D32.2

Recommendations to reduce technical barriers and improve the quality and availability of data

Project acronym:	FP5 TRANS4M-R
Starting date:	2022-07-01
Duration (in months):	45
Call (part) identifier:	HORIZON-ER-JU-2022-01 (Topic: HORIZON-ER-JU-2022-FA5-01)
Grant agreement no:	101102009
Due date of deliverable:	Month 28
Actual submission date:	2024-10-31 (M28)
Responsible/Author:	Sneha Gosavi (LSP)
Dissemination level:	PU
Deliverable Type:	R – Document, report
Doc Version & Status:	V3.0 | Submitted

Reviewed: (Yes)

"Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Europe's Rail Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them."
"The project is supported by the Europe's Rail Joint Undertaking and its members."

Document history		
Revision	*Date*	*Description*
V1	2024-02-24	First issue
V1.1	2024-05-28	Second Issue
V1.2	2024-08-14	Version sent for internal review
V2.0	2024-10-23	Final version with integrated comments from review
V3	2024-11-05	Submitted version

Report contributors		
Name	*Beneficiary Short Name*	*Details of contribution*
Sneha Gosavi	LSP	Responsible author Executive Summary Background and Objective Survey Conclusions
Nicklas Blidberg	TRV	Expert support
Maria Esther Ruiz-Capillas Mónica Pelegrín	ADIF	Expert support
Petr Šohajek	OG	Consolidation of survey results
Martin Kjellin Eddie Olsson	RISE	Data quality management Data quality and availability Data governance
Rodrigo de la Iglesia Sánchez Francisco Parrilla Ayuso	INDRA	Data standardization and interoperability Implementation plan
Felix Hildebrandt	HACON	Stakeholder survey Internal review
Saioa Arrizabalaga, Nasibeh Mohammadzadeh, Yasiru Rathsara	CEIT	Data spaces, IDS-RAM 4, Identity and access management, data auditing and monitoring, data sharing policies
Thomas Pum	OEBB RCA	Expert support

Reviewers		
Name	*Beneficiary Short Name*	*Details of contribution*
Yves Sterbak	GTS	Official Review and Feedback
De Blasio Giovanni	RFI	Official Review and Feedback

Disclaimer

The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The content of this document reflects only the author's view – the Joint Undertaking is not responsible for any use that may be made of the information it contains. The users use the information at their sole risk and liability.

The content of this report does not reflect the official opinion of the Europe's Rail Joint Undertaking (EU-Rail JU). Responsibility for the information and views expressed in the report lies entirely with the author(s).

Table of Contents

List of Figures

List of Tables

1 Executive Summary

This deliverable, D32.2 "Recommendations to reduce technical barriers and improve the quality and availability of data", is a critical output of Work Package 32 within the Flagship Project FP5 - TRANS4M-R. It builds upon the foundations laid in the previously submitted deliverable 25.1, which introduced general principles for data sharing and data structures in the rail freight sector. The primary objective of this document is to provide comprehensive recommendations for removing administrative and technical obstacles to data exchange, thereby enhancing data availability and quality across the multimodal freight transport ecosystem. This work is essential for establishing a harmonized framework for seamless data exchange within FP5 and its connection to FP1.Key aspects of this deliverable include:

- Further exploration of the Conceptual Data Model (CDM) introduced in D25.1, emphasizing its role in ensuring data interoperability and quality.
- Analysis of survey results from key stakeholders, including Infrastructure Managers (IMs) and Railway Undertakings (RUs), as well as System Suppliers highlighting current challenges and the state of data exchange in the sector.
- Detailed examination of the roles, responsibilities, and systems involvement of IMs (such as RFI, TRV, and ADIF) and RUs (like ČDC and ÖBB) and system suppliers (like HACON and GTS) in various aspects of railway operations, including planning, traffic control, and data exchange.
- Identification of critical issues in data exchange, such as lack of standardization, cybersecurity concerns, and difficulties in real-time data sharing.
- Discussion on the key aspects that are essential for data sharing such as identity and access management, auditing, monitoring and data sharing policies through a data space approach, as specified by International Data Spaces.
- Recommendations for improving data exchange, focusing on areas such as data governance, standardization, and the adoption of new technologies.
- Discussion on the importance of data quality and availability in optimizing railway operations and improving service quality.
- Coverage of data storage solutions and transversal aspects related to them, including recommendations for good practices in safety, security, and data protection.
- Collection and examination of use cases from WP26 to WP31 to identify areas requiring alignment with WP32 for data exchange purposes.

This deliverable aims to provide a comprehensive framework that will facilitate the seamless and harmonized exchange of data within the broader context of multimodal freight transport. By addressing the identified challenges and implementing the proposed recommendations, the project seeks to drive efficiency, transparency, and collaboration across systems, ultimately leading to optimized logistics, better resource management, and enhanced customer experiences in the rail freight sector. The document reflects the

process followed to align data exchange requirements across various work packages, ensuring a cohesive approach to data management and interoperability throughout the FP5-TRANS4M-R project.

