Grant Agreement Number: 101101962
Project Acronym: FP6 - FutuRe
Project title: Future of Regional Rail

DELIVERABLE D6.6
SPECIFICATION OF PASSENGER CONGESTION MONITORING

Project acronym:	FP6 - FutuRe
Starting date:	01/12/2022
Duration (in months):	48
Call (part) identifier:	Call: EU-RAIL JU Call Proposals 2022-01 (HORIZON-ER-JU-2022-01) Topic HORIZON-ER-JU-2022-FA6-01
Grant agreement no:	101101962
Grant Amendments:	NA
Due date of deliverable:	31-08-2024
Actual submission date:	06-09-2024
Coordinator:	Alessandro Mascis, Wabtec
Lead Beneficiary:	GTSP
Version:	1.0
Type:	Report
Sensitivity or Dissemination level	PU
Taxonomy/keywords:	Congestion, Mobility, Delay Estimation, Timetable Synchronization, Traveller Feedback, Optimization

This project has received funding from the Europe's Rail Joint Undertaking (JU) under grant agreement 101015423. The JU receives support from the European Union's Horizon Europe research and innovation programme and the Europe's Rail JU members other than the Union.

Document History

Date	Name	Affiliation	Position/Project Role	Action/ Short Description
01/07/2024	Tiago Fonseca	GTSP	Task leader	Author, Initial version and structure.
18/07/2024	Anna Beszczyńka	PKP	Task participant	Author, input for Task 6.5
18/07/2024	Piotr Chyliński	PKP/IK	Task participant	Author, input for Task 6.5
18/07/2024	Agnieszka Łukasiewicz, Phd.	PKP/IBDiM	Task participant	Author, input for Task 6.5
18/07/2024	Tiago Fonseca	GTSP	Task leader	Review, internal delivery version
30/07/2024	Ira Kataria, Matthias Walter	Hacon	WP leader, WP co-leader	Review
04/08/2028	Fabian Hois	d-Fine	Task participant	Review
05/08/2024	Tiago Fonseca and Rui Eirinha	GTSP	Task leader and Task participant	Final version
20/08/2024	Noelia Medrano Buendia and Fabrizio Burro	ENYSE/Wabtec	FP6 reviewer	FP6 internal technical review
21/08/2024	Fabrizio Burro	Wabtec	Deputy Project Coordinator	Quality review
23/08/2024	Tiago Fonseca; Rui Eirinha; Agnieszka Łukasiewicz, Phd.; Trenitalia	GTSP,	Task leader and participants	Final version
06/09/2024	Fabrizio Burro	Wabtec	Deputy Project Coordinator	Coordination of Steering Committee review

Table of Contents

Executive Summary 6

List of Abbreviations, Acronyms, and Definitions 8

List of Figures 9

List of Tables 10

1. Introduction 11

2. Objective/Aim 13

3. Background 14

4. Methodology 16

5. Use Cases 18

5.1. System Actors Identification 19

5.2. UC-FP6-WP6-5.01 20

5.3. UC-FP6-WP6-5.02 21

5.4. UC-FP6-WP6-5.03 22

5.5. UC-FP6-WP6-5.04 23

6. Capabilities and Requirements 25

6.1. T6.5_CA01 Weather and Train Data Collection 26

T6.5_UC5.1_FRQ01 26

T6.5_UC5.1_FRQ02 27

T6.5_UC5.1_PRQ01 27

6.2. T6.5_CA02 Predict Schedule Delays 28

T6.5_UC5.1_FRQ03 28

T6.5_UC5.1_FRQ04 28

T6.5_UC5.1_FRQ05 29

T6.5_UC5.1_FRQ06 29

T6.5_UC5.1_FRQ07 30

T6.5_UC5.1_NFRQ01 30

6.3. T6.5_CA03 Schedule Data Collection 31

T6.5_UC5.2_FRQ01 31

T6.5_UC5.2_PRQ01 31

6.4. T6.5_CA04 Schedule Synchronization Analysis 32
T6.5_UC5.2_FRQ02 32
T6.5_UC5.2_FRQ03 33
T6.5_UC5.2_FRQ04 33
T6.5_UC5.2_PRQ02 34
6.5. T6.5_CA05 Traveller Feedback Collection 34
T6.5_UC5.3_FRQ01 34
T6.5_UC5.3_FRQ02 34
T6.5_UC5.3_FRQ03 35
T6.5_UC5.3_FRQ04 35
T6.5_UC5.3_NFRQ01......36
T6.5_UC5.3_NFRQ02......36
T6.5_UC5.3_NFRQ03......37
6.6. T6.5_CA06 Feedback Data Aggregation and Analysis 37
T6.5_UC5.3_FRQ05 37
T6.5_UC5.3_FRQ06 38
T6.5_UC5.3_FRQ07 38
T6.5_UC5.3_NFRQ04......39
6.7. T6.5_CA07 Timetable and Demand Data Collection 39
T6.5_UC5.4_FRQ01 39
T6.5_UC5.4_FRQ02 40
T6.5_UC5.4_FRQ03 40
T6.5_UC5.4_FRQ04 41
T6.5_UC5.4_NFRQ01......41
6.8. T6.5_CA08 Passenger Flow Analysis......42
T6.5_UC5.4_FRQ05 42
T6.5_UC5.4_FRQ06 42
T6.5_UC5.4_FRQ07 43
T6.5_UC5.4_FRQ08 43
T6.5_UC5.4_NFRQ02......44
7. Logical Architecture 45

7.1. High-level Architecture 45
7.2. Components and Functions 46
7.2.1 Data Manager 46
7.2.2 Machine Learning DI Model 47
7.2.3 Synchronization Component 47
7.2.4 Reporting Tool 48
7.2.5 Feedback Processor 48
7.2.6 Reporting Backoffice 49
7.2.7 Database 49
7.2.8 Optimization Software 50
7.2.9 Data Analysis Tool 50
7.3. Exchange Scenarios 51
7.3.1 UC-FP6-WP6-5.01 52
7.3.2 UC-FP6-WP6-5.02 53
7.3.3 UC-FP6-WP6-5.03 54
7.3.4 UC-FP6-WP6-5.04 55
8. Algorithm Description 56
8.1. Machine Learning Delay Impact Model (UC-FP6-WP6-5.01) 56
Classification Approach 56
8.2. Timetable Synchronization Analysis (UC-FP6-WP6-5.02) 57
8.2.1 Temporal Analysis 57
8.2.2 Spatio-Temporal Analysis 57
8.3. Traveller Feedback Processing (UC-FP6-WP6-5.03) 58
Confidence Analysis 58
8.4. MILP Model for Train Platforming Problem (UC-FP6-WP6-5.04) 59
9. Conclusions 62

Executive Summary

This deliverable is developed as part of Task 6.5 within Work Package 6 (WP6) of the EU-Rail FP6 FutuRe project, a strategic initiative aimed at revitalizing regional railway networks across Europe. WP6 is dedicated to the specification phase, laying the groundwork for the solutions that will later be implemented and demonstrated by WP11. Task 6.5 emphasizes passenger congestion monitoring, providing information that enables operators and authorities to explore different scheduling approaches and adapt them to various transport services.

This document specifically focuses on the passenger congestion monitoring specifications necessary to enhance the efficiency and service quality of regional railway lines. The goal of this deliverable is to create a comprehensive specification for a passenger congestion monitoring system and its associated algorithms. This system aims to enhance responsiveness to fluctuating passenger volumes in regional rail services.

The structure of this deliverable includes the following key components:

1. **Introduction (Section 1)**: A concise overview of the document's purpose, objectives, and scope.
2. **Objective/Aim (Section 2)**: Outlines the objectives and scope of this deliverable clearly and concisely.
3. **Background (Section 3)**: Overview of congestion and monitoring, highlighting key challenges and current solutions.
4. **Methodology (Section 4)**: The methodology used for the specification, focusing on the Architecture Analysis & Design Integrated Approach (ARCADIA).
5. **Use Cases (Section 5)**: Detailed scenarios illustrating various passenger flow and congestion situations to provide practical context for system application.
6. **System Requirements (Section 6)**: A thorough outline of both functional and non-functional requirements that the system must meet to be effective and reliable.
7. **System Components (Section 7.2)**: Granular descriptions of each component's functions and their alignment with system requirements, facilitating modular development and integration.
8. **Exchange Scenarios (Section 7.3)**: Elaborated interactions between system components, depicting data flows during different operational conditions to ensure accurate and efficient data processing.

9. **Algorithm Descriptions (Section 8)**: Specifications of algorithms designed to predict and manage passenger congestion using advanced data analytics and machine learning techniques.
10. **Conclusion (Section 9)**: Summarizes the specifications for the passenger congestion monitoring system and its impact on regional railways.

By defining these elements, this deliverable provides a robust framework for developing and implementing a passenger congestion monitoring system. This framework will guide subsequent development phases in WP11, ensuring the practical application of these specifications and thereby supporting the broader mission of the FutuRe project to revitalize regional railway networks across Europe.

List of Abbreviations, Acronyms, and Definitions

Abbreviation / Acronym	Definition
ARCADIA	Architecture Analysis & Design Integrated Approach
DI	Delay Impact
DRT	Demand Responsive Transport
FP	Flagship Project
FRQ	Functional Requirement
IVVQ	Integration, Validation, Verification, Qualification
KPI	Key Performance Indicator
MBSE	Model-Based System Engineering
MILP	Mixed-Integer Linear Programming
ML	Machine Learning
NFRQ	Non-functional Requirement
PRQ	Performance Requirement
QoS	Quality of Service
TCO	Total Cost of Ownership
TMS	Traffic Management System
TSP	Transport Service Provider
UC	Use Case
WP	Work Package

List of Figures

Figure 1 - Passenger flows on railway lines in Tomaszów Mazowiecki area. 15
Figure 2 - Use Case and Actor identification 19
Figure 3 - Passenger Congestion Monitoring high level architecture 45
Figure 4 - Exchange Scenario for UC-FP6-WP6-5.01 52
Figure 5 - Exchange Scenario for UC-FP6-WP6-5.02 53
Figure 6 - Exchange Scenario for UC-FP6-WP6-5.03 54
Figure 7 - Exchange Scenario for UC-FP6-WP6-5.04 55
Figure 8 - Example of synchronization analysis of train arrivals and bus departures 58
Figure 9 - MILP Model output visualization 61

List of Tables

Table 1 - System actors 19

Table 2 - List of System Capabilities 26

Table 3 - Data Manager component 46

Table 4 - Machine Learning Delay Impact (DI) model component 47

Table 5 - Synchronization Component..47

Table 6 - Reporting tool component 48

Table 7 - Feedback Processor component 48

Table 8 - Reporting Backoffice component 49

Table 9 - Database component 49

Table 10 - Optimization Software component 50

Table 11 - Data Analysis Tool component 50

1. Introduction

This deliverable is developed as part of Task 6.5 within Work Package 6 (WP6) of the EU-Rail FP6 FutuRe project, a strategic initiative aimed at revitalizing regional railway networks across Europe. WP6 is dedicated to the specification phase, laying the groundwork for the solutions that will later be implemented and demonstrated by WP11. This document specifically focuses on the passenger congestion monitoring specifications necessary to enhance the efficiency and service quality of regional railway lines and their interconnection with the main lines.

Regional railways, often seen as secondary networks, are vital for connecting European regions and acting as feeders for both passenger and freight traffic to the core railway network. They are essential not only for their environmental benefits but also for their integration with other public transport services such as buses, and first and last-mile services like bike-sharing, cycling, and walking. However, many of these lines have been neglected or abandoned in the past due to high operational costs, necessitating a strategic effort to renew and revitalize them. The FP6 FutuRe project seeks to ensure the long-term viability of these railways by reducing Total Cost of Ownership (TCO) while maintaining high service quality and operational reliability, thereby increasing customer satisfaction and making rail a preferred mode of transport.

