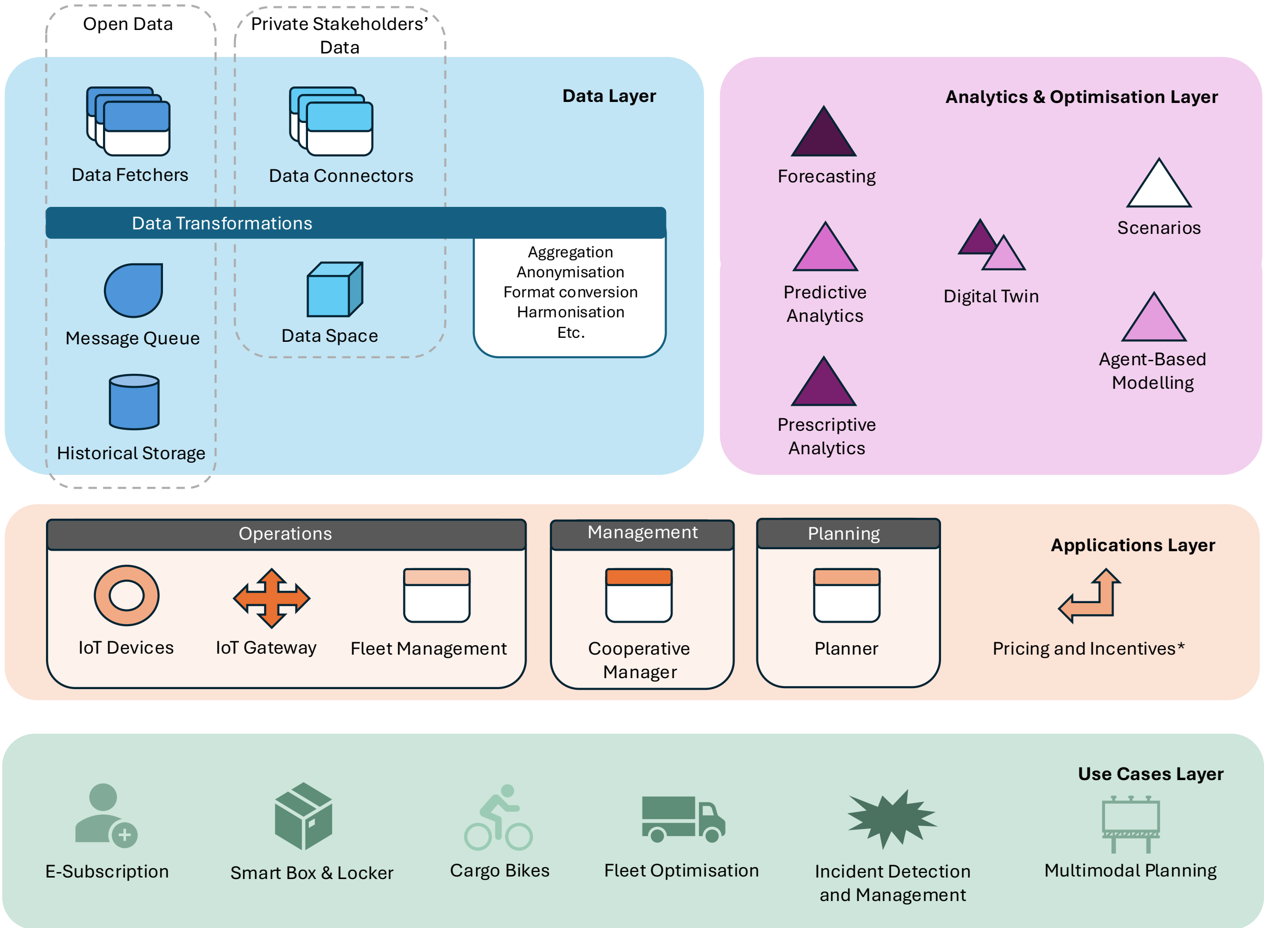


Figure 17 MODALSHIFT functional architecture

In Figure 17, no interactions between modules are depicted to avoid over-complexity. Generally, the Optimisation and Applications Layers interact with the Data Layer to collect and exchange data. Modules within these two layers also interact between them using the Message Queue of the Data Layer. The Use Cases Layer is conceptual, providing a key functional view of the implementation and each use case functionality is implemented by a combination of applications.

5.1.1 DATA LAYER

The Data Layer is responsible for collecting data from external systems, namely open data sources and private stakeholders' data. Two different data sharing mechanisms are planned for this layer: a Data Space (for private data) and a Message Queue (for open data).

Different formats and data organisation will exist in the fetched data sources. Data transformations in the Data Layer (such as aggregation, anonymisation, format conversion, and harmonisation) are essential to ensure that diverse data sources can be integrated, securely shared, and efficiently processed across project partners. These processes enable consistent, privacy-compliant, and interoperable datasets, forming the foundation for reliable analytics and decision-making.

Instead of having different partners fetching and repeating the transformation process on their own, the project will adopt a single transformation pipeline for each data source, sharing the resulting data through the two main mechanisms. A Data Space will be implemented to provide adequate data management and governance for private stakeholders' data. The Data Space will catalogue and manage access to data, ensuring that data does not leave its owners, guaranteeing full sovereignty over any shared data. Open data, in contrast, does not require these kinds of measures, any partner could fetch it, except this would be duplicating work, especially when each partner might have to transform this data, converting its format, aggregating it and anonymizing it. To avoid this, a Message Queue was selected as a complementary data sharing mechanism to the project's Data Space. The Message Queue will receive all fetched and transformed open data in the project. It will also be used to communicate between the project modules, namely, to activate models in the Optimisation from Applications and to receive their results. Live and processed data can be fetched directly by subscribing to Message Queue topics, whereas Historical data will be stored in an auxiliary filesystem. Message Queue topic naming and historical data file naming will follow project-wide definitions. Access to the data will be controlled through authentication and access control lists. Together, these measures provide an adequate level of data governance, although still below that of the Data Space.

The Data Space can be a sophisticated solution, requiring the definition of many aspects regarding the data handled. Even the initial step of injecting new data items into an existing Data Space can be challenging. The technical integration process is demanding and a lot of elements need to be defined regarding the data characteristics, namely metadata models, data schemas, semantic interoperability, identity management, contractual terms, consent management and usage control, in order to preserve data sovereignty. This is unnecessary for Open data which is free from access control mechanisms and is usually catalogued in a rather ad-hoc manner, using web documentation, examples and sometimes REST APIs. This justified having a simpler, more ad-hoc, Message Queue mechanism, which still offers access control and many other security features, but is intrinsically simpler than the Data Space. The Message Queue will also act as an accelerator for private data sharing within the project from an early stage when the Data Space will not yet be available due to the complexity of its initial definition and setup. Data sources initially shared with the Message Queue will, in time, be migrated into the Data Space, achieving a higher maturity in data governance involved.

5.1.2 OPTIMISATION LAYER

The Optimisation Layer offers several distinct forecasting, analytic and simulation models that use specific data elements from the Data Layer to produce their results. Most models require training; this is usually done using representative sets of historical data. To cater to this need, the project is preparing to start collecting historical data from an early stage. When data sources only contain live data (very common for most open data sources), the Data Layer will implement a storage mechanism that saves all fetched live data in the Historical Storage filesystem.

The Optimisation Layer will include forecasting models which use historical and real-time data, graph neural networks, and transfer learning to predict traffic flow and detect anomalies in transport systems. On the simulation side, this layer will include agent-based, activity-based, and digital twin models that enable detailed simulations of transport systems, scenario analysis and the evaluation of mobility and logistics strategies. These models can also implement optimisations by simulating various scenarios and adjusting parameters to identify the most effective transport strategies or configurations. Through dedicated optimisation and heuristic models, this layer will develop these methods based on scenarios and data created by simulation and forecasting techniques and find direct optimal (sometimes best) solutions/decisions for complex planning problems, such as dynamic route planning and capacity booking, quickly generating feasible and efficient solutions under operational constraints.

The Optimisation Layer will implement digital twins for some of the use cases. By processing real-time data and providing both forecasts and scenario-based results, the modules in this layer implement a typical digital twin architecture where the real-time status of a digital twin is extended with future states predictions and scenarios can be explored within the digital twin to determine the impact of specific optimisation measures. The Message Broker can act as the element holding both the real-time state of the digital twin and the produced data for forecasts and the results of simulating specific scenarios. Digital Twins typically have message queues at their core for state representation and exchange, and this fits perfectly with the proposed architecture for MODALSHIFT.

5.1.3 APPLICATION AND USE CASES LAYERS

Coming to the Application Layer there is an Operations group where IoT devices enable E-Subscription and Smart Box & Locker use cases driven by an IoT Gateway which integrates the devices with the platform using the services of the Data Layer. The E-Subscription use case will be a low-power IoT device that can be used by vulnerable users. It can provide a seamless onboarding and offboarding experience in public transport and can track these users and warn them and other transportation elements about safety risks. This will not only ease the life of vulnerable users but can also save their lives in situations like wheelchair users moving around stopped buses, where they can be totally invisible to drivers. The Smart Box & Locker uses other types of IoT devices to implement Capacity-as-a-Service for small scale freight using available capacity in buses.

The Fleet Management application in the Operations group will implement the Cargo Bikes and Fleet Optimisation use cases. The Fleet Optimisation will be implemented in the shunting activities in the Port of Trieste where it will improve delays and service disruptions and automate their communication. This functionality may be further supported, subject to ongoing technical and operational assessments, by the deployment of NGS IoT devices on ADF locomotives to enable enhanced real-time monitoring and data integration.

In all previous use cases, which are mostly operational, by integrating with the Data Layer, the IoT devices and Fleet Management application, will also provide enhanced situational awareness and multimodal data integration across the transport ecosystem by means of the Management and Planning components of the Applications Layer.

The Cooperative Management application will provide live and forecasted situational awareness in all the case studies, spanning all use cases. It will also include cooperative features like chat and cooperative incident management, where, building together with functionalities from the Optimisation Layer, it will be possible to interface with stakeholders from the case studies and interactively prepare and activate optimisation plans. The live and forecast situational awareness functionalities will be used to monitor the results of activating optimisation plans.