Keywords - Data exchange, Multimodal freight transport, Conceptual Data Model (CDM), Interoperability, Standardization, harmonization

2 Abbreviations & Acronyms

Abbreviation / Acronym	Description
DAS	Direct Attached Storage
EU-Rail	Europe's Rail Joint Undertaking
ERC	European Railway Checkpoint
ETA	Estimated Time of Arrival
EVN	European Vehicle Numbering
FP	Flagship Project
API	Application Programming Interface
CA	Certification Authority
CDM	a) Conceptual Data Model b) Collaborative Decision Making
DAPS	Dynamic Attribute Provisioning Service
DAT	Dynamic Attribute Token
DTM	Dynamic Trust Monitoring
FDFTO	Full Digital Freight Train Operation
GCU	General Contract of Use for Wagons
JSON	JavaScript Object Notation
IDSA	International Data Space Association
IDS-RAM	International Data Space – Reference Architecture Model
ILU	Intermodal Loading Unit
IM	Infrastructure Manager
JAAS	Java Authentication and Authorization Service
NAS	Network Attached Storage
ODRL	Open Digital Rights Language
ParIS	Participant Information Service
PAP	Policy Administration Point
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PIP	Policy Information Point
PXP	Policy Execution Point
RFID	Radio Frequency Identification
RID	Regulations concerning the International Carriage of Dangerous Goods by Rail
RNE	RailNetEurope
RU	Railway Undertaking
TAF	Telematic Applications for Freight
TSI	Technical Specification of Interoperability
TT	Transversal Topics
UC	Use Case

UN	United Nations
XACML	eXtensible Access Control Markup Language
WP	Work package

3 Background

The present document constitutes the Deliverable D32.2 “Recommendations to reduce technical barriers and improve the quality and availability of data” in the framework of the Flagship Project FP5 – TRANS4M-R as described in the EU-RAIL MAWP and contributes as well to the Flagship Project FP5 – TRANS4M-R as described in the EU-RAIL MAWP.

This deliverable is a critical output of work package 32, which aims to establish a comprehensive framework to facilitate the seamless and harmonized exchange of data within the broader context of multimodal freight transport. This framework is essential not only for the internal activities of Flagship area 5 but also for ensuring seamless connectivity with Cluster 1: FDFTO

In deliverable 25.1 Seamless Freight Specifications, the specifications and use case definitions for Seamless Data Exchange were introduced.

This deliverable constitutes the work from tasks 32.1, 32.2, 32.4. The overall objective of work package 32 - Specification of Seamless data availability/exchange is to enable the introduction of technical enabler.

4 Objectives

4.1 Link between GA objectives and chapters in the Deliverable

The table below links the tasks described in the Grant Agreement for the related WP32 objectives to the respective chapters of this deliverable that feature these elements.

Task (GA)	Task definition from GA (WP32)	Output of deliverable
32.1	General principles for Data Sharing and Data Structure (M45) First and basic approach	Chapter 6 and 7
32.2	Defining data exchange (M24) for use cases	Chapter 9
32.4	Conceptual data model - alignment with FA1 TT (M45) First and basic approach	Chapter 8

5 Introduction

The primary objective of Deliverable 32.2 is to offer thorough suggestions for removing administrative and technical obstacles to data exchange, hence enhancing data availability and quality throughout the multimodal freight transport ecosystem. This deliverable, which is a part of WP32, is essential to creating a harmonized framework for seamless data exchange inside FP5 and its link to FP1.

To assist this data-sharing process, Flagship Area 1 is developing and implementing a Conceptual Data Model (CDM), which guarantees that data represents the same concepts for the entities exchanging the information and helps gaining interoperability when transferring the data, hence increasing the data quality as it prevents from data misunderstanding problems._This study expands on previous projects that focus on lowering barriers and facilitating interoperability among systems by addressing the technology obstacles and standardization requirements.

Data - its collection, processing, and interpretation are fundamental. Data exchange is a powerful tool that can drive efficiency, transparency, and collaboration across systems. In the context of multimodal transport, for example, seamless data exchange can lead to optimized logistics, better resource management, and enhanced customer experiences.

5.1 Survey Results and Analysis of Stakeholder Involvement

To gain insights into the current challenges and status of data exchange within the rail freight sector, we conducted a comprehensive survey. The objective was to identify the difficulties, pain points, and barriers faced by infrastructure managers, operators, and system suppliers in relation to data sharing. A set of targeted survey questions was developed to address these concerns. The responses received were then consolidated and analysed, providing a clear understanding of the key issues impacting data sharing in the sector.

5.1.1 Infrastructure Managers Results

The project aims to enhance the capabilities of Infrastructure Managers (IMs) in various critical aspects of planning, traffic control and reporting based on data exchange. This part of document outlines the key roles and responsibilities of IMs, the current state of data exchange, the challenges faced, and proposals for improvement. By leveraging new technologies and standardising data exchange protocols, the project seeks to optimise railway operations and improve service quality across the network.