Task 6.5 within this context addresses passenger congestion monitoring, focusing on regional lines and their connections to main lines, alongside the associated passenger flow. By gathering and processing relevant data, and through effective interaction with travellers, this task aims to provide valuable insights that will enable operators and authorities to explore and implement various scheduling approaches. This will help adapt services to fluctuating transport demands, thereby reducing congestion at stations and improving the overall Quality of Service (QoS).

The content of this deliverable includes detailed descriptions of use cases, system actors, system capabilities, requirements, high-level architecture, exchange scenarios, and algorithm descriptions. These elements collectively form the foundation for a robust passenger congestion monitoring system tailored to the unique challenges and opportunities of regional railways.

The deliverable is organized as follows:

- Section 2 states the aims of this deliverable, providing a clear overview of the objectives and scope.
- Section 3 offers background information regarding congestion and congestion monitoring, including key challenges and existing solutions.

- Section 4 details the methodology used, focusing on the Architecture Analysis & Design Integrated Approach (ARCADIA) and its application in system and functional aspects through modelling and defining use cases, requirements, and logical architectures.
- Section 5 presents various use cases that demonstrate practical applications related to passenger congestion monitoring.
- Section 6 specifies the capabilities and requirements necessary for effective implementation.
- Section 7 elaborates on the logical architecture, outlining the system components and their interactions.
- Section 8 focuses on the description of the algorithms addressing different elements, including delay prediction, timetable synchronization analyses, traveller feedback processing, and optimization for train platforming.
- Section 9 concludes by summarizing the specifications for the passenger congestion monitoring system and its impact on regional railways.

2. Objective/Aim

The objective of this deliverable is to create a specification for a passenger congestion monitoring system and its associated algorithms, aimed at enhancing responsiveness to fluctuating passenger volumes in regional rail services. This specification will cater to both real-time and predictive congestion management, addressing not only rail services but also their connections to main lines.

Accurate passenger congestion monitoring for regional rail lines and connected transport modes presents multiple advantages. Firstly, it facilitates a more responsive approach to varying passenger flows, allowing operators to adjust schedules and deploy additional services as needed. This leads to improved operational efficiency and cost savings. Secondly, it aids in the optimal allocation of resources, ensuring that transportation services are synchronized with actual passenger demand, thereby elevating overall service quality. Additionally, possessing real-time and predictive congestion data enhances decision-making during disruptions, providing a better basis for contingency planning. Ultimately, the system aims to improve passenger experience by minimizing waiting times and preventing overcrowding.

The methodology used in this deliverable incorporates advanced data analytics and model-based engineering methods. This involves developing Classical and Machine Learning (ML) models to predict passenger congestion, using historical and real-time data. The specifications will serve as a blueprint for developing a system that effectively addresses the unique challenges of regional rail services and their connections to main lines.

In summary, the establishment of a robust passenger congestion monitoring system under Task 6.5 will enhance operational efficiency, resource allocation, and user satisfaction. This deliverable sets the stage for the practical solutions to be implemented and demonstrated in WP11 of the FutuRe project, aligning with its mission to revitalize regional railways across Europe.

3. Background

Overcrowding and pedestrian congestion in public transportation occurs in connections and type of traffic where there is high demand and high density of population[1]. A factor that increases the likelihood of passenger congestion is short rush hours (narrow time interval when everyone in area serviced wants to get to schools and workplaces at the same time). Places, where passenger traffic from different origins and different modes of transport merges, intersects and recombines, like large railway stations, bus stations, intermodal and interchange nodes are natural places where such congestion will occur. Overcrowding in vehicles usually rises during journey when new passengers board at subsequent stops wanting to reach the final station during morning peak hours. And on the other hand, in the afternoon, overcrowding will decrease at each subsequent stops as you move away from the starting station in the local economic center.

Local and capillary lines, due to their location in a non-urbanized and sparsely populated area, are not likely to be subjected to this problem, as passenger flows are usually small and distributed more evenly throughout the day than in agglomeration and long-distance transport. However, it is important considering that capillary lines are feeders for main lines, and passengers arriving at interchange nodes are subject of traffic merge and recombination, too. In this case, information provided to passengers about seat occupancy in trains on main lines is a useful added value to the trip planning process, especially for occasional trips. Local, capillary, secondary and main lines should be treated as connected vessels, where the passenger stream spreads evenly to the "thickness" of the inflow to finally meet at the widest point of the connection network. So, it is important to treat the transport system as one organism, whose circulatory system consists of individual connections, and the heart of which will be an urban centre or agglomeration, where all public transport lines converge. Therefore, blocking (meaning: congestion) of one of the arteries affects the entire system and impedes the functioning of even its distant connections. Considering local lines in isolation from the fact that they are part of a larger system can lead to harmful simplifications and erroneous conclusions.

Between the 1990s and 2000s, several railway lines in this category were taken out of service in Poland, which are now being revitalized, rebuilt and restored as a minor but important part of the larger system. Experience collected during field research in central Poland carried out for Deliverable 6.2 confirms the above theses. In the exemplary location in Tomaszów Mazowiecki, local lines from Drzewica, Opoczno and Spała are converging, and then in Koluszki the passenger flow redirects to the main line towards Łódź, a supra-regional center and voivodeship capital city, and Warsaw, a center of national importance. Trains on the main line between Warsaw and Łódź are overcrowded in peak hours in both directions, because a lot of people from Łodź, due to its weak labour market, are commuting daily to Warsaw, and many young people from Tomaszów vicinity are studying in Łódź. On the passenger flow map in Figure 1, distribution of passenger flow

[1] Podoski J., "Transprt w miastach" (*Transport in cities*), Warszawa 1985

in Koluszki node can be seen. The width of line is proportional to daily passenger flow; teal – regional trains, blue – long distance trains. Łódź city is on the upper left corner of the map.

Figure 1 - Passenger flows on railway lines in Tomaszów Mazowiecki area.[2]

[2] Source: Passenger Traffic Model "PMT", 2019.

4. Methodology

The methodology used here is the Architecture Analysis & Design Integrated Approach (ARCADIA). It is becoming popular, especially while dealing with complex systems, equipment, software or hardware architecture definition. ARCADIA is a method applied to systems and architecture engineering, supported by an open-source Capella modelling tool. It can be practiced while dealing with complex systems, equipment, software or hardware architecture definition, often taking into consideration constraints such as performance, cost, safety, security.

The reasons to use this method are to[1]:

- understand the customer needs,
- define and share the product architecture among all relevant stakeholders,
- validate its design and justify it early,
- ease and master Integration, Validation, Verification and Qualification (IVVQ).

The methodology here is employed with an explicit focus on the system and functional aspects. The upcoming subsections delineate the key factors and steps involved in the methodology[2]:

1 Definition of Use Cases and Actors:
The initial stage ~~in~~of the methodology ~~involves defining~~ defines the use cases and actors of the system. Here the functionality of the system is outlined ~~in order~~ ~~-~~to support the process of use cases identification. Use cases represent the various interactions or functionalities that the system needs to perform. There are also defined actors who will interact with the system. This step provides a clear understanding of the system's functional requirements.

2 Capabilities and Requirements:
The next step ~~is to~~ ~~identify~~identifies the capabilities and requirements of the system. Capabilities represent the high-level functionalities that the system should possess, while requirements define the specific features and constraints that need to be met. Requirements and capabilities are reciprocal; it means each requirement has a matching capability. This step helps in establishing a clear set of goals and objectives for the system.

3 Logical Architectures:
When moving to the logical architecture one of the first activities to perform is to carry on the work performed at the system level. Logical architectures represent the structural organisation of the system, including the different components and their interactions. Such

step involves model’s creation in order to depict the system's structure, as well as enabling a better understanding of its overall design.

4 Exchange Scenarios (Per Use Case):
For further improvement of the system's functionality, exchange scenarios are developed ~~taking into account~~considering each Use Case. Exchange scenarios represent the sequence of interactions between the actors and the system components to achieve specific outcomes. These scenarios help in ~~capturing~~ catch the dynamic behaviour of the system and validating its functionality.

The Model-Based System Engineering (MBSE), which is the pillar of ARCADIA methodology, not only ensures that the information can be properly used to support systems engineering, it also provides a representation of information in different ways that show specific analysis capabilities for various types of users with different needs and interests. Further, it enables refining systems and system models into subsystem models[3].

[1] https://mbse-capella.org/arcadia.html

[2] https://www.sciencedirect.com/book/9781785481697/model-based-system-and-architecture-engineering-with-the-arcadia-method

[3] https://www.mdpi.com/2079-8954/11/8/429

5. Use Cases

This section provides the definition of Use Cases (UC) and actors within the context of the specification of passenger congestion monitoring algorithms in the regional rail scope. The purpose of this section is to establish a clear understanding of stakeholders involved in the development of the passenger congestion monitoring behaviour.

In the regional rail scope, congestion monitoring can play a crucial role in understanding and predicting passenger demands, optimizing resource allocation, and enhancing overall operational efficiency. Real time information about possible congestion during journey can help passengers to plan the trip in a flexible way. To ensure the successful development of these algorithms, it is essential to define the use cases and actors that will interact with the system.

The actors are the entities or stakeholders (persons or external systems) that interact with the congestion monitoring algorithms. Defining the actors helps identify their roles, responsibilities, and perspectives, ensuring the algorithms meet their needs.

The use cases represent the scenarios where the various functionalities are exploited and/or the interactions of the congestion monitoring system occur. They include information like the description of the use case, related project tasks/subtasks, involved actors, trigger, pre-conditions, input, result/requirement, sequence of steps, involved components and the responsible partner to develop the use case.

All the detailed descriptions of the identified use cases and actors should be considered in the regional rail domain.

The diagram in Figure 2 identifies the use cases developed in Task 6.5, and related use cases from other tasks. This provides a good overview of the actors and objectives of the work developed on this deliverable. Use cases specified in Task 6.5 are indicated by orange background colour, while use cases from other tasks are highlighted by green background colour. Supporting functions have white background.

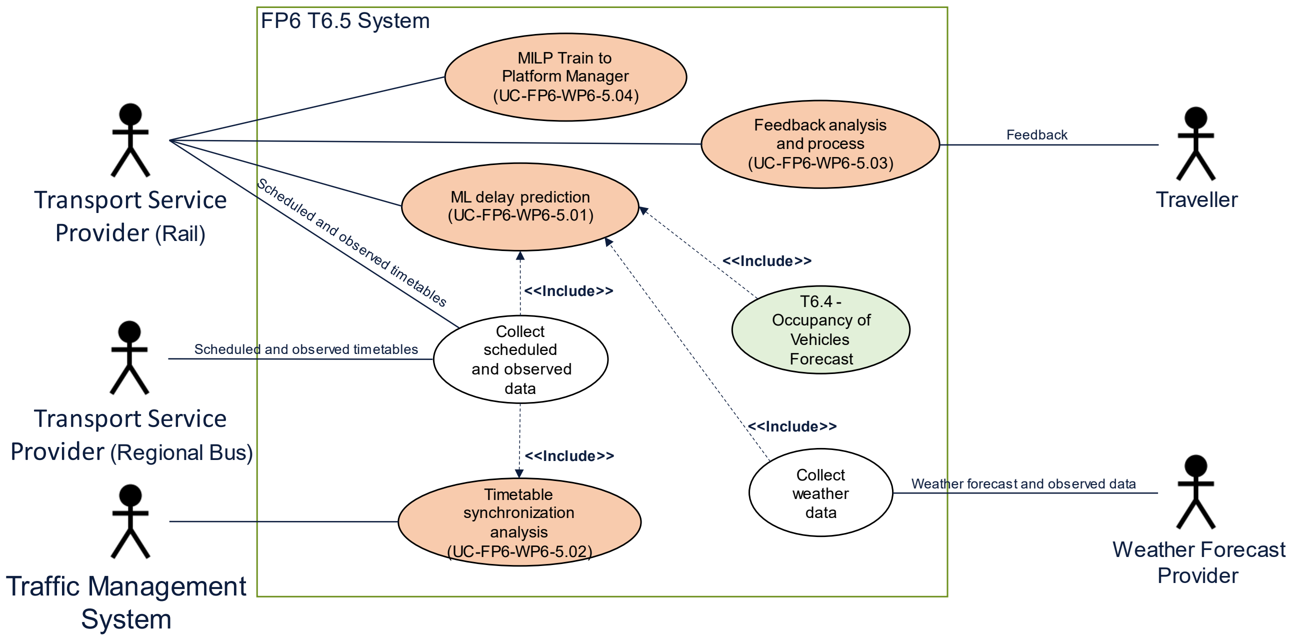

Figure 2 - Use Case and Actor identification

5.1. System Actors Identification

The following table provides the description of the actors involved in the use cases outlined in this section.