The Planner application will be the main interface with the Simulation and Scenario tools within the Optimisation Layer. It will bring tools for transport planners to enhance decision-making: from data analysis to what-if simulation, giving access to easy-to-use agent-based-modelling simulation, providing both simplified and more sophisticated simulation results.

Finally, the MODALSHIFT platform also plans for a Pricing and Incentives application which will be conceptually explored and planned for in the project, although there is no implementation or planned trial for it. This module has a key role in adopting project results at a later stage. While many optimisations can rely on coordinating services, changing service scheduling or capacity and conveying recommendations to passengers and drivers, the most impactful changes may only be achievable by changing behaviours outside of the operational, tactical and strategic control of the operators. These changes will require not only the optimisation tools but also mechanisms to influence passenger and drivers' options. This Pricing and Incentives application brings exactly that mechanism, bridging optimisation modules to actual pricing and incentive mechanisms applied to users, through apps, dynamic pricing or other adequate means.

5.2 CORE MODULES

In order to analyse each system component in more detail and present prioritised requirements the project will consider the following core modules associated with the Services presented in Deliverable 1.2 Section 6.

Module Name	Description	Services Involved
Vulnerable Users Monitoring & Safety Module	Integrated OBU/Gateway, proximity transceivers, and Bluetooth beacons enabling presence detection, alighting monitoring, and proactive V2X safety alerts for vulnerable users.	A: E-subscription device for public transport

Smart Box & Smart Locker Module	Smart Box and Smart Locker infrastructure for parcel tracking, access control, and Capacity-as-a-Service implementation in public transport.	E: Smart lockers & boxes Capacity-as-a-service
Cargo Bikes Monitoring Module	IoT-based tracking and monitoring system for cargo bikes supporting booking, usage analysis, and urban logistics optimisation.	B: T&T&M cargo-bike reservation tool
Shunting Locomotives Monitoring Module	IoT-based monitoring infrastructure for locomotives supporting real-time positioning, utilisation analysis, and integration with shunting management systems.	J: Shunting operations
Fleet Management & Optimisation Module	Centralised fleet management, tracking, and optimisation tools for cargo bikes, locomotives, and operational assets.	B: T&T&M cargo-bike reservation tool C: Dynamic fleet management and drivers' traffic optimisation tool J: Shunting operations
Cooperative Management Module	Unified cooperative Operational Management platform bringing multimodal data together.	I: Cooperative Multimodal Traffic Management System
Planning Module	Transport planning tool with visualisation, what-if analysis and agent-based modelling.	H: Transport planning, what-if & visualisation tool
Transport Modelling Module	Agent-based simulation models to evaluate different what-if scenarios.	G: Agent-based transport modelling tool
Data Integration & Standardisation Module	Data harmonisation, validation, anonymization and conversion into interoperable formats (e.g. GS1 EPCIS) for ingestion into the Data Layer.	A: E-subscription device for public transport B: T&T&M cargo-bike reservation tool E: Smart lockers & boxes Capacity-as-a-service J: Shunting operations K: Anonymization
Forecasting Module	Road and rail forecasting to be used by other modules to improve simulations and other tools.	D: Road & Rail traffic forecasting tool
Data Space & Message Queue Module	Secure data sharing, governance, access control, and distribution mechanisms for private and open data.	F: Multimodal, trust-by-design & secure data space And all the other services

Table 3: Core Modules in MODALSHIFT

5.3 SPECIFICATIONS

In this section each core module is described in more detail and requirements are presented and prioritised according to the MoSCoW analysis approach.

5.3.1 SHARED CARGO-BIKES SERVICE – GGI

A shared cargo-bike service, when offered by a municipality, can provide significant benefits to both the city and its population. By facilitating the transition from fossil fuel-based logistics to lower-emission alternatives, the service helps reduce air pollution and improve the overall quality of urban life. It also alleviates traffic congestion in city centres and Low Emission Zones (LEZ) by replacing traditional delivery vans with compact, eco-friendly cargo bikes. In addition to these environmental advantages, the service supports the local economy by offering small and medium-sized businesses a cost-effective and sustainable means of competing with larger players in the market.

The software developed by GGI builds upon it's widely used T&T&M platform, with key enhancements to support the cargo-bike service. It features an integrated reservation system that allows registered users to book cargo bikes for their deliveries. Additionally, the platform includes functionality to track the location of the cargo bikes when in use and to locate them when not in use. This ensures that users can easily find and monitor the bikes during their rental period. Tracking will be unavailable if the bike is undergoing maintenance or has been booked by another user. As the owner of the vehicles, the municipality can always monitor their location in real time and access detailed usage reports that show how frequently each shop has used the service.

The tracking devices installed on the cargo bikes are produced and maintained by NGS, as detailed in a previous section of this document. These devices transmit location data to the central platform, where it is collected and transmitted to NGS' software where it's made available to users and the municipality after further processing. The communication between the two platforms is facilitated through a standard API, ensuring smooth data exchange. The data is formatted using the GS1 EPCIS 2.0 standard, a widely adopted format for item-level traceability.

In terms of deployment, the software will initially be tested and used in the Varna's CS. However, the system is not specifically customised for that location, making it easy to replicate and implement in other cities. The architecture of the platform is designed to be scalable and flexible, enabling other municipalities to adopt the service with minimal adjustments.

ID	Requirement	Priority
TTM01	Users can book a vehicle through a calendar that shows availabilities	Should
TTM02	The system keeps record of which user is using the vehicle	**Must**
TTM03	Administrators can always track the vehicles	**Must**
TTM04	Users can track the vehicle only when it is in use by them	Should
TTM05	The system is connected to the Data Space for Single-Sign-On and retrieving real-time data	Should
TTM06	GTFS format will be supported (gtfs.org)	*Could*
TTM07	The system will expose API to be called for pushing real-time updates of the vehicles	Should
TTM08	Bikes routes will be computed with Open Route Service instead of OSRM	*Could*

Table 4: Shared cargo bikes service's requirements

5.3.2 DYNAMIC FLEET MANAGEMENT – KEDGE, GGI

KEDGE and GGI will jointly provide a dynamic fleet management and traffic optimisation service aimed at improving the efficiency, sustainability and robustness of urban freight and mobility operations. The service builds on GGI's fleet management and tracking solutions, enhanced to support multiple types of vehicles (freight and passenger, including trucks, vans, cars and cargo bikes) equipped with NGS or external trackers. Real-time vehicle positions and operational data are integrated through secure and standardised data pipelines (e.g. via EPCIS and AMQP), ensuring interoperability with logistics operators' management systems while respecting strict privacy and anonymization requirements. The tool enables real-time or near real-time visibility of shared vehicle availability, routing conditions and operational constraints, without allowing companies to track competitors.

On top of this infrastructure, KEDGE provides advanced optimisation and decision-support algorithms for dynamic routing, load consolidation and capacity-aware distribution, embedded within the MODALSHIFT Data Space and MTM tools. These algorithms process heterogeneous logistics and mobility data, including shipment metadata, routing plans, vehicle utilisation, urban regulations and smart city infrastructure, to generate optimal or near-optimal solutions depending on operational, tactical or strategic needs. The service will be validated through digital twins and project demonstrators, enabling scenario testing, performance benchmarking and KPI evaluation (e.g. CO_2 emissions, congestion reduction, delivery reliability, robustness and decision time). Together, KEDGE and GGI provide a scalable and interoperable solution supporting logistics operators, public authorities and mobility stakeholders in deploying data-driven, low-emission and efficient transport systems.

ID	Requirement	Priority
PTA01	The app displays real-time information from algorithms developed in T3.4 and T3.5	**Must**
PTA02	Optimal routes are computed leveraging WP3's algorithms	**Must**
PTA03	Personalised routes are computed based on user's preferences	**Must**
PTA04	Optimal and personalised routes are displayed on a map	Should
PTA05	User receives an alert if delays occur on their routes	*Could*
PTA06	GTFS format will be supported (gtfs.org)	Should
PTA07	The app displays timetables from public transports operators	Should

Table 5: Dynamic fleet management's requirements

5.3.3 ROAD & RAIL TRAFFIC FORECASTING – IMEC

The software provided by IMEC in MODALSHIFT on road and rail forecasting will be integrated in the Cooperative Management Module inside ATOBE's Atlas Traffic Management product.
The forecasting software will result in the ability for traffic administrators to consult the latest traffic predictions, as well as to detect recent anomalies. Moreover, as accurate forecasting predictions models mostly require a large amount of training data, it could be beneficial for administrators to understand which data sources are used for processing, especially in the context of multimodality and with possible anonymized data.

ID	Requirement	Priority
RRTF1	Administrators can ask for the latest traffic predictions.	**Must**
RRTF2	Administrators can ask for the recent detected anomalies.	**Must**
RRTF3	Administrators can ask for the processed data.	*Could*

Table 6: Road & Rail traffic forecasting's requirements

5.3.4 DATA SPACE – ITA

The technical specification of the Sovity Data Space is depicted in the diagram below. The diagram shows the trusted data-sharing architecture in a Data Space. A data provider exposes its data only through Apache APISIX, acting as an API gateway. Only APISIX endpoints are registered in the Sovity EDC. The provider uses Sovity EDC to publish data offers and manage contracts. Access control relies on Verifiable Credentials and Verifiable Presentations. The provider issues credentials and verifies presentations from consumers. The consumer uses Sovity EDC to discover offers and negotiate access. It builds a verifiable presentation proving compliance with access policies. Trust is anchored in EBSI, enabling identity and credential verification. If verification succeeds, APISIX authorizes access and data is delivered securely.