5.1.1.1 Involvement of Infrastructure Managers in the FP5-TRANS4M-R Project

Infrastructure Managers play a crucial role in planning and traffic control. This requires the exchange of data between IMs and RUs as well as between rolling stock and

infrastructure, which involves alerts for damage and the capturing of logistic data. Effective traffic management is necessary to minimise delays and optimise the use of infrastructure.

IMs RFI, TRV and ADIF systems involvement

- Planning
- Traffic control
- Data exchange between rolling stock and wayside incl. damage alerts
- Logistic data capturing
- Railway checkpoints/Intelligent video gates
- Reporting

5.1.1.2 State of the art

Data exchange is vital for the efficient functioning of the railway sector. E.g., RFI, as an infrastructure manager, exchanges a significant amount of data with other sector actors. This data may be required for various purposes, including planning, operational needs, and reporting. Information on section availability, train status, composition, and other details are crucial for effective planning. Reporting unplanned delays and other operational data helps maintain smooth operations, while data is also important for creating reports and analyses that can be used to improve services.

At the national level, data is exchanged and consulted via proprietary information systems (in case of RFI web platforms) that includes all information useful for traffic management and planning. System interconnection involves Standardized European Checkpoints, TMS, RFID, and sensor stations, which collect information from the infrastructure and send it to the cloud for analysis. Connecting different systems is key to ensuring smooth data exchange and effective traffic management.

5.1.1.3 Issues

There are various issues related to data exchange that need to be addressed are following.

- The main criticalities resulting from the lack of a standardisation protocol can be identified as the following: difficulties in train tracking, ETA forecasting, real time rescheduling, transmission of information to the end customer. The information barrier is the non-compulsory provision of information and the lack of standardised structuring of interfaces. There should be definition of a standard valid for all transport modes involved, from origin to destination, would help data sharing and improve service quality. There are difficulties at the interface between main-line and terminal, for example, resulting in increased time for train pick-up operations. Data

exchange - Impact on network performance and use operations - problems related to the lack of data exchange could result in loss of local capacity and non-optimisation in the planning and execution of activities. Additionally, train numbers may change when crossing borders, causing identification problems and subsequently delays.

- Determining vehicle damage levels and presenting this information in a unified manner/format.
- Cybersecurity compliance with the National Security Scheme and internal processes. For example, ADIF, as a public sector company, is obliged to comply with the National Security Scheme, established in article 42 of Law 11/2007, of June 22, on electronic access by citizens to Public Services and regulated by Royal Decree 3/2010, of January 8, with an update in Royal Decree 311/2022, of May 3. Also, there is a cyber-security internal process base on internal regulation stating that the degree of sensitivity of the information owned by ADIF is defined in the internal document "ADIF-PG-108-007-N08-01 Classification of Information and Labelling of Documents." and is classified based on the impact that its loss or improper use would have on the business. Due to this internal regulation, it is necessary to establish and sign agreements with the company with which information will be shared.

5.1.1.4 Requirements

To improve data exchange, several steps should be taken. Infrastructure managers should be responsible for collecting and making traffic data available on their infrastructure. Cooperation within the entire European rail community among others RNE, UIC, and in case of RFID, GS1, can help with data standardisation. New technologies can help collect and make more data available to all actors, potentially leading to the creation of a single integrated platform for data exchange and consultation. The fields of improvement are following:

- Data Governance – the IM should be responsible for collecting and making available (subject to authorisation by the railway undertakings, where applicable) traffic data on its infrastructure. In the context of the project, e.g. RFI plans to exchange data necessary for the development of Use Case 33.3 – the New Transport Configurator. Specifically, this will use residual network capacity identifiable thanks to the catalogue of available paths.
- Data Standardisation – cooperation within the European rail community, e.g. GS1 has a recent released solution for exchange of RFID data I Europe, a server in place. The barrier of a lack of standardization could be overcome with the mandatory agreements, since confidentiality and treatment are established to guarantee data security.
- New Technologies **–** collecting and making more data available, creating an integrated

platform. Except of the standardization also the new technologies can help to collect and make more data available to all actors involved, thus they could help in the construction of a single, integrated platform where data from the different railway operators can be exchanged and consulted.

Lack of data exchange can negatively impact network performance. It can lead to capacity loss and non-optimisation of planning and execution activities. Standardisation and new technologies can help overcome these barriers and improve service quality. In conclusion, the effective involvement of Infrastructure Managers in the FP5-TRANS4M-R project is essential for the successful transformation of the railway sector. Addressing the challenges in data exchange through standardisation and the adoption of new technologies can significantly enhance network performance and service quality. By fostering collaboration within the European rail community and ensuring robust data governance, the project can pave the way for a more integrated and efficient railway system. Continued efforts in these areas will be crucial for achieving the project's goals and delivering tangible benefits to all stakeholders involved.