Table 1 - System actors

Actor	Description
Transport Service Provider (Rail)	Organization providing both Rail physical services and means of transport.
Transport Service Provider (Regional Bus)	Organization providing both Regional Bus physical services and means of transport.
Weather Forecast Provider	Organization providing data about weather conditions and forecast.
Traveller	The traveller is the person making or planning a travel. In this document, the terms Traveller and Passenger may be used interchangeably.
Traffic Management System (TMS)	A TMS is a technology platform used in the rail industry to manage and control train operations. It serves as a centralized system that provides real-time monitoring, control, and coordination of trains, tracks, and related infrastructure.

5.2. UC-FP6-WP6-5.01

Name	Impact of weather and train composition on train schedules and delays
ID	UC-FP6-WP6-5.01
Description	Using a Machine Learning model that integrates weather data (temperature, precipitation, wind speed, etc.) with train schedules and compositions, the goal is to predict potential delays caused by adverse weather conditions and its impact on congestion.
Related to task/subtask(s)	T6.5
Impact on other task(s)	T6.4
Interactions SP/FP	No interaction.
Actor(s)	TSP (Rail), Weather Forecast Provider.
Trigger	Scheduled, based on weather forecast data frequency.
Pre-Condition(s)	Historical weather data; Historical train schedule data; Historical train composition information; Observed train delays;
Input	Train scheduled data; Train composition information; Weather condition data (forecast and observed).
Result/Requirement	Characterization (e.g., expected impact, time delay, congestion period, station affected, ...) of a predicted delay, including the impact, in terms of forecasted congestion variation. Future delays and congestions which might result from current and predicted weather are forecasted and can be used to identify actions to avoid these congestions.
Sequence	1. Weather forecast data is collected depending on data source release frequency; 2. The ML model is applied to a dataset, which was updated with the latest weather data, predicting potential delays; 3. If a delay is predicted, the system should perform the characterization of forecasted congestion;

	4. If no delay is predicted, no output from the system is expected.
Involved components (System)	Data Manager, Machine Learning Delay Impact (DI) model
Responsible partner/person	GTSP
Notes	Real time train data and forecasted vehicle occupancy information (UC-FP6-WP6-4.02) would improve the accuracy of the results.

5.3. UC-FP6-WP6-5.02

Name	Synchronization between train and regional bus schedules
ID	UC-FP6-WP6-5.02
Description	The synchronization of interfacing operators' timetables highly affects congestion at stations and the perceived Quality of Service (QoS). This use case involves a detection system for timeframes where synchronization between interfacing operators could be improved. These detected timeframes can then be delivered to TMSs to further improve timetables synchronization.
Related to task/subtask(s)	T6.5
Impact on other task(s)	No impact.
Interactions SP/FP	No interaction.
Actor(s)	TSP (Rail), TSP (Regional Bus), TMS
Trigger	Periodically, depending on train and regional bus schedule updates.
Pre-Condition(s)	Availability of identified input data.
Input	Train scheduled and observed data; Bus scheduled and observed data.
Result/Requirement	The identified timeframes, where synchronization between train

	and bus schedules could be improved, are delivered to the TMSs.
Sequence	1. New train or regional bus schedule received; 2. Collect train scheduled and observed data; 3. Collect bus scheduled and observed data; 4. Analyse collected data to identify time gaps (timeframes) between interfacing services; 5. Deliver the identified timeframes to TMSs to further improve timetables synchronization.
Involved components (System)	Data Manager, Synchronization Component
Responsible partner/person	GTSP
Notes	The accuracy of these UC result will depend on the quality and completeness of the available data.

5.4. UC-FP6-WP6-5.03

Name	Traveller feedback for congestion analysis
ID	UC-FP6-WP6-5.03
Description	Using the traveller as a sensor, the system collects information about occupancy on train vehicles and stations. The collected information is then aggregated and processed, improving its accuracy and later usability. This processed information is made available for the TSP, empowering it with more information related to network congestion.
Related to task/subtask(s)	T6.5
Impact on other task(s)	No impact
Interactions SP/FP	No interaction.
Actor(s)	Traveller and TSP
Trigger	Periodically, depending on the number of traveller's feedback received.

Pre-Condition(s)	Available mobile application
Input	Traveller feedback
Result/Requirement	Processed congestion information based on traveller feedback is available for TSP.
Sequence	1. User sends feedback about the surroundings (train vehicle or station): a. After receiving a request for feedback notification. b. After the traveller takes the initiative to provide feedback; 2. After receiving the traveller feedback, the collected information is analysed and processed according to specific rules (location, service, …); 3. The processed information is made available for the TSP to use.
Involved components (System)	Feedback Processor, Reporting Tool
Responsible partner/person	GTSP
Notes	-

5.5. UC-FP6-WP6-5.04

Name	Train Platform Allocation Problem
ID	UC-FP6-WP6-5.04
Description	It is a system that efficiently assigns trains to railway station platforms using MILP optimization algorithms with the aim of optimizing operational efficiency, reducing the distance between connecting trains' platform and therefore improving the overall user experience in passenger hubs dealing with regional and main lines. Using MILP algorithms to model the optimal platforming problem, starting from a timetable and a geometric configuration of a railway station, it is computed the setting able to maximize the time available to passengers for inter-platform movement, also considering train punctuality.

Related to task/subtask(s)	T6.5
Impact on other task(s)	No impact
Interactions SP/FP	No interaction
Actor(s)	TSP (Rail)
Trigger	Periodically, depending on agreements with the infrastructure manager.
Pre-Condition(s)	Access to timetable, punctuality, and demand data; possibility of proposing new configurations to infrastructure manager.
Input	Timetable, punctuality data, infrastructure data and number of travellers.
Result/Requirement	Maximize the available transfer time for passengers.
Sequence	1. Acquisition of timetable data from GTFS format, demand and infrastructure data from various sources to the database. At the same time, acquisition of punctuality and platforming data in a data analysis tool; 2. Ingestion of the above-mentioned data in the optimization model and current scenario evaluation; 3. Run of the optimization algorithm; 4. Generation of the optimized scenario; 5. KPIs and visualizations for results' validation.
Involved components (System)	Database, Optimization Software, Data Analysis Tool
Responsible partner/person	Trenitalia
Notes	-

6. Capabilities and Requirements

Coherently with the methodology presented in section 4, in this section, the capabilities of the system are outlined. Each capability is identified by a capability ID in the format of T6.5_CAxy.

The system is seen as a set of components (described in section 7.2) necessary to accomplish the task of generating the passenger congestion monitoring information related to the usage of a transportation network.

The capabilities provide a concise overview of the high-level functions that the system must possess. By defining and understanding these capabilities, it is ensured that the system meets the task objectives and defined requirements for each of the capabilities, facilitating the development of the solution.

This Chapter includes also the requirements for each capability. These requirements are defined using a template for technical requirement descriptions. The template consists of the following fields:

- Requirement ID: A unique identifier for the requirement in the format T6.x_UCx.x_RequirementType0x.
- Requirement Name: The name of the requirement.
- Use Case ID: A link to the corresponding Use Case.
- Category: The category of the requirement.
- Priority: The priority of the requirement, either "MUST" or "Nice-to-have with high priority."
- Main goal: A detailed description of the requirement, explaining what it is and why it is needed.
- Assumptions: Any assumptions made for the requirement.
- Specification: A detailed description of the requirement and how it should work.
- Additional Notes: Any additional notes about the requirement, such as risks, dependencies, or constraints.

By including the requirements for each capability, it is ensured that the system is developed to meet the specific needs and objectives of the project. These requirements provide a detailed understanding of the functionality and performance expected from the system, guiding the development process effectively. Table 2~~Table 2~~ enumerates system capabilities alongside their associations with the UCs.

Table 2 - List of System Capabilities

UC ID	Use Case name	Capabilities
UC-FP6-WP6-5.01	Impact of weather and train composition on train schedules and delays	T6.5_CA01 Weather and train data collection T6.5_CA02 Predict schedule delays
UC-FP6-WP6-5.02	Synchronization between train and regional bus schedules	T6.5_CA03 Schedule data collection T6.5_CA04 Schedule synchronization analysis
UC-FP6-WP6-5.03	Traveller feedback for congestion analysis	T6.5_CA05 Traveller feedback collection T6.5_CA06 Feedback data aggregation and analysis
UC-FP6-WP6-5.04	Train Platform Allocation Problem	T6.5_CA07 Timetable and Demand Data Collection T6.5_CA08 Passenger Flow Analysis

6.1. T6.5_CA01 Weather and Train Data Collection

T6.5_UC5.1_FRQ01

Requirement ID	T6.5_UC5.1_FRQ01
Requirement Name	The system must be able to fetch train data from TMS and data from weather data sources to predict schedule delays.
Use Case ID	UC-FP6-WP6-5.01
Category	Functional
Priority	MUST
Main goal	The system must be able to gather relevant data to train and to apply the trained Machine Learning Model to predict delays in regional train services.
Assumptions	Availability of the data sources.
Specification	The system must be able to gather data relevant to predict schedule delays: • Weather data (forecast and observed weather conditions); • Train scheduled and observed data; • Train composition data.
Additional Notes	-

T6.5_UC5.1_FRQ02

Requirement ID	T6.5_UC5.1_FRQ02
Requirement Name	The system may be able to fetch forecasted occupancy data to improve predictions in schedule delays.
Use Case ID	T6.5_UC5.1_FRQ02
Category	Functional
Priority	Nice-to-have with high priority
Main goal	The system may be able to gather additional data to improve the training of the Machine Learning Model to predict delays in Regional Train services.
Assumptions	Availability of the additional data source.
Specification	The system may be able to gather additional data relevant to predict schedule delays: • Forecasted occupancy data.
Additional Notes	-

T6.5_UC5.1_PRQ01

Requirement ID	T6.5_UC5.1_PRQ01
Requirement Name	The system must be able to process the collected data within a reasonable time.
Use Case ID	UC-FP6-WP6-5.01
Category	Performance
Priority	MUST
Main goal	The processing time must be kept minimal for the output (predicted delays) of this module to be relevant and used by the TSP.
Assumptions	Reasonable processing hardware available.
Specification	The implementation of this processing module must be performed in an efficient manner, optimizing resource utilization, and minimizing processing time. A reasonable time frame for processing the collected data is less than an hour.
Additional Notes	The time measurement starts with the collection of data and ends when the output is available.