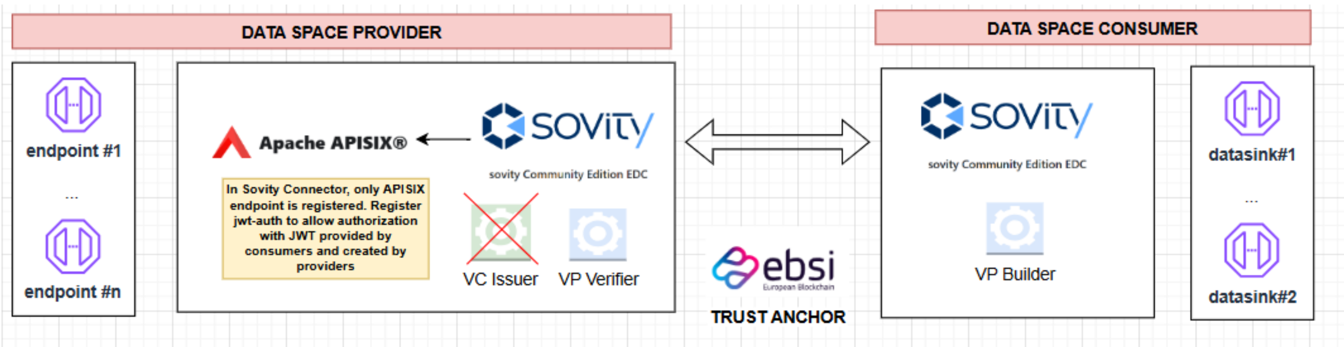


Figure 18 Sovity Data Space

The main requirements for this component are listed below.

ID	Requirement	Priority
SDS01	It is possible to add a dataspace membership and trust anchor configuration	**Must**
SDS02	It is possible to create assets, policies and data offers	**Must**
SDS03	Participants can share data with others with policies and rules	**Must**
SDS04	Control participant identifier with ID, firewall and credentials	**Must**
SDS05	Should offer HTTPS	Should
SDS06	DNS to facilitate communication and search of participants	Should
SDS07	Provide an API to automatically execute CRUD operations	Should
SDS08	Log registry to trace transactions	*Could*

Table 7: Data Space requirements

5.3.5 DATA EXCHANGE MECHANISM – ATOBE, TELLAE

The Open Data exchange mechanism described in the Data Layer of the MODALSHIFT functional architecture will consist of the setup of an historical storage online filesystem and a message broker aligned with the development of a set of data collection tools and data transformers, including the Data Anonymisation by OCTO.

The main requirements for this component are listed below.

ID	Requirement	Priority
DEM01	A Message Queue is available to deliver and subscribe transformed open data in the project	**Must**
DEM02	Historic storage is kept for collected data sources	**Must**
DEM03	It is possible to apply multiple transformations such as aggregation, anonymisation and harmonisation	Should
DEM04	Different data formats are supported such as JSON and Parquet	Should
DEM05	It is possible to add new data sources through configuration	*Could*

Table 8: Data exchange mechanism's requirements

5.3.6 COOPERATIVE MANAGEMENT MODULE – ATOBE

The Cooperative Management Module is supported by ATOBE's Traffic Management Platform (Atlas) which will be extended with results from the TANGENT funded project. The adoption of this module will be divided into two main milestones, the first will be focused on data integration and visualisation. The second will extend the platform with cooperative functionalities supported by models and features from the Optimisation components (see project Functional Architecture – Section 5.1 in this document).

ID	Requirement	Priority
MV01	It is possible to display real-time data from the Case Studies in maps and KPIs	**Must**
MV02	It is possible to display data from different modes in the same map	**Must**
MV03	Open data is available for each Case Study	**Must**
MV04	Private data is available with access controlled according to its specific policies	**Must**
MV05	Operational data from the other services is available	Should
MV06	It is possible to configure the layout of the data displays and which data is present	Should
MV07	Access is controlled by individual user credentials and limited with a role-based permissions mechanism	Should
MV08	It is possible to configure how the data is displayed and easily add new data sources	*Could*

Table 9: Multimodal Visualisation requirements

ID	Requirement	Priority
CMM01	Forecasts and detected anomalies data is available in the platform	**Must**
CMM02	Calculated data is available in the platform	*Could*
CMM03	Historic data is available in the platform	*Could*
CMM04	Historic data is available in the form of trends and histograms	*Would*
CMM05	Users can cooperate via a chat mechanism	Should
CMM06	Optimisation features can be invoked from this module	Should
CMM07	Incident response based on Optimisation features is available	*Could*

Table 10: Cooperative Multimodal Management requirements

5.3.7 AGENT-BASED TRANSPORT MODELLING TOOL – TELLAE, AIT

EQASIM, Starling, ABM-light and MATSim are the Open-Source software tools which will be used.

EQASIM algorithms will be adapted to non-French data where information might be different. Also, due to the lack of some data, new algorithms might be needed in order to replace the missing information. Thanks to the availability of mobile phone data in Madrid new algorithms with this kind of data will be developed.

Regarding Starling, the agent-based simulator, the framework will be benefit from several innovation. This will allow the simulation of logistic use cases by introducing logistics agents (parcels, vehicles, hubs, cargo bikes) and algorithms (dispatch, hubs choice, hubs access choice). Linked to existing algorithms for passenger simulation, multimodal route choice for individuals be introduced and design for passengers and logistics. In addition, algorithms will be introduced in order to manage parcels transport in public transport. Finally, road traffic congestion will be introduced in the framework for more realistic scenarios.

In addition, agent-based demand modelling approaches with different levels of behavioural and network resolution will be developed. A lightweight activity-based (ABM-light) modelling framework, rooted in the classical four-step model and enriched with selected features from agent-based models (e.g. individual trip chaining, temporal dynamics, and multimodal behaviour), will be used to efficiently support scenario design and policy testing. Where higher behavioural detail or detailed network dynamics are required, full-scale agent-based simulations using MATSim can be performed. Depending on the use case, outputs from the ABM-light framework and MATSim can be used to enrich, calibrate, or validate Starling simulations, ensuring consistency across scales while preserving computational efficiency.

ID	Requirement	Priority
ABTM01	Modellers can generate a synthetic population of trips using non-French input data	**Must**
ABTM02	Modellers can generate a synthetic population of trips using mobile phone data	Should
ABTM03	The modelling framework includes parcels, vehicles transport rules and hubs	**Must**
ABTM04	The modelling framework includes a dispatch algorithm for parcels delivery	**Must**
ABTM05	The modelling framework includes parcels hubs choice and hubs access choice	*Could*
ABTM06	The modelling framework includes multimodal route choice for individuals	*Could*
ABTM07	The modelling framework includes cargo bikes logistics transport vehicles and rules	Should
ABTM08	The modelling framework includes parcels transport rules in public transport	Should
ABTM09	The modelling framework includes road traffic congestion with changes over the day	Should
ABTM10	The modelling framework includes road traffic congestion differentiated for each link with changes over the day	*Could*

Table 11: Agent-based transport modelling tool's requirements

5.3.8 TRANSPORT PLANNING, WHAT-IF & VISUALIZATION TOOL – TELLAE

MODALSHIFT will enhance the existing functionalities of the KITE tool regarding the following three themes:

- Data source by giving access to large variety of high-quality data from Open Data or specific data providers;
- Visualisation though the map, tables or charts for enhancing the design of multimodal networks;
- Simulation integrating a variety of what-if scenarios directly into the platform.

Regarding the data theme, TELLAE's existing algorithms and data management will be improved in order to give access to OpenStreetMap data at EU-scale. The data management will have to be robust to huge amounts of data. Algorithms will be adapted and tested to be robust for whole EU, and also to reduce processing time. Additionally, available Open Data will be added for the three pilot sites. Then, more flexibility will be introduced for the algorithms for managing GTFS data from EU countries.

Data visualisation will significatively been improved by adding interactive visualisation of synthetic populations and trips generated by the simulation process. Regarding these heavy detailed data, new forms of data visualisation will be introduced and new indicators will be created. More generally, new KPIs and visualisations will be added. It will lead to plots, tables and spatial indicators. Cross analysis between data will be enhanced. A dedicated zoom in multimodal hubs will be developed with dedicated KPIs.

Finally, simulation use cases will be introduced. This will allow a direct and easy to use connection between the platform and the simulators. Simulators engine will be selected from the agent-based modelling tools developed during the project or other simulation tools which would be relevant. A specific widget will be developed in order to offer functionalities for scenarios exploration and comparison.

ID	Requirement	Priority
KITE01	Running and accessing results from agent-based models / simulation engines or other simulators	Should
KITE02	Widget for exploring simulation scenarios outputs (strengthen the exploration of scenarios)	*Could*
KITE03	Improve bringing together all data around transportation, flows and amenities	Should
KITE04	New KPIs and visualisations for exploring multimodal hubs	*Could*

KITE05	Extended access of OpenStreetMap data at EU-scale	Should
KITE06	Interactive visualisation of synthetic population	**Must**
KITE07	New data from Open Data platform of the three pilot sites	*Could*
KITE08	New data connections and input formats (e.g. parquet, db storage, API, ...)	*Could*
KITE09	New algorithms for strengthening raw data from Open Data (OSM, GTFS, ...)	Should
KITE10	New algorithms for cross analysis between data for territory analysis	**Must**
KITE11	New plots, tables and spatial indicators for KPIs visualisation	Should

Table 12: Transport planning, what-if & visualization tool's requirements

5.3.9 SHUNTING OPERATION ENHANCER – GGI

This software is designed to support and enhance multimodal freight logistics by improving the way shunting operations are planned and managed. Its primary objective is to provide shunting operators with clearer visibility and easier management of the rolling stock under their responsibility, with a particular focus on shunting locomotives.

Within the scope of MODALSHIFT, the software will be specifically adapted to the operational context of ADF, which will act as the shunting operator, and to the Port of Trieste, which represents the physical location where shunting activities take place.

Improved management of shunting locomotives can significantly increase the efficiency of shunting operations. By reducing delays in the movement, positioning, and assembly of unpowered wagons into complete trains, the overall turnaround time for freight movements can be shortened. This increased efficiency supports the broader objective of encouraging rail-based and environmentally sustainable ("green") logistics, making such solutions more attractive and competitive for logistics providers and end customers.

The software will combine ADF internal operational data with relevant open data sources to provide operators with actionable insights into current and near-future resource utilization. In particular, the system will identify shunting locomotives that are expected to be idle or underutilized over the coming days and shunting locations or time periods where a high operational workload is anticipated.