5.1.2 Railway Undertaking

This part outlines the key roles and responsibilities of RUs, the current state of data exchange, the challenges faced, and proposals for improvement. By leveraging new technologies and standardising data exchange protocols, the project seeks to optimise railway operations and improve service quality across the network.

5.1.2.1 Involvement of Railway Undertakings

Railway Undertakings (RUs) such as ČDC and ÖBB play a crucial role in planning and traffic control, including train running prediction and logistic data capturing. Effective traffic management is necessary to minimise delays and optimise the use of assets. Poor data quality is a significant problem for daily operations, as high-quality data is essential to fulfil daily tasks and satisfy customers. When data quality is poor or unavailable at the required time, it incurs personnel expenses and additional costs.

RUs (ČDC and ÖBB) involvement within the project FP5-TRANS4M-R

- Planning, assets management
- Traffic control including train running prediction
- Logistic data capturing
- Reporting

5.1.2.2 ČDC involved systems:

- ELITE – Short-term planning system for dynamic transport planning. Creates a link between the transport request (order) and the trains. The plan for the dispatch system is created here. The system also provides information about possible train

connections for other systems

- GPPS – System for planning of technological and operational processes in stations or processes in marshalling yard.
- DISCOR – Dispatching system for transport management - train dispatching and running, resource capacity management (path, loco, driver). The system also allows real-time train management.
- Transport Planner (Assets Warehouse) – A newly developed system. The system allows for easy planning of resource capacities (locos, drivers, other personnel) based on real customer/transport order requirements (dynamic allocation). It can also be used as a central resource warehouse for transport planning or other operations for sharing purposes among RUs.

5.1.2.3 State of the Art

A major challenge in a daily operation is the poor quality of data. High-quality, timely data is essential for fulfilling daily tasks and meeting customer expectations. When data quality is subpar or unavailable at the required time, it results in increased personnel expenses and costs.

5.1.2.4 Issues

Several issues related to data exchange need to be addressed. One of the primary concerns is the development of multiple interfaces. There is a strong desire for a single, standardised interface set of interfaces based on TSI. In case of European checkpoints there is necessary to have one standardized interface. Additionally, the versioning of interfaces poses a problem, especially when different stakeholders use different versions of the same interface definition. A clear and comprehensible description of the interface is crucial, as is the reliable and swift implementation of interfaces and quick response times to our requests. The issue of different versions for different partners using the same interface remains a significant challenge.

- Cybersecurity is paramount in daily operations. Data transfer must be secure, ensuring that only the intended recipient receives the data and that it does not fall into the hands of unauthorised third parties.
- Systems and Real-Time Data Sharing – in case of a new interface implementation, we always evaluate whether to integrate functionality into the system or create a new system. This approach helps us scrutinise systems and replace them with new system when necessary.
- Cooperation with Terminal Operators and System Suppliers – currently the communication is performed with some terminal operators via standardised interfaces (such as TAF/TSI, Hermes, etc.) whenever possible. It is necessary to use the

latest version of each interface. In case the standardised interfaces are unavailable, it is necessary to develop the proprietary interface.

5.1.2.5 Requirements

To improve data exchange would help defining obligatory European standards for all interfaces. Stakeholders, mainly infrastructure managers should be required to use standard interfaces whenever available and always use the latest versions of these interfaces.

The involvement of railway undertakings in the FP5-TRANS4M-R project highlights the importance of high-quality data exchange, cybersecurity, and the need for standardised interfaces. Addressing these issues will enhance operational efficiency and ensure better service delivery to customers.

5.1.3 Intermodal Operator Results

This part outlines the various in-house systems and public applications utilised by Kombiverkehr, the current state of data exchange, and proposals for future improvements.

Intermodal Operator – Kombiverkehr involvement

- In-house systems CAT (intermodal operator system),
- Public timetable application
- Public customer portal
- Internal train monitoring system
- Data hub DXI (KV4.0)
- RNE TIS train running information
- DB Leidis (The operations information system LeiDis-FI collects operational information on train runs (current positions, timetable data), train formation or routes).

5.1.3.1 State of the art

The state of data exchange within the FP5-TRANS4M-R project is evolving. All systems are, or will be, connected via electronic interfaces. This integration facilitates the creation of new features for all involved intermodal actors. Historically, obtaining data from other sources and partners was challenging. However, with the implementation of KV4.0 and TIS, there has been a notable shift in participant behaviour. There is now inside the intermodal transport chain a willingness to exchange data upon request, making it an opportune time to establish connections for data exchange. Currently, the primary obstacles to data exchange are technical barriers and a lack of standardisation.

5.1.3.2 Requirements

At the moment it is just a matter of time and recourses to implement new interfaces. Data is available and the willingness to exchange has arrived in the minds of the involved actors.

In most cases there are mostly technical barriers to exchange date, but the biggest problem of missing standardization will get smaller with the new railway and intermodal standards TAF TSI and EDIGES. So, it is important to fund projects that are planning to implement new features to use the new data treasure and adapt their systems to be part of the new railway standard.