6.2. T6.5_CA02 Predict Schedule Delays

T6.5_UC5.1_FRQ03

Requirement ID	T6.5_UC5.1_FRQ03
Requirement Name	The system must be able to train a Machine Learning model using weather data, train schedule data and, optionally, forecasted occupancy data.
Use Case ID	UC-FP6-WP6-5.01
Category	Functional
Priority	MUST
Main goal	To predict delays, a ML model will be applied to the newly collected data. This ML model must be trained using relevant data.
Assumptions	The relevant data is available.
Specification	• The received data should be processed and transformed into a suitable format for training the model • The system should utilize appropriate machine learning algorithms and techniques to train the model based on the provided data • The trained model should be able to analyse and make accurate predictions based on training data
Additional Notes	-

T6.5_UC5.1_FRQ04

Requirement ID	T6.5_UC5.1_FRQ04
Requirement Name	The system must be able to retrain the machine learning model with newly collected observed data.
Use Case ID	UC-FP6-WP6-5.01
Category	Functional
Priority	MUST
Main goal	The output quality (predicted delays) may be improved with new data. New data will continuously be provided to update the ML model.
Assumptions	Newly collected observed data and previously collected is available.

Specification	• The system should have the capability to receive weather, train composition, planned schedule and observed schedule data, for updating the occupancy model. • The system should utilize appropriate techniques and algorithms to update the occupancy model based on the new and historical data. • The updated occupancy model should take into account the patterns and trends observed in the historical data, as well as the new data, to provide accurate predictions. • The updated delay prediction model should overwrite old models and be made available for future delay forecasting and analysis.
Additional Notes	-

T6.5_UC5.1_FRQ05

Requirement ID	T6.5_UC5.1_FRQ05
Requirement Name	The system must be able to apply the trained model to predict delays.
Use Case ID	UC-FP6-WP6-5.01
Category	Functional
Priority	MUST
Main goal	The system must be able to apply the trained model to collected data in order to predict delays.
Assumptions	The system must have a trained ML model and new relevant data.
Specification	• The received request should be processed and utilized by the trained model to predict delays
Additional Notes	-

T6.5_UC5.1_FRQ06

Requirement ID	T6.5_UC5.1_FRQ06
Requirement Name	The system must be able to store the predicted delays.
Use Case ID	UC-FP6-WP6-5.01
Category	Functional
Priority	MUST

Main goal	The system must have the capability to store predicted delays so that this information can be accessed later.
Assumptions	The system must have enough storage capacity to store the predicted delays for a year.
Specification	A continuous storage mechanism must be implemented in order to store the predicted delays.
Additional Notes	Increasing the storage capacity for a longer period may improve the output accuracy of the system.

T6.5_UC5.1_FRQ07

Requirement ID	T6.5_UC5.1_FRQ07
Requirement Name	The system must provide a specific interface for the predicted delays to be fetched.
Use Case ID	UC-FP6-WP6-5.01
Category	Functional
Priority	MUST
Main goal	The main goal of UC-FP6-WP6-5.01 is to empower the TSP with improved information about the operation. The TSP must be able to collect the latest predicted delays.
Assumptions	Predicted delays must have been stored and available to be accessed.
Specification	An external authenticated interface must be implemented to facilitate access to the output of this module. The external authenticated interface provides TSPs access to the delay predictions
Additional Notes	-

T6.5_UC5.1_NFRQ01

Requirement ID	T6.5_UC5.1_NFRQ01
Requirement Name	The system should keep the delay prediction model as updated as possible.
Use Case ID	UC-FP6-WP6-5.01
Category	Non- Functional
Priority	Nice-to-have with high priority

Main goal	The system should retrain the delay prediction model with the most recent data available periodically. This period should be kept minimal in order to provide the best predictions possible.
Assumptions	The system must have a trained ML model and new relevant data.
Specification	When sufficient amount of new data is available the model should be retrained The performance of the updated model is tested and compared to performance of the previous model After retraining the model, it is applied as the new default model if its performance exceeds the one from the old model
Additional Notes	-

6.3. T6.5_CA03 Schedule Data Collection

T6.5_UC5.2_FRQ01

Requirement ID	T6.5_UC5.2_FRQ01
Requirement Name	The system must be able to collect relevant data from Train's TMS and Bus Services.
Use Case ID	UC-FP6-WP6-5.02
Category	Functional
Priority	MUST
Main goal	The system must be able to collect relevant data to perform the identification of possible synchronization issues between timetables.
Assumptions	Availability of the data sources.
Specification	The system must be able to gather timetable data for: • Train services; • Bus services.
Additional Notes	The focus of this requirement lies in the data collection of the input data.

T6.5_UC5.2_PRQ01

Requirement ID	T6.5_UC5.2_PRQ01

Requirement Name	The system must be able to process the collected data in a reasonable time.
Use Case ID	UC-FP6-WP6-5.02
Category	Non- Functional
Priority	MUST
Main goal	The processing time must be kept minimal for the output of this module to be relevant and used by the TMS. The output should be available within an hour of receiving new input data.
Assumptions	Reasonable processing hardware available.
Specification	The implementation of this processing module must be performed in an efficient manner, optimizing resource utilization, and minimizing processing time.
Additional Notes	The measured process starts with the receival of new data and ends when the output is available.

6.4. T6.5_CA04 Schedule Synchronization Analysis

T6.5_UC5.2_FRQ02

Requirement ID	T6.5_UC5.2_FRQ02
Requirement Name	The system must be able to analyse the collected data and identify possible synchronization issues automatically.
Use Case ID	UC-FP6-WP6-5.02
Category	Functional
Priority	MUST
Main goal	The collected data must be analysed and processed so that synchronization issues can be identified.
Assumptions	The relevant data must have been collected.
Specification	The analysis is triggered by the arrival of new timetables The results of the analysis will be the identified timeframes where potential synchronization improvements can be made
Additional Notes	-

T6.5_UC5.2_FRQ03

Requirement ID	T6.5_UC5.2_FRQ03
Requirement Name	The system must be able to store the possible synchronization issues.
Use Case ID	UC-FP6-WP6-5.02
Category	Functional
Priority	MUST
Main goal	The system must have the capability to store the output so that this information can be accessed later.
Assumptions	The system must have enough storage capacity to store the output.
Specification	A continuous storage mechanism must be implemented in order to store the output.
Additional Notes	The data should be stored for at least 2 years to have two full annual cycles.

T6.5_UC5.2_FRQ04

Requirement ID	T6.5_UC5.2_FRQ04
Requirement Name	The system must be able to keep the identified synchronization issues available to be manually collected by the TMS.
Use Case ID	UC-FP6-WP6-5.02
Category	Functional
Priority	MUST
Main goal	The main goal of UC-FP6-WP6-5.02 is to empower the TMS with improved information about the operation. The TMS must be able to collect the identified synchronization issues.
Assumptions	The identified synchronization issues must have been stored and must be available to be accessed.
Specification	An external interface must be implemented to facilitate access to the output of this module.
Additional Notes	-

T6.5_UC5.2_PRQ02

Requirement ID	T6.5_UC5.2_PRQ02
Requirement Name	The system should be able to identify possible synchronization issues in reasonable time.
Use Case ID	UC-FP6-WP6-5.02
Category	Non- Functional
Priority	Nice-to-have with high priority
Main goal	The system should be able to identify possible synchronization issues in under an hour to be relevant and used by the TMS.
Assumptions	Reasonable processing hardware available and relevant data available.
Specification	The processing time must be kept minimal for the output of this module to be relevant and used by the TMS.
Additional Notes	-

6.5. T6.5_CA05 Traveller Feedback Collection

T6.5_UC5.3_FRQ01

Requirement ID	T6.5_UC5.3_FRQ01
Requirement Name	The system must be able to collect traveller feedback.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional
Priority	MUST
Main goal	The system must be able to collect traveller feedback so that it can be analysed and processed.
Assumptions	An interface with the traveller must exist and able to receive the traveller feedback.
Specification	A WebApp must be implemented to collect traveller feedback.
Additional Notes	-

T6.5_UC5.3_FRQ02

Requirement ID	T6.5_UC5.3_FRQ02

Requirement Name	The traveller must be able to send feedback via the WebApp at any time.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional
Priority	MUST
Main goal	The traveller must be able to send feedback at any time in order to improve traveller engagement and confidence on the interface.
Assumptions	The WebApp must be available to collect traveller feedback.
Specification	The WebApp is available via the most popular browsers at any point in time Unplanned downtimes should be minimized
Additional Notes	-

T6.5_UC5.3_FRQ03

Requirement ID	T6.5_UC5.3_FRQ03
Requirement Name	The system must have an interface for the TSP.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional
Priority	MUST
Main goal	The Reporting Backoffice provides an interface for the TSP to interact with the system.
Assumptions	The interface must be available. The TSP must be previously registered.
Specification	The Reporting Backoffice allows the TSP to visualize the reporting output and to request feedback from travellers based on a defined region.
Additional Notes	

T6.5_UC5.3_FRQ04

Requirement ID	T6.5_UC5.3_FRQ04
Requirement Name	The system should be able to request feedback from travellers.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional

Priority	MUST
Main goal	The TSP should be able to request traveller feedback in order to have a more complete knowledge of the network status.
Assumptions	Reporting Backoffice must be available and the WebApp must be able to receive the traveller feedback. The traveller must allow web browser notifications and location sharing for the WebApp, in order to receive notifications.
Specification	The TSP will have an interface (Reporting Backoffice) to notify users. The Reporting Backoffice enables the TSP to define regions in a map. All users within that region will receive the notification in the WebApp for user feedback.
Additional Notes	

T6.5_UC5.3_NFRQ01

Requirement ID	T6.5_UC5.3_NFRQ01
Requirement Name	The WebApp should be user-friendly and intuitive.
Use Case ID	UC-FP6-WP6-5.03
Category	Non- Functional
Priority	MUST
Main goal	To ensure that users can easily understand the interface and provide meaningful feedback.
Assumptions	The WebApp must be available to collect traveller feedback.
Specification	The WebApp adapts itself to the screen size of the device on which the WebApp is opened The WebApp is designed with an intuitive UI
Additional Notes	-

T6.5_UC5.3_NFRQ02

Requirement ID	T6.5_UC5.3_NFRQ02
Requirement Name	The WebApp should be compatible with most mobile devices.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional
Priority	MUST

Main goal	To guarantee the highest possible feedback rate, the WebApp must be compatible with most mobile devices.
Assumptions	The WebApp must be available to collect traveller feedback.
Specification	A WebApp must be implemented to minimize the obstacle of reaching travellers due to a wide range of mobile Operative Systems.
Additional Notes	-

T6.5_UC5.3_NFRQ03

Requirement ID	T6.5_UC5.3_NFRQ03
Requirement Name	The collected traveller feedback scope should focus on (trains and/or station) congestion.
Use Case ID	UC-FP6-WP6-5.03
Category	Non- Functional
Priority	Nice-to-have with high priority
Main goal	To guarantee the highest possible feedback rate, the scope of the interface should be focused on congestion.
Assumptions	The WebApp must be available to collect traveller feedback.
Specification	Travellers will provide information in a stepwise form. Questions will be about the amount of persons in trains and on stations e.g., “Are there many seats available?”, “Are passengers standing?”, etc.
Additional Notes	-

6.6. T6.5_CA06 Feedback Data Aggregation and Analysis

T6.5_UC5.3_FRQ05

Requirement ID	T6.5_UC5.3_FRQ05
Requirement Name	The system must be able to store, analyse and process the collected traveller feedback.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional
Priority	MUST

Main goal	To improve the knowledge about the network status, the collected traveller feedback must be analysed and processed based to specific parameters. Aggregating feedback from multiple travellers increases confidence in the resulting information.
Assumptions	The traveller feedback must have been collected.
Specification	Specific parameters can be time, location, date, user history, etc. The reliability of an aggregated feedback will increase with the number of travellers reporting the same feedback. A single feedback will still be a valuable information.
Additional Notes	-

T6.5_UC5.3_FRQ06

Requirement ID	T6.5_UC5.3_FRQ06
Requirement Name	The system must be able to store processed traveller feedback information.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional
Priority	MUST
Main goal	The system must have the capability to store the output so that this information can be accessed later.
Assumptions	The system must have enough storage capacity to store the output.
Specification	A continuous storage mechanism must be implemented in order to store the output.
Additional Notes	-