By making this information readily available and easy to interpret, the software will enable ADF to plan and coordinate the reallocation of powered rolling stock between different shunting locations. This forward-looking planning capability helps ensure that locomotives are available where and when they are most needed, reducing inefficiencies, minimizing idle time, and supporting more reliable freight operations overall.

ID	Requirement	Priority
RSM01	The system keeps track of where each managed rolling stock is	**Must**
RSM02	Rolling stocks are displayed on a map	Should
RSM03	The system displays the forecasted workload for each location	Should
RSM04	Rolling stocks location is updated in real-time	*Could*
RSM05	Main statistics and KPIs are displayed to the user	*Could*

Table 13: Shunting operation enhancer's requirements

5.3.10 DATA ANONYMISATION – OCTO

OCTO anticipates 4 key areas for improvement of its software and services:

- broadening the scope of application of Avatar

- integrating data augmentation functionalities to Avatar
- providing interfaces and user access to geolocation trace anonymization functionality
- providing connectors for interaction between Avatar and other MODALSHIFT software

Those themes are detailed below.

Broadening the scope of application of Avatar

During the MODALSHIFT project, OCTO will execute anonymization for a range of datasets identified by the consortium as having potential to improve or enable the proposed services. The variety of services and use cases considered brings a variety of characteristics and constraints on the associated data. While to date, Avatar is a mature solution, the major part of the improvement to the avatar solution will be focused on broadening the scope of application of the solution to efficiently and effectively treat all datasets and needs.

Integrating data augmentation functionalities to Avatar

The technology at the core of avatar enables augmentation to be done with only a few adaptations of the algorithms. Most research and development activities on augmentation has been carried out prior to the project but robust integration to the tool remains. Data augmentation should cover cases where the number of data samples is insufficient for robust modelling and cases where parts of a population are under-represented (class balancing).

Providing interfaces and user access to geolocation trace anonymization functionality

Recent activities of OCTO have resulted in a novel approach able to anonymize geolocation trace data, a type of data particularly difficult to anonymize due to its high re-identification potential and the challenge to alter a trace while preserving its coherence. The geolocation trace solution is currently operational but only proposed as a service carried out by OCTO to its users on demand. Work is required to provide this functionality to its user or to provide minimum interfaces to allow its use within some identified MODALSHIFT pipelines.

Providing connectors

In the section §6.4.3.4 *File formats and connectors* of the Privacy guidelines in this document a more detailed description of current and planned Avatar connectors is present.

ID	Requirement	Priority
AVAT01	Users can anonymize tabular (and time series) datasets with the Avatar tool	**Must**
AVAT02	Users can anonymize geolocation trace data with custom tool	**Must**
AVAT03	Users can anonymize geolocation trace data with the Avatar tool	Should
AVAT04	Users can augment (in number) the number of data points in their tabular data with the Avatar tool	**Must**
AVAT05	Users can augment data in under-represented region of their tabular data with the Avatar tool	**Must**
AVAT06	Users can augment geolocation trace data with custom tool	Should
AVAT07	Users can augment geolocation trace data with Avatar tool	*Could*

Table 14: Data anonymisation's requirements

6 DATA PRIVACY GUIDELINES

In this chapter, the guidelines that the MODALSHIFT partners must follow when processing internal and stakeholders' data are set out, ensuring that all activities are conducted ethically and in full compliance with current regulations.

Following a brief introduction highlighting the importance of data protection and the responsible handling of sensitive information, a dedicated section outlines the key principles and requirements of the General Data Protection Regulation (GDPR) and the Artificial Intelligence Act (AI Act), explaining how these frameworks directly influence and shape OCTO's work within the project.

The final section presents the solution developed by OCTO and made available to the project partners to anonymise data effectively. This solution is designed to ensure full legal compliance while preserving the complete statistical value of the data, thereby enabling meaningful analysis without compromising privacy or regulatory obligations.

6.1 INTRODUCTION

The increasing complexity of data-driven research and cross-border collaborations within European projects necessitates robust privacy frameworks. This privacy guidelines chapter is designed to provide partners with clear, actionable instructions on the considerations of privacy measures and use of the Avatar method, which leverages anonymous synthetic data. The chapter will serve as a foundational reference, ensuring that all stakeholders understand not only the technical implementation but also the regulatory, ethical, and practical considerations of using this technology, particularly in the contexts of structured and geolocation data. The objectives of this chapter are:

- To explain regulatory requirements, focusing especially on the European General Data Protection Regulation (GDPR) and related legislative frameworks.
- To clarify processes and best practices for generating, sharing, and managing anonymous synthetic data.
- To instruct project partners on the appropriate circumstances for using the avatar method.
- To address the specificities of handling structured datasets and geolocation data using the avatar method.
- To promote a culture of privacy-by-design and privacy-by-default among consortium members and their sub-partners.

Note that most topics presented here are also addressed in the extensive Avatar documentation aimed at both technical and non-technical profiles.

6.2 WHY PROTECTING DATA?

Several reasons can motivate data protection in a project like MODALSHIFT:

- Legal frameworks such as GDPR and AI Act are in place and require personal data to be handled and processed with specific considerations. Not complying with those requirements can result in significant financial penalties.
- Beyond the financial impact of poor data protection measures, the reputation of all parties involved in data processing can be impacted. The long-term impact of a poor reputation should not be neglected.
- Complying with the data protection regulations and potentially going further demonstrates the willingness of the data processing parties to adopt an ethical position.
- While data protection measures are initially motivated by the need to protect individuals' data, they can also benefit industry data. In this context, protecting the data refers to the idea of removing sensitive information that could potentially harm a company (by providing an advantage to a competitor for

example). This data protection principle is called desensitization. This argument can serve to convince industrial or use case partners to share their data without risks.

In terms of business impact, data protection measures such as anonymization provide opportunities that would not exist without.

- Access to data becomes possible. Anonymization enables datasets that are confined to a specific use and context to be taken out of this context and used without restrictions. This can be translated in the context of the MODALSHIFT project by an increase of datasets to consider developing decision-making tools. Since MODALSHIFT focuses on multi-modal solutions, those are highly likely to rely on or benefit from the inclusion of many data sources.
- Increase the number of experts that can gain access to a source of data (in its anonymized form). Access to sensitive or personal data is often restricted to a specific number of people. Removing this restriction enables more people, hence more data experts to gain access to the data. Different data expertise can work together with a likely positive impact on resulting algorithms.
- GDPR requires data to be collected and kept for a specified period. By removing the GDPR restrictions, anonymization removes data retention limits.

These arguments represent some key reasons to dedicate some energy to data protection measures. Additional motivations can be drawn from understanding the regulatory frameworks in place and the constraints removed by data protection measures and in particular data anonymization. This is covered in the next section.

6.3 REGULATORY CONSIDERATIONS

The legal management of personal data is a crucial aspect of data processing, ensuring compliance with various regulations while balancing privacy protection and data utility. Personal data regulations exist to establish rights for individuals and responsibilities for organizations handling such data.

Several key principles govern the legal processing of personal data:

- Lawfulness, Fairness, and Transparency: Data must be processed lawfully, with clear and legitimate purposes, and individuals must be informed.
- Purpose Limitation: Data collection and processing must serve specific, well-defined purposes.
- Data Minimization: Only necessary data should be collected and retained.
- Accuracy: Data must be kept accurate and up to date.
- Storage Limitation: Personal data should not be kept longer than necessary.
- Integrity and Confidentiality: Organizations must ensure security and protection against unauthorized access.
- Accountability: Organizations must be able to demonstrate compliance with applicable laws.

Data protection regulations vary across jurisdictions, with key frameworks including:

- European Artificial Intelligence Act (AI Act) – European Union
- General Data Protection Regulation (GDPR) – European Union
- California Consumer Privacy Act (CCPA) – California, USA
- Health Insurance Portability and Accountability Act (HIPAA) – United States (health data)
- New Swiss Federal Act on Data Protection (nFADP / nLPD) – Switzerland

Focusing on the European context, a more detailed data protection is presented as considered in the GDPR and AI Act frameworks.

6.3.1 GENERAL DATA PROTECTION REGULATION (GDPR)

The General Data Protection Regulation (GDPR) is a comprehensive data privacy law enacted by the European Union to protect individuals' personal data. It applies:

- to all organizations that process personal data of EU residents, regardless of where the organization is based
- and to all organizations based in the EU and processing data regardless of the origin of the data (i.e. possibly non-EU data).

According to the Article 4 of the GDPR, personal data refers to:

“*Any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.*”

Examples of personal data include name, email, phone number, IP addresses, financial information and health records.

Under Article 6 of the GDPR, processing personal data is only lawful if it meets one of the following conditions:

- Consent: The individual has given explicit consent.
- Contractual necessity: Processing is necessary for contract performance.
- Legal obligation: Compliance with legal requirements.
- Vital interests: Processing is required to protect life or health.
- Public interest: Processing is necessary for tasks carried out in interest of the public.
- Legitimate interests: The organization has a legitimate reason, provided it does not override the individual's rights.

Article 9 of the GDPR defines sensitive data as information relating to race, religion, political opinions, genetics, biometrics, health, or sexual orientation. Processing such data is generally prohibited, except under specific conditions:

- Explicit consent
- Medical or scientific research
- Public health reasons
- Legal claims

6.3.2 ANONYMIZATION IN THE CONTEXT OF GDPR

In terms of data protection, there are two key complementary steps: pseudonymization and anonymization. With pseudonymization, the data is modified so that it cannot be attributed to a person without additional information. In practice, pseudonymization is often performed by removing or replacing direct identifier variables. Although this step is necessary for data protection, pseudonymized data is still considered personal data under GDPR. With anonymization, the data is altered in an irreversible way that ensures that re-identification is impossible in practice. Anonymized data is no longer subject to GDPR.

The European Data Protection Board (EDPB), formerly known as Working Party 29 (WP29), considers that:

"*The anonymization process, which involves processing personal data to render it anonymous, constitutes further processing.*"

This interpretation highlights that anonymization is not an initial data processing activity but rather an additional transformation applied to personal data.