Kombiverkehr's involvement in the FP5-TRANS4M-R project highlights the importance of data exchange and standardisation in the intermodal transport sector. By overcoming technical barriers and embracing new standards, Kombiverkehr and its partners can enhance their systems and contribute to a more efficient and integrated transport network.

5.1.4 System Supplier Results

This part outlines Hacon's contributions, focusing on the development of functionalities for existing systems, the creation of a prediction system for freight, and the implementation of a multimodal transport planner.

Hacon involvement

Development of functionalities for existing Hacon systems within WP26 and WP27 regarding the exchange of information and messages between yard systems and mainline systems. Development of a prediction system for freight within WP28 and finally a multimodal transport planner within WP31.

5.1.4.1 State of the art

The current state of data exchange involves connecting various systems via electronic interfaces. These systems include

- Terminal Operating Systems,
- Yard Operating Systems,
- Main-Line Transport Management Systems,
- Commercial Timetable Information,
- RNE Train Information System,
- Data Warehouses

5.1.4.2 Issues

There is a notable reluctance among actors to share data, particularly performance-

related data such as punctuality information. Additionally, there is a lack of information regarding existing systems that can provide the required data. Legal and administrative constraints also pose significant challenges, requiring considerable time and effort to resolve. Furthermore, different standards regarding data exchange, including data quality, data format, ontologies, data content, and data structure, complicate the process.

5.1.4.3 Requirements

To address these issues, stronger enforcement of mandatory data standards, especially TAF TSI, is proposed. This includes ensuring the general fulfilment of the provision of certain message types and the processes by which these messages and the information within them are generated and sent. A central overview of data standards, platforms, ontologies, and public data sources is also recommended, detailing the types of data available from each source. Additionally, joint sector agreements emphasising the importance of data sharing and transparency are suggested, including a general, non-binding commitment towards a more open exchange of data.

Hacon's involvement in the FP5-TRANS4M-R project is pivotal in advancing the integration and efficiency of transport systems. By addressing the current challenges in data exchange and proposing robust solutions, Hacon aims to foster a more collaborative and transparent environment within the transport sector. This will ultimately lead to improved performance and reliability across various transport systems.

5.1.5 Terminal Operator and System Supplier

This part outlines the current state and issues related to data exchange and particularly focusing on TSI data.

GTS systems involvement

- Terminal Management System,
- Booking System,
- Track & Trace System

The integration of terminal management systems, booking systems, and track and trace systems is crucial for the efficient operation of terminal operators. These systems are, or will be, connected via electronic interfaces to traffic management systems, European checkpoints, and other relevant systems.

5.1.5.1 State of the art and issues

The current state of data exchange, especially concerning TSI data, presents several challenges. It is almost mandatory to use RNE, which poses a commercial hurdle, and the distribution of TSI data is very limited. For instance, in some countries, traffic management systems struggle to send out TSI data, and each railway undertaking can only direct the

data stream to one company. This situation becomes particularly problematic for larger RUs, making it nearly impossible to receive this data without incurring costs from multiple stakeholders.

5.1.5.2 Requirements

To address these issues, it is proposed that all TSI data should be made freely available to all stakeholders. This approach would eliminate the commercial barriers and ensure that all relevant parties have access to the necessary data, thereby improving the overall efficiency and effectiveness of terminal operations and system integrations.

In conclusion, the current limitations in TSI data exchange hinder the optimal functioning of terminal management systems and related interfaces. By opening up TSI data for free access to all stakeholders, we can overcome these challenges and foster a more collaborative and efficient environment for terminal operators and system suppliers. This proposal aims to enhance data accessibility, reduce costs, and improve the overall performance of the integrated systems.

5.2 Process Analysis

To ensure business value for involved stakeholders in transport systems, an efficient data sharing ecosystem is required. Importantly, different involved actors use the data being shared for different purposes to gain value in their operations.

However, to ensure high-performing business practices, business operations need to be aligned with the use of supporting technologies. As it is, with few exceptions, the countries that own the railway infrastructure and the authorities appointed by them that maintain, manage, and regulate the use, rules and policies for the sharing of the data must be established and met by the different actors involved, as e.g. a train operator applying for and receiving license etc. to be allowed to operate the railway.

As this project focuses on the shipping of goods between countries in Europe, the complexity of an aligned approach to data sharing and applied business operations is complex as the goods that are transported by rail have very likely used/will use adjacent modes of transport of the goods during the journey from origin to the final destination. The transport buyer, i.e. the shipper as well as the end customer is examples of important stakeholders in the transport chain, of which actors engaged in rail transport need to position themselves in relation to, as those have an interest in certain information details. Freight by rail is not necessarily the end but is one of many important legs, not the least a sustainable option, in a supply chain and where the exchange of information is vital – both between actors within the railway sectors and with actors associated to the railway sector.