T6.5_UC5.3_FRQ07

Requirement ID	T6.5_UC5.3_FRQ07
Requirement Name	The Reporting Backoffice must provide an interface to facilitate access to the output for TSP.
Use Case ID	UC-FP6-WP6-5.03
Category	Functional
Priority	MUST

Main goal	The main goal of UC-FP6-WP6-5.03 is to empower the TSP with a more complete knowledge about the network status. The TSP must be able to collect the processed traveller feedback information.
Assumptions	The processed information must have been stored and available to be accessed.
Specification	An external interface must be implemented to facilitate access to the output of this module. Each TSP will only have access to the information for the area/stations they operate. The TSP will have access to the conducted feedback reporting analysis.
Additional Notes	

T6.5_UC5.3_NFRQ04

Requirement ID	T6.5_UC5.3_NFRQ04
Requirement Name	The system should keep the processed congestion information as updated as possible.
Use Case ID	UC-FP6-WP6-5.03
Category	Non- Functional
Priority	Nice-to-have with high priority
Main goal	The system should keep the processed congestion information as updated as possible. The processed congestion information should be processed within thirty minutes of receiving new traveller feedback data.
Assumptions	Traveller feedback collected.
Specification	The system should process the received traveller feedback as it is collected.
Additional Notes	-

6.7. T6.5_CA07 Timetable and Demand Data Collection

T6.5_UC5.4_FRQ01

Requirement ID	T6.5_UC5.4_ FRQ01

Requirement Name	The system must be able to collect demand data.
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	MUST
Main goal	The system must be able to collect real-time or modelled passenger flow data in a database to store it and make it available for the analysis.
Assumptions	Availability of demand data.
Specification	The system must be able to collect the data from the identified sources and store it in a database.
Additional Notes	-

T6.5_UC5.4_FRQ02

Requirement ID	T6.5_UC5.4_ FRQ02
Requirement Name	The system must be able to collect timetable data.
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	MUST
Main goal	The system must be able to collect data of a timetable scenario in a database to store it and make it available for the analysis.
Assumptions	Availability of timetable data.
Specification	The system must be able to collect the timetable and train platforming data and store it in a database.
Additional Notes	-

T6.5_UC5.4_FRQ03

Requirement ID	T6.5_UC5.4_ FRQ03
Requirement Name	The system should be able to consider train punctuality.
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	Nice-to-have with high priority

Main goal	The system should be able to consider train punctuality in the analysed station, store it in a database and make it available for the analysis.
Assumptions	Availability of train punctuality data.
Specification	The system should be able to consider train punctuality in the analysed station in order to conduct the most fitting analysis.
Additional Notes	-

T6.5_UC5.4_FRQ04

Requirement ID	T6.5_UC5.4_ FRQ04
Requirement Name	The system must be able to consider the station geometry and the platform configuration.
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	MUST
Main goal	The system must be able to consider the station geometry and the platform configuration, store it in a database and make it available for the analysis.
Assumptions	Availability of infrastructure data.
Specification	The system must be able to consider the station geometry and the platform configuration in the analysed station in order to conduct the most fitting analysis.
Additional Notes	-

T6.5_UC5.4_NFRQ01

Requirement ID	T6.5_UC5.4_ NFRQ01
Requirement Name	The system should be able to perform on different scenarios considering adapted input data updates, which are provided by railway undertakings
Use Case ID	UC-FP6-WP6-5.04
Category	Non- Functional
Priority	Nice-to-have with high priority
Main goal	The system should be able to simulate different scenarios and conduct gap analysis.

Assumptions	Generation of different scenarios or data update request are possible.
Specification	The system should be able to take into account time-sensitive data updates or different configuration scenarios.
Additional Notes	-

6.8. T6.5_CA08 Passenger Flow Analysis

T6.5_UC5.4_FRQ05

Requirement ID	T6.5_UC5.4_ FRQ05
Requirement Name	The system must be able to analyse and process the collected demand data.
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	MUST
Main goal	The system must be able to analyse and process the collected passenger flow data from the database and use it for the analysis.
Assumptions	Collected and stored demand data.
Specification	The system must be able to analyse and process the collected passenger flow data from the database and use it for the optimization first, then for the performance analysis.
Additional Notes	-

T6.5_UC5.4_FRQ06

Requirement ID	T6.5_UC5.4_ FRQ06
Requirement Name	The system must be able to analyse and process the timetable data.
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	MUST
Main goal	The system must be able to analyse and process the collected timetable data from the database and use it for the analysis.

Assumptions	Collected and stored timetable data.
Specification	The system must be able to analyse and process the collected timetable data from the database and use it for the optimization first, then for the performance analysis.
Additional Notes	-

T6.5_UC5.4_FRQ07

Requirement ID	T6.5_UC5.4_ FRQ07
Requirement Name	The system must be able to optimize the allocation of trains to platforms to allow maximum connection time for passengers
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	MUST
Main goal	Train platform allocation is optimized for maximizing the connection time for passengers
Assumptions	Relevant data is available
Specification	The system must be able to process the input data, run the optimization algorithm and provide the best platform configuration as result
Additional Notes	-

T6.5_UC5.4_FRQ08

Requirement ID	T6.5_UC5.4_ FRQ08
Requirement Name	The system must be able to evaluate congestion of passengers based on the optimized platform configuration.
Use Case ID	UC-FP6-WP6-5.04
Category	Functional
Priority	MUST
Main goal	The system must be able to evaluate congestion of passengers according to optimized platform configuration
Assumptions	Collected and stored timetable data, demand and infrastructure data are available

Specification	The system must be able to evaluate congestion of passengers according to optimized platform configuration by visualization and performance measurement tools.
Additional Notes	-

T6.5_UC5.4_NFRQ02

Requirement ID	T6.5_UC5.4_ NFRQ02
Requirement Name	The system should be able to rerun the process upon every change of the scenario or input data.
Use Case ID	UC-FP6-WP6-5.04
Category	Non- Functional
Priority	Nice-to-have with high priority
Main goal	The system should be able to efficiently rerun the process upon every update to obtain an optimized configuration every time.
Assumptions	The scenario or input data is changed.
Specification	The system should be able to efficiently rerun the process upon every update, in an automated way.
Additional Notes	-

7. Logical Architecture

7.1. High-level Architecture

This section explores the high-level architecture of the passenger congestion monitoring system within the FutuRe project. High-level architecture serves as the foundational blueprint that outlines the system's key components (described in section 7.2), their interactions, and the overarching framework that guides the development and integration processes.

By understanding this architecture, stakeholders can ensure cohesive integration and efficient operation, aligning with the project's goals. This architecture is directly based on the exchange scenarios provided in section 7.3, which detail the sequence of interactions and data flows necessary for system functionality.

The logical architecture diagram (Figure 3) use blue and white components to differentiate their roles. Blue components show external elements, such as systems or interfaces, not developed within this task's scope. White components represent the elements specifically developed for this deliverable, including core functionalities and modules, which will be detailed in the following section.

<

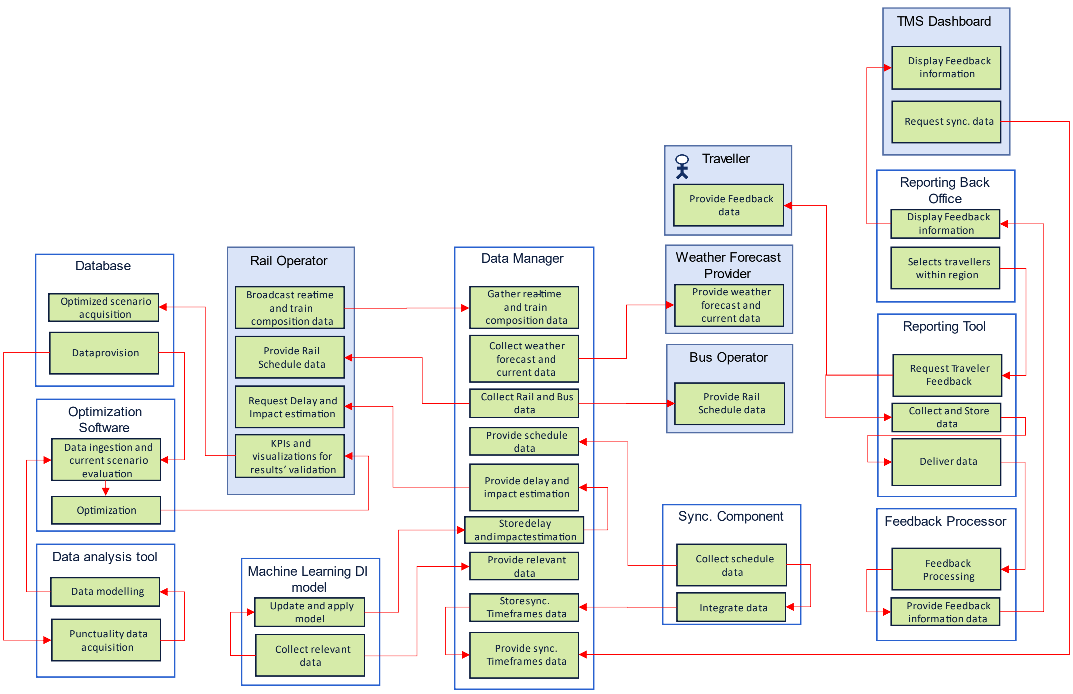

Figure 3 - Passenger Congestion Monitoring high level architecture

7.2. Components and Functions

This section describes each passenger congestion monitoring system component introduced in the high-level architecture. Each component's specific functions and traceability against with the system requirements are also outlined. This detailed description provides a clear understanding of the system's structure and functionality.

7.2.1 Data Manager

Table 3 - Data Manager component

Component Name	***Component Description***	***NFRQ ID***
Data Manager	Data Manager is responsible for collecting data from multiple sources, its preparation, storage, and availability for internal processes.	
Function Name	***Function Description***	***FRQ ID***
Gather real-time and train composition data	Collects both scheduled and real-time train timetable data as well as train composition data. It is also responsible for preparing and storing the collected data to ensure its availability for future processing	T6.5_UC5.1_FRQ01
Collect weather forecast and current data	Collects both current weather data and forecasts from a specified source. It is also responsible for preparing and storing the collected data to ensure its availability for future processing	T6.5_UC5.1_FRQ01
Provide relevant data	When requested, this function provides data related to both scheduled and real-time train timetable data, as well as train composition data and weather data.	
Store delay and impact estimation	After applying the ML model to the newly collected data, this function is responsible for storing the estimated delays and their impact.	T6.5_UC5.1_FRQ06
Provide delay and impact estimation	When requested, this function provides information related to estimated delays and their impact.	T6.5_UC5.1_FRQ07
Collect Rail and Bus data	Collects both train and bus schedule data. It is also responsible for preparing and storing the collected data to ensure its availability for future processing	T6.5_UC5.2_FRQ01
Provide schedule data	When requested, this function provides data related to train and bus schedules.	