Under Article 5 of the GDPR, personal data cannot be processed for purposes incompatible with those initially intended.

In other words, any new processing purpose must be compatible with the original purpose communicated to the data subjects at the time of collection. If deemed compatible, further processing is permissible under GDPR regulations.

According to Recital 50 of the GDPR, if the new purpose of further processing is deemed compatible with the original processing purpose, additional notification to the data subjects is not required.

This means that when anonymization aligns with the original processing intent, organizations are not obligated to re-inform individuals about the data transformation.

Authorities such as the French CNIL consider that data subjects must be informed in cases of a "substantial modification" to data processing or in the event of a significant change.

A data processing change is not considered substantial if it falls within the following categories:

- Refinements or clarifications regarding the purpose of data processing.
- Modifications in technical and organizational measures that maintain or enhance data security.

As anonymization could be considered as a treatment aiming to enhance data security, it is not considered as a substantial modification. Organizations are therefore not obligated to re-inform individuals about anonymization.

The concept of "personal data" as defined in Article 4(1) of the GDPR is not unlimited: it requires that the data subject be identified or identifiable. The Court of Justice of the European Union (CJEU) reaffirmed this principle (SRB v. EDPS, September 4, 2025), confirming that identifiability is the central criterion for qualification.

Recital 26 of the GDPR specifies that information is considered anonymous where "the data subject is not or no longer identifiable," taking into account all the means reasonably likely to be used, considering cost, time required, and the technological means available at the time of processing.

Therefore, the status of data as anonymous is not determined by a simple binary threshold. Instead, it depends on:

- the nature of the data,
- their level of sensitivity,
- and the context of use.

An anonymization process must therefore be assessed and documented to demonstrate that the residual risk of re-identification remains reasonable in light of the context and purposes.

The technical assessment of risk should consider the three criteria defined by the EDPB:

- Singling-out: the possibility of isolating a single individual within a dataset,
- Linkability: the possibility of linking two datasets concerning the same individual,
- Inference: the possibility of deducing new, sensitive information about an individual.

In addition, a PIA (Privacy Impact Assessment) must be carried out to evaluate the likelihood of potential attacks against the anonymization process, as well as their impact if successful. In simple terms, the technical report evaluates the residual re-identification risk posed by certain threats, while the PIA measures the probability and severity of those threats in the given context.

Where residual risk is considered too high, additional technical and organizational safeguards can be implemented to further strengthen the anonymization process.

When all these criteria are satisfied, the resulting data is no longer considered personal data under the GDPR and thus falls outside its scope of application.

6.3.3 AVATAR METHOD UNDER GDPR

The EDPB (formerly WP29) acknowledges that anonymization, as further processing of personal data, can be considered compatible with the original processing purpose. However, this is contingent on the anonymization process meeting the criteria outlined in Opinion 05/2014.

The anonymization method developed by OCTO has been awarded a CNIL compliance certificate, confirming its classification as an authentic anonymization technique.

Furthermore, anonymization enhances data security by making identification irreversible. Therefore, it is categorized as a technical and organizational security measure, not a substantial modification of data processing.

- Anonymization of personal health data can be considered a compatible processing activity aligned with the original purposes of data collection.
- Anonymization does not constitute a substantial modification of data processing under GDPR.
- No additional notification or consent from data subjects is required when implementing anonymization.
- Avatarized data resulting from this anonymization process is no longer classified as personal data and thus falls outside the scope of the GDPR.

The data generated using the Avatar Method are not considered anonymous by default. Their anonymity is systematically and automatically evaluated and documented through a rigorous process involving:

- A technical report measuring the residual re-identification risk based on EDPB criteria (Singling out, Linkability, Inference) and metrics simulating state-of-the-art attack scenarios.
- A Privacy Impact Assessment (PIA) evaluating the likelihood and severity of these attacks based on data sensitivity and usage context.

Under all these considerations, the synthetic data generated using the Avatar method is no longer considered personal data and therefore falls outside the scope of the GDPR.

6.3.4 AI ACT

The Artificial Intelligence Act (AI Act) is the world's first comprehensive legislation on AI. Entering into force in August 2024, it aims to ensure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory, and environmentally friendly, while encouraging innovation.

The AI Act classifies AI systems according to the level of risk they pose to individuals' rights and safety:

- Unacceptable Risk: Prohibited systems (e.g., social scoring, cognitive manipulation).
- High Risk: Systems used in critical infrastructure, education, employment, essential services, or law enforcement.
- Limited Risk: Transparency obligations (e.g., chatbots).
- Minimal Risk: The vast majority of systems (e.g., video games, spam filters).

For model developers, the core regulation is the data governance for high-risk systems. According to Article 10 of the AI Act, high-risk AI systems trained on data must comply with strict data governance and management requirements:

"Training, validation and testing data sets shall be subject to appropriate data governance and management practices."

These practices must ensure that data sets are:

- Relevant, representative, free of errors and complete (Art. 10, paragraph 3).
- Examined to detect possible biases (Art. 10, paragraph 2, f).

Training models on real data poses a double legal challenge:

1. GDPR Compliance: Processing personal data for training requires a solid legal basis and compliance with the minimization principle.
2. AI Act Compliance: Article 10 requires high-quality and representative data sets, which often necessitates large volumes of varied data.

Paragraph 5 of Article 10 specifies a notable exception regarding sensitive data (special categories of personal data):

"*To the extent that it is strictly necessary for the purposes of ensuring bias monitoring, detection and correction […], providers of high-risk AI systems may process special categories of personal data […] subject to appropriate safeguards for the fundamental rights and freedoms of natural persons.*"

However, the text implicitly encourages the use of privacy-preserving techniques. Using raw personal data remains a high-risk factor for non-compliance.

This also constitutes a major security risk identified by ANSSI (French National Cybersecurity Agency). In its security recommendations for generative AI systems, the agency highlights that trained models can memorize and leak sensitive data in the event of an attack. To reduce these risks, ANSSI explicitly mentions the use of synthetic data as a relevant protective measure, recommending that models be trained only with data whose sensitivity is consistent with the access rights of end-users (Recommendation R18).

Training high-risk models therefore raises a business challenge: how to freely exploit the full potential of data without exposing the company to legal or security risks.

Article 53 and following articles establish regulatory sandboxes for AI. These are controlled environments that facilitate the development and testing of innovative AI systems under the supervision of competent authorities before they are placed on the market.

The goal is to enable innovation while ensuring compliance with the GDPR and the AI Act. Access to quality data in these environments is crucial. The provision of anonymous data facilitates the implementation of these sandboxes.

6.3.5 AVATAR METHOD UNDER THE AI ACT

OCTO's synthetic data generation method positions itself as a strategic compliance lever for AI providers. It allows meeting the requirements of Article 10 while activating the value of personal data. The anonymous nature of the generated data unlocks the full business potential of the original data by removing barriers related to personal data processing.

The AI Act requires data to be "relevant, representative, and complete." Unlike classic anonymization techniques (such as masking or strong generalization) which strongly impact data utility, the Avatar method preserves original statistical properties. Models trained on Avatar synthetic data perform equivalently to those trained on real data, guaranteeing the technical compliance required by Europe.

Synthetic data generation allows for rebalancing data sets (e.g., increasing the representation of a minority class) to mitigate algorithmic biases. This directly addresses the obligation to detect and correct biases imposed by the AI Act, while protecting the privacy of the individuals concerned.

In summary, using the Avatar method for the AI Act helps to:

- Eliminate legal risks related to the use of personal data (GDPR).
- Guarantee model performance thanks to statistical preservation.
- Demonstrate responsible data governance, a key criterion for CE certification of high-risk AI systems.
- Accelerate Time to Market by reducing the R&D cycle by 50% (access to data often takes 6 to 18 months).

- Increase AI Performance. The Avatar method allows for the generation of augmented and balanced datasets, improving model robustness and accuracy.
- Reduce Operational Costs. Using anonymous synthetic data lowers cybersecurity costs and secure storage expenses.

6.4 AVATAR METHOD

6.4.1 WHAT IS AVATAR?

The avatar method is a patented privacy-preserving data synthesis technique that replaces real individuals' data points with artificial "avatars"—synthetic records statistically like the originals, but non-identifiable. This approach enables organizations to analyse and share datasets without exposing sensitive personal information. By generating avatars that mirror the key distributions and relationships of the source data, the method supports compliance with privacy regulations such as GDPR while maintaining analytical value for research and development purposes. In particular, it has been shown that avatars can be used to train statistical and machine learning models with no or minimal utility loss.

The Avatar solution can handle data in a single table or relational databases (multiple tables), and different types ranging from demographic and time series data to geolocation. In this section, the focus is on the single table case to illustrate the core mechanism of the method. This concept is adapted to handle other types of datasets but the principle remains the same.

In a single table context, the core of the Avatar method follows three major steps.

1. **Input**: the input data takes the form of a pseudonymized tabular dataset where each row corresponds to an independent individual.
2. **The Avatar transformation**:
    a. Individual observations are projected into a multidimensional space where distances between individuals can be computed.
    b. Using the first dimensions of this space, pairwise distances are computed between all sensitive individual observations to find the k nearest neighbours of each individual. Those neighbours are used to define a local area.
    c. For each individual, a set of coordinates is drawn in their local area. Those coordinates represent the avatar in the projected space. The number of neighbours k defines how large the local area around each individual is and consequently controls the expected privacy level. The coordinates sampling follows a procedure described in detail in a scientific article. While describing it in full length is not justified in the present document, it is important to understand that the sampling relies on two distinct sources of randomness. This ensures that the overall avatar procedure is irreversible, a legal requirement for an anonymization solution.
    d. The produced avatars are evaluated in terms of utility and privacy.
3. **Output**: the avatars are reverse transformed from their coordinates in the projected space into values in the original space corresponding to the initial dataset. Synthetic observations (rows) are shuffled to remove the link between the original individuals and the avatars.