Given the above, each actor and stakeholder in the entire supply chain needs to "speak

the same language" (semantics), to ensure that the information is of the right quality and that it is available to each actor and stakeholder at the right time. Of course, cyber security also needs to be considered. Admitting that transports are conducted within a self-organizing ecosystem, requires a collaborative approach to cyber security.

Over the years, several different projects have analysed and evaluated how different business practices, in this case everything from land, sea to air transport, individually and together should be able to collaborate in an effective way to address the concerns of the client of the transport ecosystem being the physical cargo flow and its digital representation for enhanced visibility to follow the goods. A cargo carrier being the container can be transported by ship, truck and rail, which is identified by a unique ID which is a necessity for the digital representation of the container. This container carries a large amount of goods that also have unique digital ID's emerging from the customer order. A project to mention in this context is the EU project FEDeRATED, https://www.federatedplatforms.eu/, which proposes a semantic model to digitally represent essential aspects of the supply chain (see http://185.87.184.112/~federated/images/Library/Activity2/TechnicalSpecs/1_Technical_Specifications_Semantics.pdf).

In this project, FP5 TRANS4M-R, the Conceptual Data Model, CDM, is used for this purpose and is developed in another project, FP1 MOTIONAL, with the purpose to handle the semantic part regarding rail transports and its relation to other modes of transport. This too is described in more detail in another chapter of this document.

Another important part for railroad transport is TAF/TAP (Telematic Applications for Freight/Telematic Applications for Passenger services) and TSI (Technical Specifications for Interoperability), which are technical specifications that primarily regulate relationships between infrastructure owners and railway undertakings both for freight and passenger traffic, e.g. DB, Green Cargo and others. RNE (RailNetEurope) has the mandate to coordinate the implementation of those specifications. More details to be found in https://rne.eu/it/taf-tap-tsi/, and also, at http://taf-jsg.info/.

5.3 Data Quality and Availability

Data quality and availability are critical concerns for railway freight services in Europe. Addressing these issues is essential for ensuring efficient operations, compliance with regulations, and maintaining customer satisfaction. Some of the main concerns are the following:

5.3.1 Data Accuracy and Consistency

- **Errors in Data Entry:** Inaccurate data can result from manual entry mistakes and also from inabilities of technical devices, leading to discrepancies in schedules, cargo details, or customer information. For example, lacking image analysis capabilities and obscured or damaged vehicle markings may lead to a checkpoint (developed in WP29) reporting

incorrect vehicle numbers.

- **Inconsistent Data Formats**: Different operators or systems may use diverging data formats, making integration and consistency difficult.

5.3.2 Real-Time Data Availability

- **Timeliness of Information**: Delayed updates regarding freight status, location, and condition can hinder operational decisions and customer communication. For example, the prediction of arrival times in WP28 requires that changes to the operational timetable are quickly made available.

- **Integration Challenges**: Lack of real-time data sharing across the supply chain can lead to miscommunication and inefficiencies.

5.3.3 Data Silos

- **Fragmented Data Sources**: Different entities within the freight ecosystem (e.g., rail operators, logistics providers, customs authorities) often store data in silos, making it inaccessible to other stakeholders.

- **Poor Collaboration**: Fragmentation can inhibit collaboration and data sharing, leading to an incomplete picture of freight movements.

5.3.4 Compliance with Regulations

- **Regulatory Requirements**: Maintaining data quality to comply with various European regulations and standards related to safety, security, and environmental impacts is essential.

- **Tracking and Reporting**: Accurate data is necessary for compliance reporting, which may include safety audits, customs declarations, and environmental assessments.

5.3.5 Data Security and Integrity

- **Cybersecurity Risks**: With increasing digitalisation, the risk of data breaches or cyberattacks poses a threat to data integrity and availability.

- **Disaster Recovery**: Ensuring data is backed up and recoverable in case of system failures or cyber incidents is crucial for operational continuity.

5.3.6 Data Standardisation

- **Lack of Industry Standards**: The absence of universally accepted data standards can lead to compatibility issues and hinder data sharing across different systems.

- **Diverse Stakeholders**: Various stakeholders in the freight supply chain may have differing standards for data collection and reporting.

5.3.7 Maintenance of Data Quality

- **Data Lifecycle Management**: Proper management of data throughout its lifecycle is crucial, including consistent monitoring, cleaning, and updating of data records.

- **Staff Training**: Ensuring that staff are adequately trained in data handling practices is vital for maintaining high data quality.

5.3.8 Impact of External Factors

- **Supply Chain Disruptions**: External factors such as political changes, economic fluctuations, and disruptions (e.g., pandemics) can affect data availability and quality, impacting planning and decision-making.

- **Changing Customer Preferences**: Understanding and adapting to evolving customer needs requires timely and accurate data on market trends and operational capabilities.

5.3.9 Technology Integration

- **Legacy Systems**: Existing legacy systems may not support modern data management practices, making it challenging to collect and analyse data effectively.