Store sync. Timeframes data	After analysing the train and bus schedule data, this function is responsible for storing the identified synchronization issues	T6.5_UC5.2_FRQ03
Provide sync. Timeframes data	When requested, this function provides information related to possible synchronization issues.	T6.5_UC5.2_FRQ04

7.2.2 Machine Learning DI Model

Table 4 - Machine Learning Delay Impact (DI) model component

Component Name	***Component Description***	***NFRQ ID***
Machine Learning Delay Impact (DI) model	Machine Learning DI model applies the trained ML model to newly collected data. When sufficient data is collected, the ML model is retrained, tested and applied if its performance exceeds that of the previous model.	T6.5_UC5.1_PRQ01 T6.5_UC5.1_NFRQ01
Function Name	***Function Description***	***FRQ ID***
Collect relevant data	Requests data related to both scheduled and real-time train timetable data, as well as train composition data and weather data from Data Manager.	T6.5_UC5.1_FRQ01 T6.5_UC5.1_FRQ02
Update and apply model	The ML model is applied to newly collected data. When sufficient data is collected, the ML model is retrained, tested and applied if its performance exceeds that of the previous model.	T6.5_UC5.1_FRQ03 T6.5_UC5.1_FRQ04 T6.5_UC5.1_FRQ05

7.2.3 Synchronization Component

Table 5 - Synchronization Component

Component Name	***Component Description***	***NFRQ ID***
Synchronization Component	Synchronization Component is responsible for analysing the collected Train and Bus scheduled data and identifying possible synchronization issues.	T6.5_UC5.2_PRQ01 T6.5_UC5.2_PRQ02

Function Name	Function Description	FRQ ID
Collect schedule data	Requests data related to train and bus timetable data from Data Manager.	T6.5_UC5.2_FRQ01
Integrate data	Train and bus scheduled data is analysed in terms of identifying possible synchronization issues.	T6.5_UC5.2_FRQ02

7.2.4 Reporting Tool

Table 6 - Reporting tool component

Component Name	Component Description	NFRQ ID
Reporting tool	Reporting tool provides all mechanisms to request and receive feedback from travellers.	T6.5_UC5.3_NFRQ01 T6.5_UC5.3_NFRQ02 T6.5_UC5.3_NFRQ03
Function Name	**Function Description**	**FRQ ID**
Request traveller feedback	When requested, this function broadcasts a request for feedback to a selection of travellers.	T6.5_UC5.3_FRQ04
Collect data	This function is responsible for collecting feedback from travellers.	T6.5_UC5.3_FRQ01 T6.5_UC5.3_FRQ02 T6.5_UC5.3_FRQ05
Deliver data	After the feedback is collected by the Reporting Tool, it is delivered to the Feedback Processor to be processed.	

7.2.5 Feedback Processor

Table 7 - Feedback Processor component

Component Name	Component Description	NFRQ ID
Feedback Processor	Feedback Processor is responsible for aggregating and processing all feedback data from users. The resulting output information is delivered to the Reporting Backoffice to be presented to the end-user (TSP).	T6.5_UC5.3_NFRQ04

Function Name	Function Description	FRQ ID
Feedback Processing	The collected feedback from travellers is aggregated and processed in order to extract valuable information about the congestion current state of the transportation network.	T6.5_UC5.3_FRQ05 T6.5_UC5.3_FRQ06
Provide Feedback information data	Responsible for delivering the processed output information to the Reporting Backoffice to be presented to the end-user (TSP).	

7.2.6 Reporting Backoffice

Table 8 - Reporting Backoffice component

Component Name	Component Description	NFRQ ID
Reporting Backoffice	Reporting Backoffice is the interface with the end-user. It is on this interface that the end-user (TSP) can define regions to request feedback from travellers and to visualize the processed output information.	
Function Name	**Function Description**	**FRQ ID**
Selects travellers within region	This function provides the mechanisms for the end-user to define a region within the transportation network	T6.5_UC5.3_FRQ04
Display Feedback information	The processed output information is visually presented to the end-user (TSP).	T6.5_UC5.3_FRQ03 T6.5_UC5.3_FRQ07

7.2.7 Database

Table 9 - Database component

Component Name	Component Description	NFRQ ID
Database	This component is a data repository that collects train schedule, demand, infrastructure and historical punctuality data. It employs tables to organize information, ensuring data integrity and facilitating queries.	T6.5_UC5.4_NFRQ01 T6.5_UC5.4_NFRQ02

Function Name	Function Description	FRQ ID
Data provision	It provides all the necessary data and reorganizes it to facilitate processing by the other components.	T6.5_UC5.4_FRQ01 T6.5_UC5.4_FRQ02 T6.5_UC5.4_FRQ03 T6.5_UC5.4_FRQ04
Optimized scenario provision	It provides data of the optimized scenario and reorganizes it to facilitate post-processing and results' validation.	T6.5_UC5.4_FRQ01 T6.5_UC5.4_FRQ02 T6.5_UC5.4_FRQ03 T6.5_UC5.4_FRQ04

7.2.8 Optimization Software

Table 10 - Optimization Software component

Component Name	Component Description	NFRQ ID
Optimization Software	This software collects data processed as a mixed integer linear programming model. After running the optimization algorithms, it returns as an output the optimal platform allocation of trains in the analysed station.	T6.5_UC5.4_NFRQ02
Function Name	**Function Description**	**FRQ ID**
Data provision and current scenario evaluation	After collecting the data from the database, it processes and analyses the data to calculate the current passenger connection times and congestion on and besides platforms.	T6.5_UC5.4_FRQ05 T6.5_UC5.4_FRQ06 T6.5_UC5.4_FRQ07 T6.5_UC5.4_FRQ08
Optimization	It implements complex mathematical algorithms to determine the distance between platforms that returns optimal connection times, therefore the optimized platform allocation.	T6.5_UC5.4_FRQ05 T6.5_UC5.4_FRQ06 T6.5_UC5.4_FRQ07 T6.5_UC5.4_FRQ08

7.2.9 Data Analysis Tool

Table 11 - Data Analysis Tool component

Component Name	Component Description	NFRQ ID
Data Analysis Tool	The Data Analysis Tool provides and processes data input, through statistical modelling, to then provide delay analysis as an output.	T6.5_UC5.4_NFRQ02

Function Name	Function Description	FRQ ID
Punctuality data provision	After requesting data from the database, it processes the data to prepare it for modelling.	T6.5_UC5.4_FRQ03
Data modelling	After implementing statistical modelling techniques, it returns the train expected delay in the analysed station.	T6.5_UC5.4_FRQ05 T6.5_UC5.4_FRQ06

7.3. Exchange Scenarios

This section presents and describes an exchange scenario for each use case. Exchange scenarios are crucial for understanding the system's behaviour, illustrating how data flows between system components. By detailing these interactions, exchange scenarios provide valuable insights into the system's operational dynamics, supporting the implementation phase. They ensure that developers and stakeholders have a clear view of data exchanges, helping to identify potential issues and streamline the integration process. This foundational understanding is essential for building a robust and efficient congestion monitoring system, ensuring that all components work seamlessly together to meet the system requirements.

The diagrams shown (Figure 4 to Figure 7) illustrate the interactions between system actors and components. Actors with a blue background represent external system actors, while components with a white background represent system components. Interactions between functions are depicted with arrows: solid arrows indicate requests, and dotted arrows indicate responses.

7.3.1 UC-FP6-WP6-5.01

Upon receiving new scheduled and real-time train timetable data, as well as train composition data from the TSP, the Data Manager initiates communication with the Weather Forecast Provider to obtain observed and forecasted weather data. Subsequently, the Machine Learning DI model applies the current ML model to the collected data, predicting potential train operation delays and their associated impact. The resultant predictions are stored by the Data Manager for retrieval by the TSP upon request. This exchange scenario is illustrated in Figure 4.

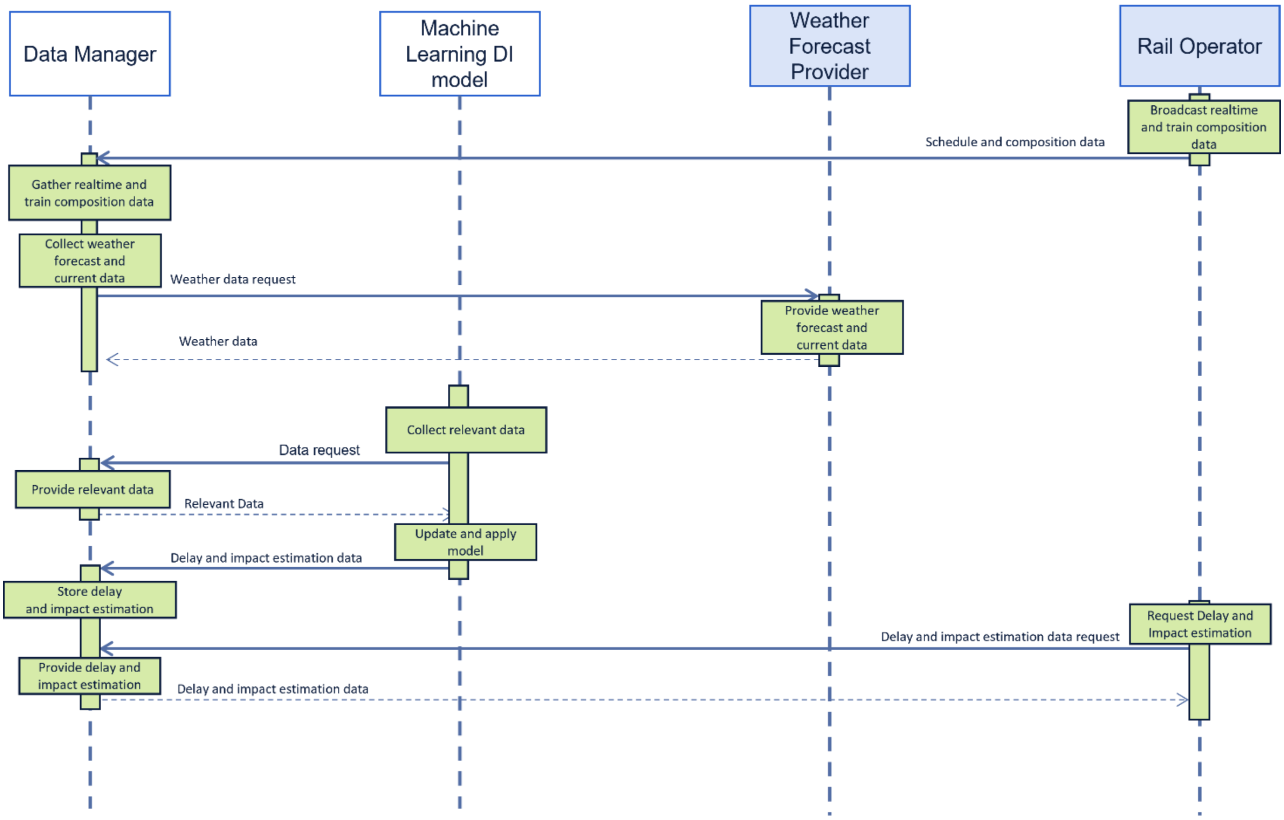

Figure 4 - Exchange Scenario for UC-FP6-WP6-5.01

7.3.2 UC-FP6-WP6-5.02

Data manager collects train and bus scheduled data from the respective TSP. This data is then delivered to the synchronization component to be analysed to identify potential synchronization issues. The resultant possible synchronization issues are stored by the Data Manager for retrieval and integration by the TMS Dashboard upon request. This exchange scenario is shown in Figure 5.