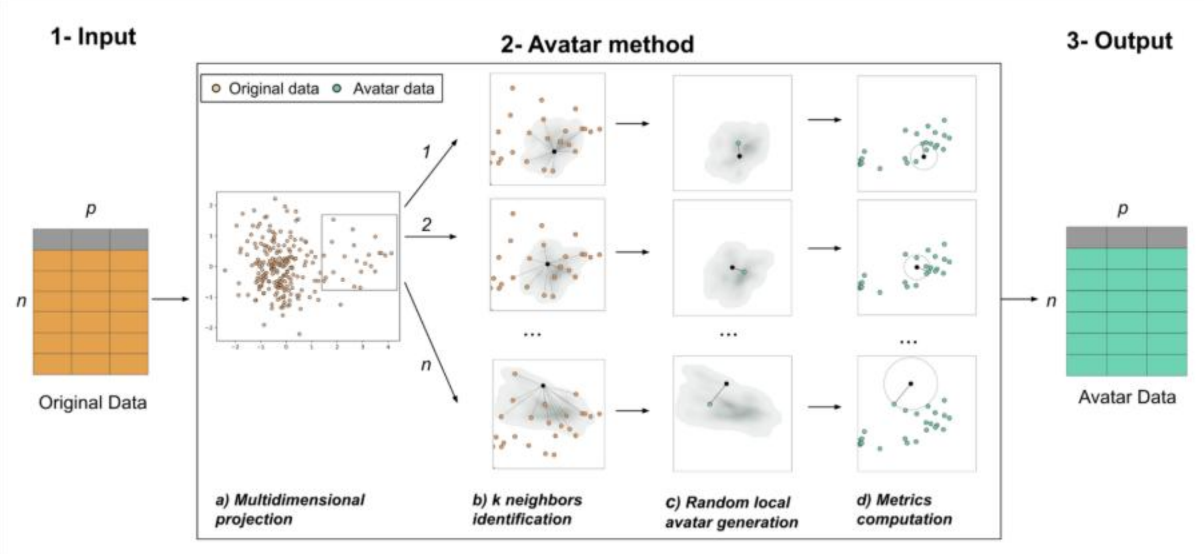


Figure 19 Avatar method

The Avatar method uses local modelling to stochastically generate a synthetic individual. (1) Original pseudonymized sensitive data. (2) The anonymization consists of four steps: (2a) individuals are projected in a multidimensional space; (2b) pairwise distances are computed to find the k nearest neighbours (here k=12) in a reduced space; (2c) a synthetic individual is pseudo-randomly generated in the subspace defined by the neighbours; (2d) privacy and utility metrics are evaluated. (3) Output of the synthetic dataset.

This anonymization method stands out due to its transparent and explainable design, allowing full control over the anonymization process and its evaluation. By generating data within a locally projected modelling space, the approach ensures the creation of plausible synthetic individuals preserving multivariate relationships.

Additionally, the introduction of controlled randomness guarantees the irreversible nature of the anonymization process. The individual-centric approach further enhances the assessment of residual risk, providing a more comprehensive evaluation of the privacy protection of the generated data.

One of the key strengths of the avatar method lies in its ability to balance privacy and data utility. By quantifying privacy risks—such as the likelihood of re-identification—and iteratively refining the synthesis process, the method allows practitioners to calibrate the level of similarity between the synthetic and original data based on risk tolerance and intended use. Furthermore, integrated validation steps, such as statistical fidelity checks and privacy audits, help ensure that the produced avatars meet both compliance and performance benchmarks. This makes the avatar method particularly suitable for collaborative environments, multi-institutional research, and scenarios requiring cross-border data sharing under strict privacy regulations.

At its core, the avatar method employs advanced statistical modelling and machine learning to capture the nuanced patterns, correlations, and outliers present in the original data. Instead of merely shuffling or masking values, it creates entirely new data points that resemble the authentic dataset in both structure and utility yet are disassociated from any real individual. This process is particularly powerful for high-dimensional or relational data, where preserving the interplay between variables is crucial for meaningful analysis.

The avatar method is extensively detailed in the aforementioned scientific article in Nature Digital Medicine. Further information and in-depth analysis of the generated anonymous data are detailed in this paper.

6.4.2 HANDLING THE PRIVACY – SIGNAL DILEMMA

Synthetic anonymous data generation aims to faithfully replicate the characteristics of the original training data while ensuring the privacy of the individuals from whom the information originates. To ensure optimal quality of the generated data, it is crucial to quantify both privacy and the preservation of utility in the synthetic dataset.

OCTO proposes a set of metrics to evaluate these two aspects. These metrics provide an objective measure of the level of privacy protection across multiple criteria while verifying that the synthetic data retains analytical value comparable to the original dataset across various evaluation levels.

The **context of use**, the data **sensitivity** and the **utility preservation goal** are critical considerations in any anonymization process. The context defines the specific objectives and potential threats to mitigate, shaping the methods and parameters employed. For instance, data intended for research purposes may require a different approach than data shared publicly. Similarly, the sensitivity of the data—whether tied to its nature (e.g., health, financial) or its granularity—dictates the level of protection needed. Finally specific uses may require keeping highly granular yet very re-identifying information. Effective anonymization must account for all these dimensions to ensure the optimal result both in terms of privacy and utility.

6.4.2.1 Privacy

The technical evaluation of synthetic data anonymity is based on three fundamental criteria defined here by the European Data Protection Board (EDPB). These criteria are essential for assessing the regulatory compliance of an anonymized dataset:

- **Singling Out**: the risk of identifying an individual within a dataset.
- **Linkability**: the ability to connect individuals to an external data source containing shared characteristics.
- **Inference**: the possibility of deducing, with significant probability, information about individuals from the anonymized dataset.

To date, the Avatar tool proposes eight metrics relevant to singling-out and one metric for linkability and inference each. The relatively high number of singling-out metrics reflects the stronger importance given to singling-out criteria by the EDPB. Below are described some key metrics.

- **Local Cloaking**. The Local Cloaking metric is specific to the Avatar method. It could be seen as a density metric. It is designed to address a distance-based record linkage attack scenario. This metric evaluates the loss of the link between an original data point and its avatar in a context where each individual generates their own synthetic data. It quantifies how isolated an individual's avatar is from their original data point by measuring the median number of other avatars positioned between the original data point and its avatar in the distance space.
- **Hidden Rate**. As for Local Cloaking, the Hidden Rate is specific to the Avatar method and can be seen as a safeguard that ensures there has been no algorithmic flaw and that the link between the original data and the synthetic data has been effectively broken. The Hidden Rate can be interpreted as follows: it represents the percentage of individuals for whom the generated avatar is not the most similar entry in the dataset.
- **Distance To Closest**. The Distance To Closest metric computes the median Euclidean distance between an individual and the closest avatar in the dataset, without taking into account the link between the original data point and its corresponding avatar. This metric provides insight into how close the synthetic data lies to the distribution of the original data, while avoiding direct one-to-one connections between original individuals and their avatars. For interpretability, the obtained value is compared to a reference value.
  The final score is the ratio between the Distance To Closest computed on the avatars and the one computed on a reference dataset. This comparison makes it possible to evaluate whether the proximity of avatars to original individuals is similar to what is observed in the reference data.
- **Closest Distances Ratio**. The Closest Distances Ratio is a local density metric. It computes the median ratio between the Euclidean distance separating an individual from their closest avatar, and the distance separating that same individual from their second closest avatar. For interpretability, and similar to

Distance To Closest, the final score is the ratio between the Closest Distances Ratio computed on the avatars and the one computed on a reference dataset.

- **Closest Rate**. The Closest Rate metric combines the results of the Distance to Closest and Closest Distances Ratio metrics to determine the percentage of individuals whose most similar avatar is not too close, and whose original data/avatar pair is not easily distinguishable from the rest of the dataset. It provides an overall measure of the dissociation between the original data and their avatars in the synthetic dataset.
- **Linkability Protection Rate**. The Linkability Protection Rate metric is specifically designed to evaluate the linkability criterion. It quantifies the risk of correlation in synthetic datasets by simulating scenarios in which different subsets of variables are used to establish matches between avatars and their corresponding original records. This metric directly addresses the risk of cross-linking, providing a robust measure of privacy protection. The simulated scenario consists of selecting two distinct subsets of variables that could realistically be available in an external database (such as demographic information). Each original individual from both subsets is then matched to their closest avatar within the feature space defined by the selected variables. If an individual is linked to the same avatar across both subsets, this reveals a linkability risk for that individual. The Linkability Protection Rate is expressed as a percentage, where higher values indicate stronger protection. A score of 100% means that no avatar was linked to the same original record across the different variable subsets — demonstrating complete dissociation between the synthetic and original data. Conversely, lower values indicate a greater linkability risk, as some avatars maintain a stable association with their original record even when the linking variables change.
- **Attribute Inference**. The Attribute Inference metric evaluates the risk that an attacker could successfully deduce the value of a target variable using a subset of known attributes. An inference is considered successful when the predicted modality matches the true modality for a categorical variable, or when the predicted value is sufficiently close to the real value for a continuous variable (~5%). In practice, each individual is assigned the target value of the synthetic record that is most similar in terms of attribute proximity. The resulting prediction performance constitutes the Attribute Inference score. This metric is particularly relevant in attribute inference attack scenarios, where attackers leverage publicly available or easily accessible demographic data to infer sensitive categorical information about individuals.

6.4.2.2 Utility

The evaluation of synthetic data quality hinges on a delicate balance between privacy and utility. While privacy ensures that individuals cannot be identified or linked, utility focuses on the preservation of the original dataset's statistical properties, enabling synthetic data to be a reliable substitute for real data.

To assess the utility of synthetic data, a framework based on three generic levels of metrics is used. These levels offer a comprehensive evaluation of how well the synthetic data retains the essential characteristics of the original dataset:

- Univariate Utility: Measures the similarity of individual variable distributions between the original and synthetic datasets.
- Bivariate Utility: Evaluates the preservation of pairwise correlations between variables.
- Multivariate Utility: Assesses the retention of the dataset's global structure, including clustering patterns and overall data relationships.

While these three levels provide a generic foundation for utility evaluation, the incorporation of use-case-specific metrics tailored to the needs of the dataset and its intended applications is encouraged. Custom metrics allow for a more targeted assessment of the synthetic data's utility by focusing on specific aspects or features crucial to the use case.