- **Interoperability Issues**: Different technologies used across the rail networks can hinder seamless data exchange and accessibility.

Addressing these concerns is imperative for improving the efficiency and reliability of railway freight services in Europe, enhancing customer satisfaction, and promoting sustainable practices within the logistics sector. Implementing robust data governance frameworks, adopting new technologies, and fostering collaboration among stakeholders can help mitigate these challenges. Strategies to improve data quality and data availability are further discussed in chapter 7.1 and chapter 7.2 , respectively.

6 General principles for Data Sharing and Data Structure

The International Data Space Association (IDSA) standard enables data sharing through data spaces characterized by uniform rules, certified data providers and recipients and trust among partners. Data spaces provide the basis for fruitful cooperation, lowered barriers to entry and limitless innovation in the data economy of the future.

In particular, the Rail Data Space, based on IDSA standard, is foreseen to be used in the demonstration phase for the data exchange among partners within the seamless use cases. The Rail Data Space is currently being deployed by WP31 in the FP1-MOTIONAL project.

All IDSA-based data spaces include several core capabilities (independent of the actual data to be exchanged) that promote the trust among the partners. In particular, the following subsections identify and technically describe how data spaces provide the following transversal functionalities:

- Identity and access management (see subsection 6.1), in order to identify and authenticate the participants and authorize their activity.
- Data auditing and monitoring (see subsection 6.2), in order to trace, log and monitor activities (e.g. contract negotiation, data transfer) that take place between the participants of a data space.
- Data sharing policies (see subsection 6.3), in order to define and enforce the restrictions on usage of such shared data.

Data spaces provide these core capabilities but are independent of the data to be exchanged. In this sense, additional mechanism is necessary for aiming data standardization and interoperability. This aspect is detailed in subsection 6.4.

6.1 Identity and access management

6.1.1 Overview

Identity and access management is one of the enablers that can make data sharing be trustful. This sub-section describes how these capabilities are provided within the data spaces. As already described in D25.1, a data space, a concept introduced by the International Data Spaces Association-Reference Architecture Model (IDS-RAM), can be defined in its simplest form, as an open platform where its participants can exchange data by complying to different access and usage policies. To achieve this, a framework is required to uniquely identify the participants in a data space and consequently establish trust among them based on the registered identities. A reliable trust framework with identity and access management does not only provide authentication and authorization,

but also enables secure data exchange with non-repudiation and auditability. Next paragraphs describe technically the approach proposed by IDS-RAM4.0 to provide identity and access management in the process of data-exchange.

6.1.2 Provision of identities

To cope with multiple existing use cases, IDS-RAM 4.0 defines identities for both participants and their devices involved in data spaces. The former can represent organizations or individual humans whereas the latter represents the technical components such as a Connector (a software through which participants can connect to a data space and exchange data) or a Metadata Broker (a catalogue of published data offers available to consumers in a data space). In both these cases, the subject should be certified prior to the registration with an identity.

A set of services has been defined in IDS-RAM 4.0 to realize the requirements associated with the identity and access management. As shown by step 1 in Figure 1, a participant initially requests an Evaluation Facility to register a technical component in a data space. This facility then carries out an assessment based on the participant's technical and organizational aspects. As a result of this evaluation process, it delivers an Operational Environment Certification (e.g. for organizational processes) and a Component Certification (e.g. security of hardware and software platforms) with the corresponding trust and assurance levels as defined by IDSA. This entire evaluation process will be governed by a Certification Body that regulates the standard evaluation procedures and supervises the activities of Evaluation Facilities. A Certificate Authority (CA) will be notified once the evaluation process is concluded (step 2). If these certifications are valid, a participant will then be granted an X.509 certificate by the CA (step 4), which primarily serves as the identity certificate of the underlying component.

Figure 1 Process of digital identity provision [5]

A distinct service called DAPS (Dynamic Attribute Provisioning Service) has been defined in IDS-RAM to incorporate dynamic trust, in addition to the static trust provided by CAs and Certification Bodies. Once a participant registers with a valid X.509 certificate (step 5), DAPS delivers a Dynamic Attribute Token (DAT) to the requested participant's component. This token contains signed claims with up-to-date information such as Software Manifests of the component, Company Description and other dynamic attributes (e.g. location, supported transport certificates). Such dynamic information is provided by a separate service known as DTM (Dynamic Trust Monitoring) that continuously monitors the participants and shares its information with DAPS such as levels of trustworthiness, security vulnerabilities and attempted attacks (step 6).

Due to its transient behaviour and decoupling from the identity certificate, DAT reduces the need for certificate revocation and enables more flexibility to include dynamic attributes whenever applicable. The participant can then share this token with any other IDS component (e.g. Connector) to authenticate and establish trust prior to their interactions. Consequently, an IDS component should also possess the capability to verify any given DAT by associating it to the identity certificate of the sender. Figure 2 shows the interactions between the participants and the Identity Provider during an authentication process. In particular, an IDS Connector acquires a certificate from CA (step 1), requests a DAT from DAPS (step 2) and shares this DAT with the other participant's Connector (step 3).