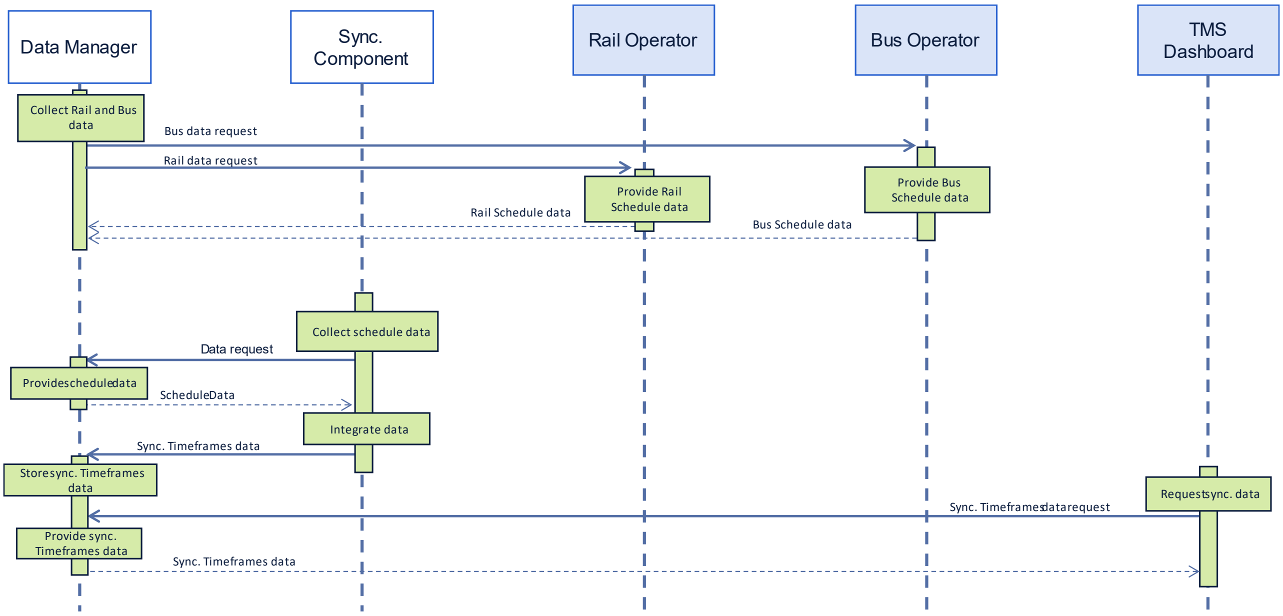

Figure 5 - Exchange Scenario for UC-FP6-WP6-5.02

7.3.3 UC-FP6-WP6-5.03

The feedback reporting process can be initiated by either the TSP or the traveller. In case the traveller decides to provide a feedback report, the traveller interacts with the Reporting Tool providing the requested information in a stepwise manner.

In case the TSP requests feedback from travellers, the TSP interacts with the Reporting Backoffice to select a region in the transportation network and send feedback requests to travellers in that area. Travellers then interact with the Reporting Tool as previously described.

All collected feedback from Reporting Tool is delivered to Feedback Processor for aggregation and processing to enhance its quality and accuracy. The processed feedback is then displayed in the Reporting Backoffice for the TSP to review. This exchange scenario is illustrated in Figure 6.

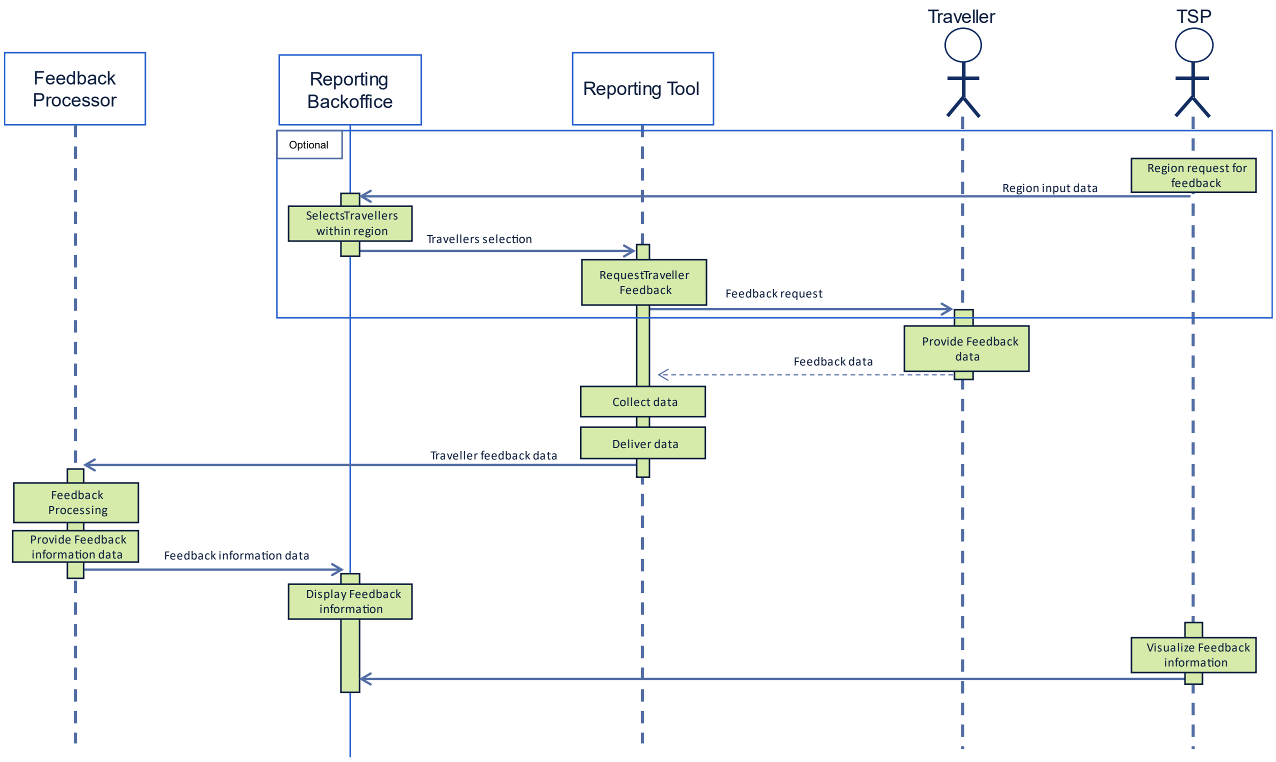

Figure 6 - Exchange Scenario for UC-FP6-WP6-5.03

7.3.4 UC-FP6-WP6-5.04

In this use case, a database is fed with timetable data, demand and infrastructure data. At the same time, provision of punctuality and platforming data are provided to a data analysis tool for modelling.

After that, the database provides the optimization software with the above-mentioned data and the optimization software valuates the current scenario, to then run the optimization algorithm. The optimization software generates the optimized platform allocation. Results' validation is done through Key Performance Indicator (KPIs) and visualization. After validation, the optimized scenario is stored in the database. This exchange scenario is shown in Figure 7.

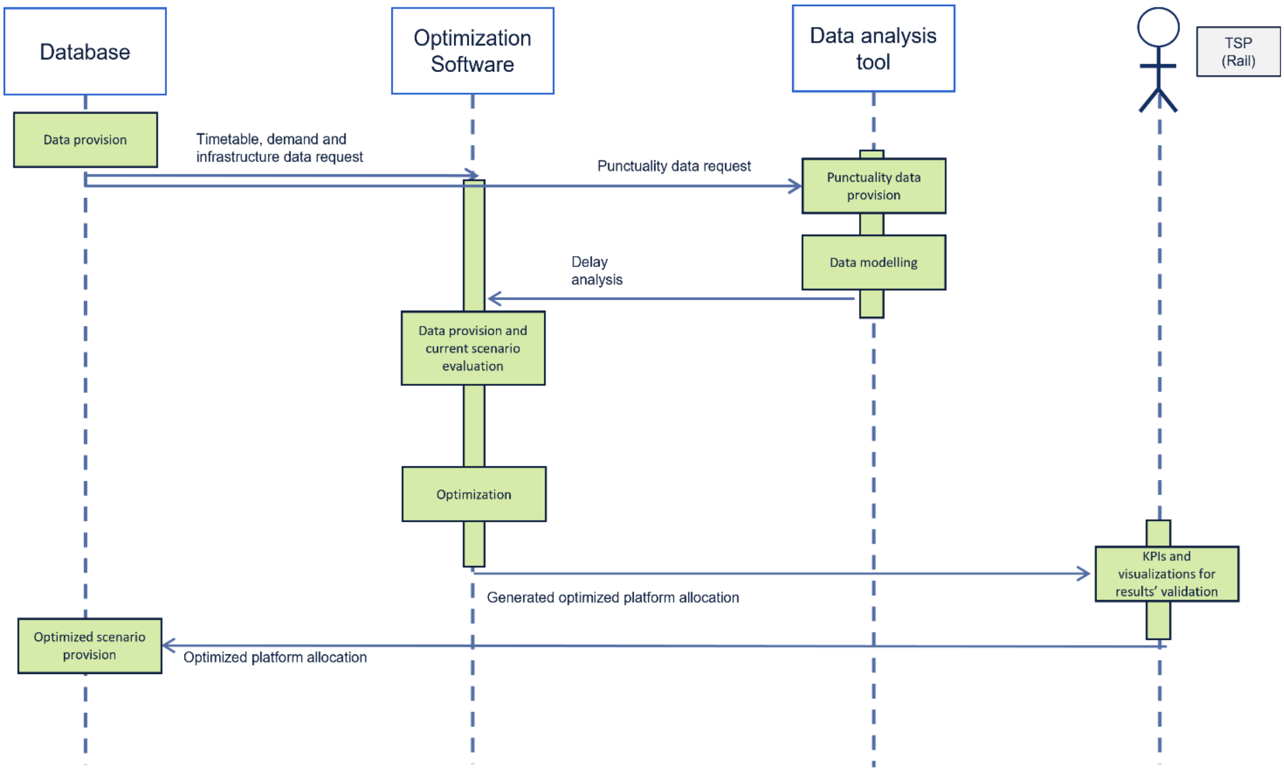

Figure 7 - Exchange Scenario for UC-FP6-WP6-5.04

8. Algorithm Description

The following sub-sections detail the expected methods to be used to achieve the objectives addressed by each use case, covering machine learning models for delay prediction, timetable synchronization analyses, traveller feedback processing, and optimization algorithms for train platforming.

8.1. Machine Learning Delay Impact Model (UC-FP6-WP6-5.01)

Train delays are a common issue influenced by various external factors, with weather conditions playing a significant role. The objective of this modelling approach is to predict potential train delays by integrating weather data with train schedules and historical delay records. By using a classification algorithm, the goal is to forecast delays that exceed specific thresholds, such as 5 or 10 minutes, and gather sufficient information to characterize these delays and perform a subsequent impact analysis.

Classification Approach

The classification approach focuses on predicting the likelihood of train delays under adverse weather conditions using a Random Forest classifier, which is a widely used machine learning algorithm that belongs to the family of ensemble learning methods. It works by building multiple decision trees and combines their outputs to improve prediction accuracy and handle complex data interactions. Specifically, Random Forest is expected to be well-suited for this task as it can manage the non-linear relationships between multiple factors, such as different weather conditions and train schedules.

The dataset comprises weather data (temperature, precipitation, wind speed, etc.), train schedules, and historical delay records. The key steps include:

- Temporal and geospatial alignment, synchronizing weather data with train schedules, and mapping them to corresponding stations and routes.
- Feature engineering to create features from weather data, such as average temperature and total precipitation, train schedules, including scheduled departure and arrival times and historical punctuality, and temporal features to address the differences between weekdays and weekends.

As the model learns the patterns within the data, it is also expected to extract feature importance insights from the Random Forest algorithm. These insights will help identify the factors that most influence delays, leading to a more interpretable impact analysis.

The impact analysis aims to understand the variation in congestion that might occur due to delays. By examining train composition details, it is possible to assess whether a delay during a particular timeframe might cause a station to become busier than usual. For example, if a delayed train with a high passenger capacity arrives at a station during peak hours, it could lead to overcrowding on platforms and in subsequent trains. In contrast a delay during off-peak hours might have a minimal impact on congestion.

8.2. Timetable Synchronization Analysis (UC-FP6-WP6-5.02)

Efficient synchronization between train and bus schedules is crucial for minimizing commuter wait times and enhancing the overall QoS. In this analysis, based on data granularity and precision, a two-stage approach will be undertaken in order to identify potential synchronization issues. The first stage comprises a temporal analysis, while the second stage is based on a spatio-temporal analysis.

8.2.1 Temporal Analysis

This stage focuses only on the temporal aspect by examining the time differences between scheduled train arrivals and bus departures, and vice versa. By calculating and analysing these time differences, the goal is to identify instances where significant gaps or overlaps exist, leading to potential synchronization issues. This initial analysis will provide a baseline understanding of how well train and bus schedules are coordinated in terms of time.