Utility must always be evaluated in tandem with privacy to ensure that the synthetic data fulfils its purpose without compromising anonymity. As with privacy, the context of use is critical in determining the adequacy of utility.

For exploratory data analysis, retaining global trends and distributions may be sufficient. For training machine learning models, the preservation of multivariate structures and complex interactions is often necessary.

By systematically evaluating utility across these three levels and incorporating use-case-specific metrics, the Avatar solution empowers users to ensure that synthetic data is both privacy-compliant and fit for its intended purpose. This dual focus ensures the creation of synthetic datasets that are secure, reliable, and practical for a wide range of applications.

6.4.3 HOW TO USE AVATAR

6.4.3.1 Prerequisites to use and access Avatar method

All MODALSHIFT partners have access to a full Avatar license provided through the project at no additional cost. This license enables partners to run anonymization tasks on the Avatar SaaS with no restrictions. To receive login credentials and instructions, partners are invited to email support@octopize.io or contact any OCTO project member.

The Avatar solution can be accessed via the web interface (address to date: www.octopize.app). The web interface does not require any prerequisite in terms of data science knowledge. For technical users with data manipulation skills, a python interface is available. Both web and python interfaces access the same SaaS server and have similar features.

While data science prerequisites are not needed, it is recommended to know the basic principles of anonymization. For this reason, OCTO proposes training sessions to its users to cover data privacy, anonymization and avatar principles. All MODALSHIFT partners are expected to attend two workshops on ethical data practices and anonymization (Task T8.3).

6.4.3.2 Setting up and running an anonymization

OCTO dedicates a lot of efforts in democratizing and simplifying data anonymization:

- An extensive documentation available at https://docs.octopize.io/ summarizes the key information related to data privacy and anonymization, provides transparent descriptions of the mechanisms around Avatar along with details on how to run the solution.
- Different modes of expertise are available when anonymizing via the web interface. The base mode ensures that only the key information and parameters are provided to the user, while an expert mode provides access to the full range of functionalities of the Avatar solution.
- Blueprints are available to ease the parametrization of classic tasks (e.g. anonymization, augmentation …)
- A series of python tutorials is available on the Avatar python GitHub public repository. Each tutorial focuses on an anonymization topic. To date, tutorials cover the following topics: basics of avatar anonymization, parametrization of avatar, multi-table anonymization, time series anonymization, and data augmentation.
- Technical support is always available at support@octopize.io
- A training workshop for all MODALSHIFT members is planned for 2026 aiming to cover the principles and use of Avatar.

- Examples of anonymization are provided via a series of anonymization demo packs written as part of the MODALSHIFT project. Each pack provides a dataset, an anonymized version and a document providing context and detailing the relevance to different topics covered in MODALSHIFT. Demo packs are available on the MODALSHIFT SharePoint.

A detailed description on how to use the Avatar solution is available in the AVATAR's documentation.

While the objective is to train the users to anonymize their data in full autonomy, anonymization can also be performed by OCTO as a service following discussions.

6.4.3.3 Geolocation trace data

To date, the Avatar solution provided to users for a usage in full autonomy covers data organized in a single table (where a row represents an individual), in a relational database (multiple tables) or in a time series format.

The Avatar solution is also compatible as is with geolocation data containing a small number of points. It is well adapted to Origin-Destination representations or to independent geolocation variables such as coarse-grain addresses.

When the number of geolocation points is high, the term geolocation traces is used. Examples of geolocation traces include for example, the tracking of a vehicle or person in movement over a period of time. Studies have demonstrated that only a handful of points (i.e. 4) sampled from a group of geolocation traces is sufficient to re-identify almost all individuals in this population (de Montjoye, YA. et al.). The retention of utility is another major challenge when anonymizing geolocation traces.

OCTO has focused on those topics in the past 4 years and developed an approach able to anonymize geolocation traces. The resulting invention is currently subject to a pending patent application.

In the MODALSHIFT context, geolocation trace data can be anonymized on demand by OCTO as a service. Secure channels and NDAs can be put in place to allow initial transfer of data. Note that the trace anonymization functionality is expected to be integrated to the Avatar solution during the MODALSHIFT project.

6.4.3.4 File formats and connectors

The Avatar solution currently supports data in csv and parquet formats. Other formatting constraints need to be respected. In particular:

- Data should be pseudonymized (direct identifiers such as names, SSN and contact details removed).
- Data should be reasonably cleaned.
- To anonymize multiple tables, tables should be linked with each other by means of coherent primary and foreign keys relationships.
- Time series tables should be provided in a flat data format (i.e. it should contain identifier, time and value variables).

To ease the interaction between Avatar and other software in the MODALSHIFT project, work on connectors will be carried out. While the list of formats is not formalized and can be adapted based on the data identified for anonymization, it is anticipated that connectors to several formats should be developed:

- GeoJSON
- EPCIS 2.0 & CBV files

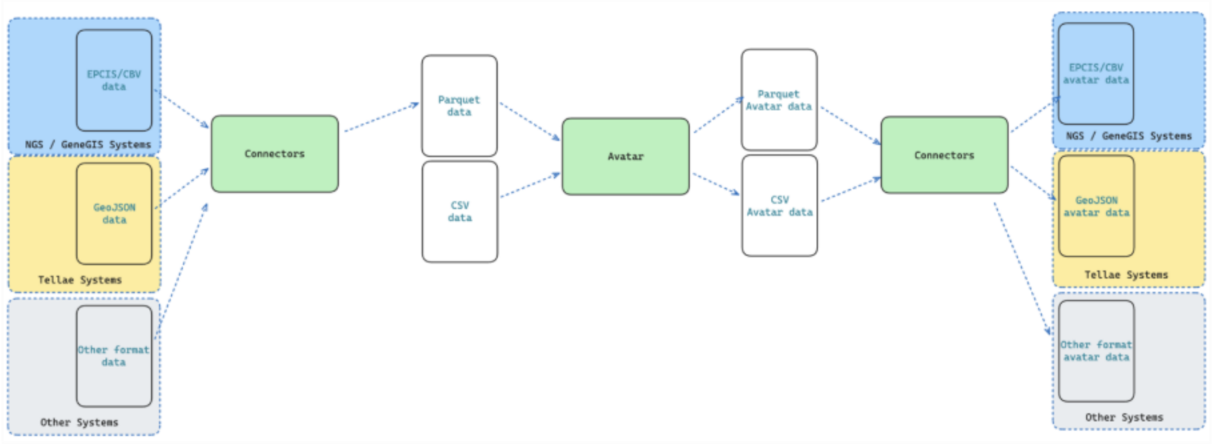


Figure 20 Avatar Connectors

Connectors ensure compatibility between Avatar and different file formats, standards and environments.

6.4.4 DOCUMENTATION AND AUDIT TRAILS

It is essential that all partners involved in the synthetic data workflow maintain rigorous documentation of the whole process. For every dataset that undergoes anonymization, partners must retain comprehensive privacy metrics as verifiable evidence of privacy-preserving processing. These records support auditability, facilitate compliance with regulatory requirements, and ensure stakeholders that privacy standards have been met throughout the data lifecycle.

The Avatar solution automatically generates two reports to summarize and document an anonymization:

- A technical report
- A Privacy Impact Assessment (PIA) report

These two reports are complementary and allow for a complete vision for the final decision: **is the residual re-identification risk reasonable regarding the context?**

- **If yes**: Data is considered anonymous in this context.
- **If no**: Additional measures can be added (encryption, limited sharing, etc.).

6.4.4.1 Technical report

The technical assessment focuses on the **residual re-identification risk** of a dataset. This report is purely technical and relies on the three criteria defined by the EDPB using privacy metrics. While crucial, this report can be complex to interpret for legal or non-technical profiles. Its goal is to provide a quantitative measure of re-identification risk against various attack scenarios.

6.4.4.2 Privacy Impact Assessment report

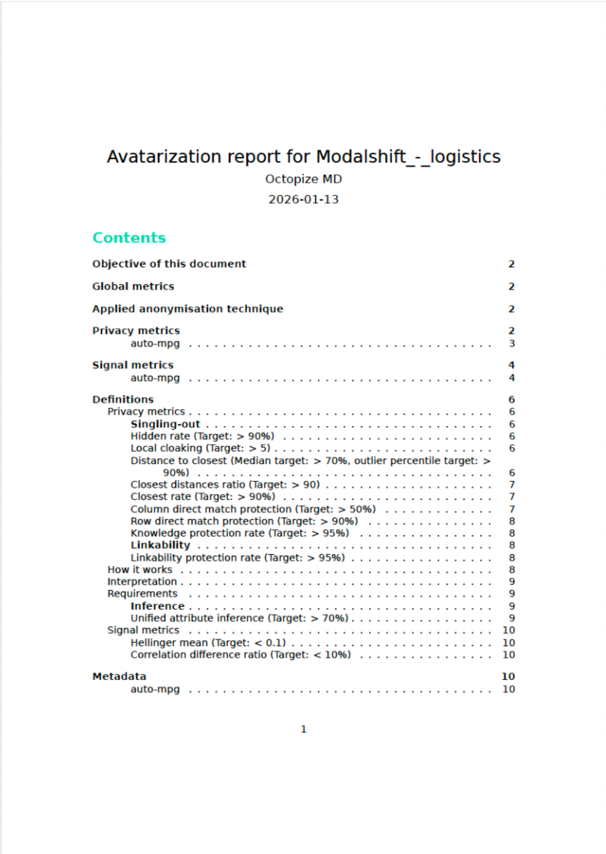
The Impact Assessment (or Privacy Impact Assessment) complements the technical assessment. Where the technical report measures a theoretical re-identification risk, the impact assessment evaluates the likelihood and severity of this risk in the real world.

This report takes into account the global context of data usage:

- **Probability of occurrence**: Are the elements needed to perform an attack reasonably accessible?
- **Impact**: What would be the consequences if an attack were successful, considering data sensitivity?
- **Trust level**: Who are the data recipients?
- **Data subjects**: Who are the individuals concerned?

- **Demographic variables**: How many variables could be used for attacks?

The automatically generated impact assessment report is adapted based on the data usage context, sensitivity, and trust level provided by the user. The format of this report is also intentionally editable (docx) to meet the specific customization and arbitration needs of each organization or industry sector.



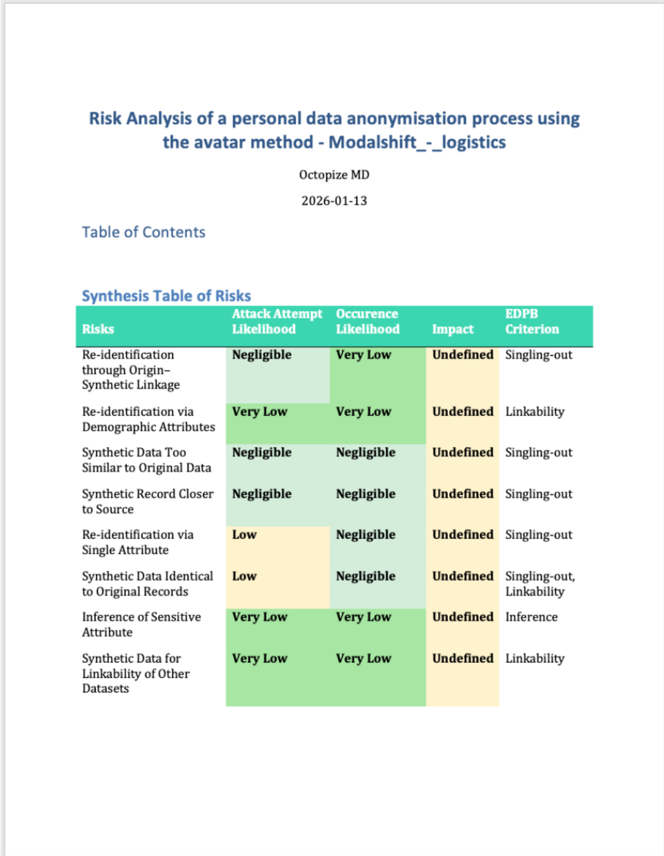
Avatarization report for Modalshift_-_logistics

Octopize MD

2026-01-13

Contents

1



Risk Analysis of a personal data anonymisation process using the avatar method - Modalshift_-_logistics

Octopize MD

2026-01-13

Table of Contents

Synthesis Table of Risks

Risks	Attack Attempt Likelihood	Occurence Likelihood	Impact	EDPB Criterion
Re-identification through Origin–Synthetic Linkage	**Negligible**	**Very Low**	**Undefined**	Singling-out
Re-identification via Demographic Attributes	**Very Low**	**Very Low**	**Undefined**	Linkability
Synthetic Data Too Similar to Original Data	**Negligible**	**Negligible**	**Undefined**	Singling-out
Synthetic Record Closer to Source	**Negligible**	**Negligible**	**Undefined**	Singling-out
Re-identification via Single Attribute	**Low**	**Negligible**	**Undefined**	Singling-out
Synthetic Data Identical to Original Records	**Low**	**Negligible**	**Undefined**	Singling-out, Linkability
Inference of Sensitive Attribute	**Very Low**	**Very Low**	**Undefined**	Inference
Synthetic Data for Linkability of Other Datasets	**Very Low**	**Very Low**	**Undefined**	Linkability

Figure 21 Technical and Privacy Impact Assessment automatic reports

7 CONCLUSION

This document defines a comprehensive framework for the design, integration, and operation of the multimodal transport solutions, drawing on the distinct technical expertise of each participating technical organisation. Collectively, the partners contribute specialized capabilities across transportation modes, terminal operations, digital platforms, tracking technologies, and systems integration, fostering seamless coordination across road, rail and sea networks.

The technical specifications establish standardized requirements for infrastructure interfaces, data exchange, tracking and tracing and monitoring, traffic forecasting, logistics optimisation and exception management across the multimodal logistics chain. By defining common data models and communication protocols, the specifications reduce fragmentation, support real-time decision-making, and enable scalable expansion of the logistics network.

The work presented in this document is sustained by a comprehensive gap analysis that examines in detail the needs of the three CSs, while also considering the wider European context and the strategic directions set by the European Commission.

Three principal gaps have been identified, aligned with the scope of MODALSHIFT. The first concerns data sharing in multimodal transport: fragmented systems and a lack of trust between stakeholders hinder the effective flow of data, resulting in underutilisation of infrastructure and a limited capacity to scale up services. A second gap relates to the slow evolution of transport infrastructure and digital services. The project seeks to address this by promoting a more flexible and forward-looking approach, beginning at the design stage. Finally, railways are recognised as playing a pivotal role in advancing sustainable freight and passenger transport. Strengthening multimodality is therefore seen as a key pathway towards achieving a greener logistics.

Data privacy guidelines are embedded throughout the specifications to protect commercially sensitive and personal data exchanged between logistics partners, platforms, and authorities. These guidelines ensure compliance with applicable data protection regulations while supporting secure data sharing, role-based access, and traceability across the multimodal ecosystem.

In summary, this document provides a unified technical and governance foundation for multimodal transport operations, enabling interoperability, regulatory compliance, operational efficiency, and resilience.

8 REFERENCES

- 5G-MOBIX, https://www.5g-mobix.com/ accessed 16 February 2026.
- **AI Act** AI Act, 'The EU Artificial Intelligence Act' https://aiactinfo.eu/ accessed 20 January 2026.
- **Avatar python GitHub repository**, https://github.com/octopize/avatar-python.
- C-ROADS, https://www.climateinteractive.org/c-roads/ accessed 16 February 2026.
- **California Consumer Privacy Act (CCPA)** State of California Department of Justice, 'California Consumer Privacy Act (CCPA)' (Office of the Attorney General) https://www.oag.ca.gov/privacy/ccpa accessed 20 January 2026.
- **CNIL** Commission Nationale de l'Informatique et des Libertés (CNIL), 'Home' https://www.cnil.fr/en/home accessed 20 January 2026.

- **de Montjoye, YA., Hidalgo, C., Verleysen, M. et al.** de Montjoye Y-A and others, 'Unique in the Crowd: The Privacy Bounds of Human Mobility' (2013) 3 *Scientific Reports* 1376 https://doi.org/10.1038/srep01376.

- **EPCIS** GS1, 'EPCIS' https://www.gs1.org/standards/epcis accessed 20 January 2026.

- European Commission's **Sustainable and Smart Mobility Strategy** https://transport.ec.europa.eu/transport-themes/mobility-strategy_en accessed 3 February 2026

- European **Green Deal** https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en accessed 3 February 2026

- **GDPR** Intersoft Consulting, 'General Data Protection Regulation (GDPR)' https://gdpr-info.eu/ accessed 20 January 2026.

- **Guillaudeux, M., Rousseau, O., Petot, J. et al.** Guillaudeux M and others, 'Patient-centric synthetic data generation, no reason to risk re-identification in biomedical data analysis' (2023) 6 *npj Digital Medicine* 37 https://doi.org/10.1038/s41746-023-00771-5.

- **Health Insurance Portability and Accountability Act (HIPAA)** Centers for Disease Control and Prevention, 'Health Insurance Portability and Accountability Act of 1996 (HIPAA)' https://www.cdc.gov/phlp/php/resources/health-insurance-portability-and-accountability-act-of-1996-hipaa.html accessed 20 January 2026.

- **Lincan Li, Hanchen Wang, Wenjie Zhang, Adelle Coster**, 2024, 'STG-Mamba: Spatial-Temporal Graph Learning via Selective State Space Model' arXiv:2403.12418 [cs.LG] https://doi.org/10.48550/arXiv.2403.12418

- **MATSim model of Vienna** https://github.com/ait-energy/matsim-model-vienna accessed 16 February 2026

- **New Swiss Federal Act on Data Protection (nFADP / nLPD)** State Secretariat for Economic Affairs (SECO), 'New Federal Act on Data Protection (nFADP)' (SME Portal) https://www.kmu.admin.ch/kmu/en/home/facts-and-trends/digitization/data-protection/new-federal-act-on-data-protection-nfadp.html accessed 20 January 2026.

- **Octopize Documentation** https://docs.octopize.io/ accessed 21 January 2026.

- **Opinion 05/2014** Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP216 https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf accessed 20 January 2026.

- **Recital 26 GDPR** Intersoft Consulting, 'Recital 26: Not Applicable to Anonymous Data' https://gdpr-info.eu/recitals/no-26/ accessed 20 January 2026.

- **Recital 50 GDPR** Intersoft Consulting, 'Recital 50: Further Processing' https://gdpr-info.eu/recitals/no-50/ accessed 20 January 2026.

- **Rudloff, C., Straub, M.** Mobility surveys beyond stated preference: introducing MyTrips, an SP-off-RP survey tool, and results of two case studies. Eur. Transp. Res. Rev. 13, 49 (2021). https://doi.org/10.1186/s12544-021-00510-5

- **SRB v. EDPS, September 4, 2025** *Single Resolution Board v European Data Protection Supervisor* (Case C-413/23 P) [2025] CJEU https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62023CJ0413 accessed 20 January 2026.

- **TANGENT**, https://tangent-h2020.eu/ accessed 16 February 2026.

- **Yuchen Fang, Yuxuan Liang, Bo Hui, Zezhi Shao, Liwei Deng, Xu Liu, Xinke Jiang, and Kai Zheng**. 2025. Efficient Large-Scale Traffic Forecasting with Transformers: A Spatial Data Management Perspective. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1 (KDD '25). Association for Computing Machinery, New York, NY, USA, 307–317. https://doi.org/10.1145/3690624.3709177

- **Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Chengqi Zhang** 2019, 'Graph WaveNet for Deep Spatial-Temporal Graph Modeling' arXiv:1906.00121 [cs.LG] https://doi.org/10.48550/arXiv.1906.00121

- **Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang**. 2020. 'Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks'. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20). Association for Computing Machinery, New York, NY, USA, 753–763. https://doi.org/10.1145/3394486.3403118