8.2.2 Spatio-Temporal Analysis

Depending on data granularity and precision, this stage will extend the analysis by incorporating spatial factors, such as the distance between train and bus stations. To achieve a better understanding of synchronization issues, it is planned to employ a direct lag analysis approach. This method involves calculating the time difference between train arrival times and adjusted bus departure times, where adjustments are made to account for the walking time between the stations.

By computing these time differences, or lags, it is possible to identify potential synchronization issues. Negative lags indicate that buses depart before trains arrive, while positive lags can be interpreted as the waiting times for passengers.

For illustration purposes, the plot in Figure 8 shows the train arrivals, original bus departures, and adjusted bus departures (taking into account walking times). The black annotations highlight the lags, or waiting times, between train arrivals and adjusted bus departures, which helps us identify potential synchronizations issues.

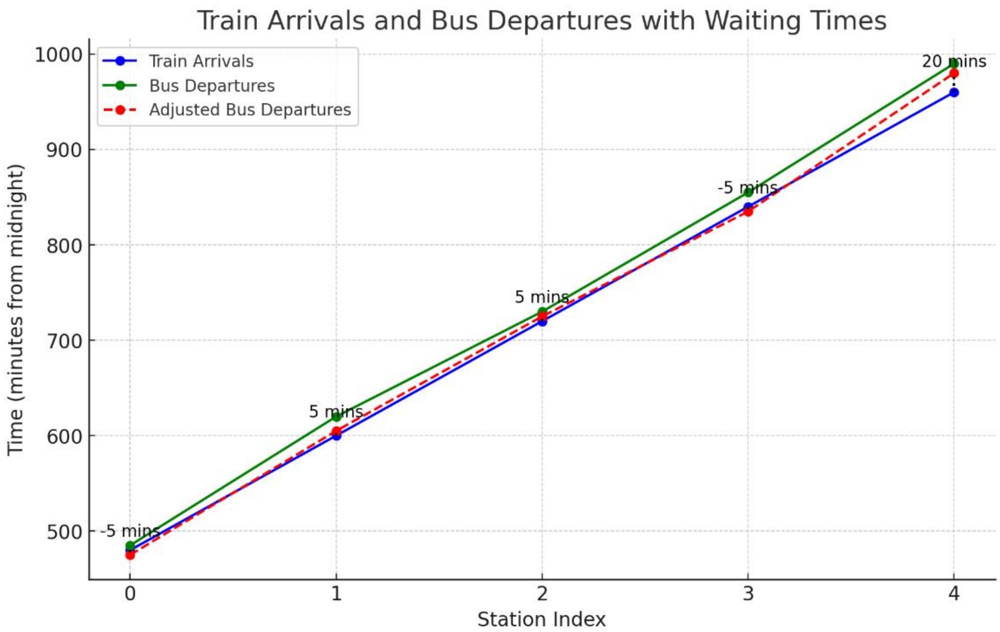

Figure 8 - Example of synchronization analysis of train arrivals and bus departures

8.3. Traveller Feedback Processing (UC-FP6-WP6-5.03)

A reliable evaluation of traveller feedback plays an important role for ensuring the accuracy of reported occupancy data. In this analysis, it is foreseen to calculate a confidence score for each report, based on the proximity of the report's coordinates to the point of interest.

Confidence Analysis

The methodology for confidence analysis involves various key steps to calculate the confidence score, ensuring that accurate and relevant reports are identified:

1. **Geographical instance creation:**
    a. **Report coordinates:** extract the geographical coordinates of the report, representing a point.
    b. **Region of interest coordinates:** define the geographical coordinates of the region of interest, representing a polygon.

2. **Score assignment process:**

a. **Exact match (score = 1):** if the report's coordinates fall within the region of interest polygon, the confidence score is assigned a value of 1.
b. **Outside threshold (score = 0):** if the distance between the report's coordinates and the region of interest polygon exceeds a predefined threshold (default is 2 km), the confidence score is set to 0.
c. **Within threshold:** for reports that are outside the polygon but within the threshold distance, the score is calculated using the formula:

$$Score = 1 - (\frac{D_p}{D_t})$$

Where D_p is the measured distance between the report location and the region of interest, and D_t is the threshold distance, the maximum allowed distance for a report to be considered reliable.

The result of this confidence calculation analysis is a confidence score ranging from 0 to 1, representing the reliability of the report based on its proximity to the point of interest. This score will be used to filter and prioritize traveller feedback, ensuring that accurate data can be utilized for occupancy analysis and network congestion assessment.

To further improve the confidence level of the reports, a subsequent analysis can be conducted, taking into account the number of users reporting the same feedback. This can be used as a weighting factor to calculate the overall confidence score, where each additional report increases the score incrementally. By aggregating similar reports from multiple users, the overall confidence in the reported data can be increased. This approach leverages the principle of consensus, where a higher number of consistent reports indicates higher reliability.

8.4. MILP Model for Train Platforming Problem (UC-FP6-WP6-5.04)

The goal of the **optimization algorithm** is to maximize the total time available for passengers to transfer between connecting trains by optimizing the placement of trains on the tracks. The optimization algorithm solves the optimization problem that is modelled as a mixed-integer linear program (MILP). The MILP comprises an objective function and constraints (also called boundary conditions). The objective function reads as follows.

$$\boldsymbol{max} \sum_{ij} \boldsymbol{p_{ij}} \, (\boldsymbol{t_{ij}} - \boldsymbol{T_{ij}}) \, \boldsymbol{w_{ij}}$$

The objective function is composed of a summation over indices *i* and *j*, that vary within the set of all trains; each term of the summation is the product of:

- p_{ij} is the number of passengers who change between train *i* and train *j* (predefined value);
- t_{ij} is the scheduled connection time between train *i* and train *j* (predefined value);
- T_{ij} is the time needed for passengers to move from train *i* to train *j* (decision variable);
- w_{ij} is the value of the connection between train *i* and train *j*, the latter value is directly proportional to the number of passengers (predefined value).

The difference between the term t_{ij} and T_{ij} is the **total connection time** that needs to be maximized.

The boundary conditions aim to define a scenario that respects the logistic constraints of space and time. These conditions serve to ensure that the time needed for passengers to move from one platform to another is always minor than the time that elapses between the arrival of the first train and the departure of the second. The time necessary for passengers is defined by considering all the possible train/platform combinations assuming a passenger walking speed of *3.5 km/h*.

The scheduled connection time t_{ij} is adjusted by considering delay associated with the connecting trains, obtained from a probability distribution based on punctuality data. To model delays, an empirical probability distribution was created using historical punctuality data for each train passing through the analysed station.

The algorithm produces an optimized platform allocation with occupancy times, enhancing connection efficiency for passengers.

The following diagram is an example of model output visualization, showing the optimized track allocation with occupancy time, where the x-axis represents the time of the day and the y-axis the platform name of the analysed station, and each colour is associated to a train brand.

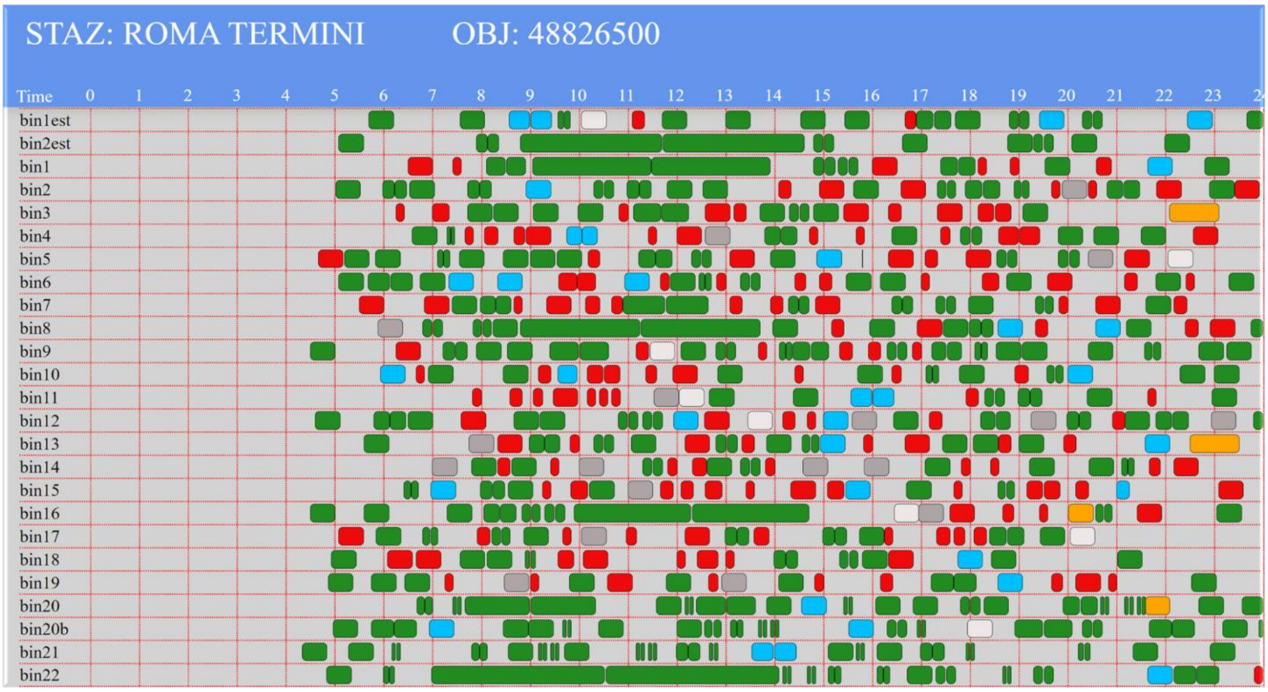

Figure 9 - MILP Model output visualization

This algorithm effectively balances train scheduling and passenger transfer times within stations that connect main and regional lines, resulting in an improvement of customer satisfaction.

9. Conclusions

This deliverable, developed as part of Task 6.5 within WP6 of the EU-Rail FP6 FutuRe project, provides a comprehensive specification for the passenger congestion monitoring system aimed at revitalizing regional railway networks across Europe.

A detailed approach was employed to define use cases, system requirements, system components, and exchange scenarios, forming a comprehensive blueprint for subsequent development, implementation, and demonstrations.

The defined use cases illustrate various scenarios related to passenger congestion, providing a practical context for the system's application. These use cases are crucial for understanding the specific challenges and operational contexts that the congestion monitoring system must address. They serve as the foundation upon which the system's functionalities are built.

System requirements have been outlined to ensure that the congestion monitoring system meets its intended objectives. These requirements encompass both functional and non-functional aspects, detailing what the system must do and the conditions it must meet to be effective and reliable.

In defining the system components, a granular view of the various elements that will constitute the congestion monitoring system, was provided. Each component's function has been described, highlighting its role and its alignment with the specified requirements.

Exchange scenarios have been elaborated to depict the interactions between different system components and the data flows that occur during various operational conditions. These scenarios are essential for understanding the dynamic behaviour of the system, ensuring that data is accurately and efficiently processed and exchanged to manage passenger congestion.

The definition of algorithms is a critical aspect of this deliverable. By leveraging advanced data analytics and machine learning techniques, these algorithms are designed to predict and manage passenger congestion with high accuracy. They process historical and real-time data to provide actionable insights, enabling operators to make informed decisions. These algorithms form the intelligence behind the system, driving its capability to adapt to varying passenger flows and improving the overall QoS.

In summary, this deliverable provides a thorough and detailed framework for the passenger congestion monitoring system, setting the foundation for the development, implementation and demonstration phase in WP11. By addressing the unique challenges of regional railways and enhancing their operational capabilities, this system will contribute significantly to the FutuRe project's mission of revitalizing Europe's regional railway networks.

End-of-document