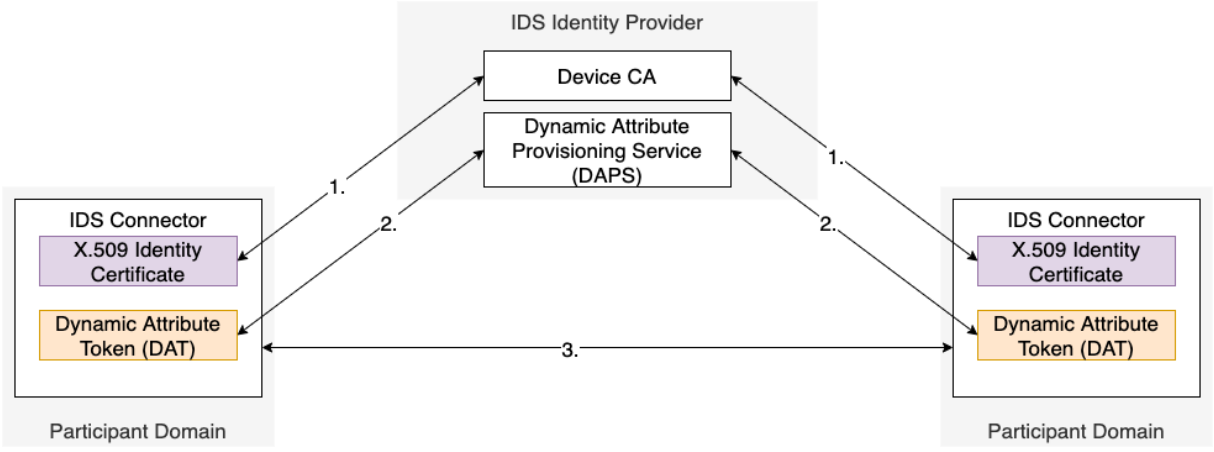

Figure 2 Interaction between Connectors and Identity Providers [5]

In addition, a complementary service known as ParIS (Participant Information Service) provides further information about the participants (e.g. business attributes such as registered address), which can be referenced during an initial connection setup between two components. This also provides a higher level of trust as the information stored in ParIS is verified by a Support Organization, usually through a manual human intervention.

In summary, IDS services such as Certification Bodies, Evaluation Facilities, CAs, DAPS, DTM and ParIS together form an identity management system that can be used by IDS components to authenticate and exchange data while upholding trust.

6.1.3 Access management

IDS follow a resource-oriented and attribute-based approach for access control, where policies can be defined on targeted resources or endpoints to include the constraints that should be satisfied by the subjects (e.g. identities, attributes, security profiles, environmental contexts of the participant). As an example, a data owner may enforce a specific dataset to be accessed only if the requesting Connector is located within the EU and has a security profile of TPM >= 1.2 (TPM stands for Trusted Platform Module). The requesting participant may provide such information via a token, a certificate or a Verifiable Credential. In addition to the actual exchangeable data, access policies can also be defined on other resources owned by the participant, such as Self-Descriptions hosted on a Metadata-Broker.

The enforcement of such access policies can be implemented within the Connector (or as an external service) using technologies such XACML (eXtensible Access Control Markup Language) or JAAS (Java Authentication and Authorization Service), depending on the internal requirements. However, IDS do not impose any restrictions on the policy language or its specific implementation. These aspects related to the policy engine and its subcomponents, are described in the subsection 6.3 regarding data sharing policies.

6.2 Data auditing and monitoring

The process of tracing, logging and monitoring activities (e.g. contract negotiation, data transfer) that take place between the participants of a data space is defined as its observability. IDS-RAM 4.0 delegates these observability functions to a separate trust service known as the Clearing House. A Clearing House can be based on the architecture of an IDS Connector with several components such as Clearing and Settlement, Logging, Usage Control Claim Validation and Billing services.

A participant involved in a data exchange process may log its activities and transactions in a Clearing House such that they can be referred in future to resolve any possible conflicts (e.g. whether a data asset has been received by the Consumer) and to provide clearing and settlement services for financial transactions based on the actual usage of data. As a trust service that provides logging functionality, the Clearing House enhances transparency of data space activities by enabling verification methods for their compliance with usage policies, business contracts and legal regulations. These message logs can also be digitally signed by the relevant participants to provide non-repudiation. Furthermore, they maintain data provenance and traceability across the same context

(e.g. a data asset) such that their messages can be used to find out when, how and by whom the data were modified.

Figure 3 shows an example use case of the Clearing House. During a contract negotiation process, a provider may validate the consumer's request, sign the contract agreement and store it locally. This agreement can then be sent to a Clearing House which will validate the participants' signatures, add its own signature, persist the agreement and return it to the relevant participants. In this case, the Clearing House extends the aforementioned functions to provide separate trust and regulation for the negotiation process.

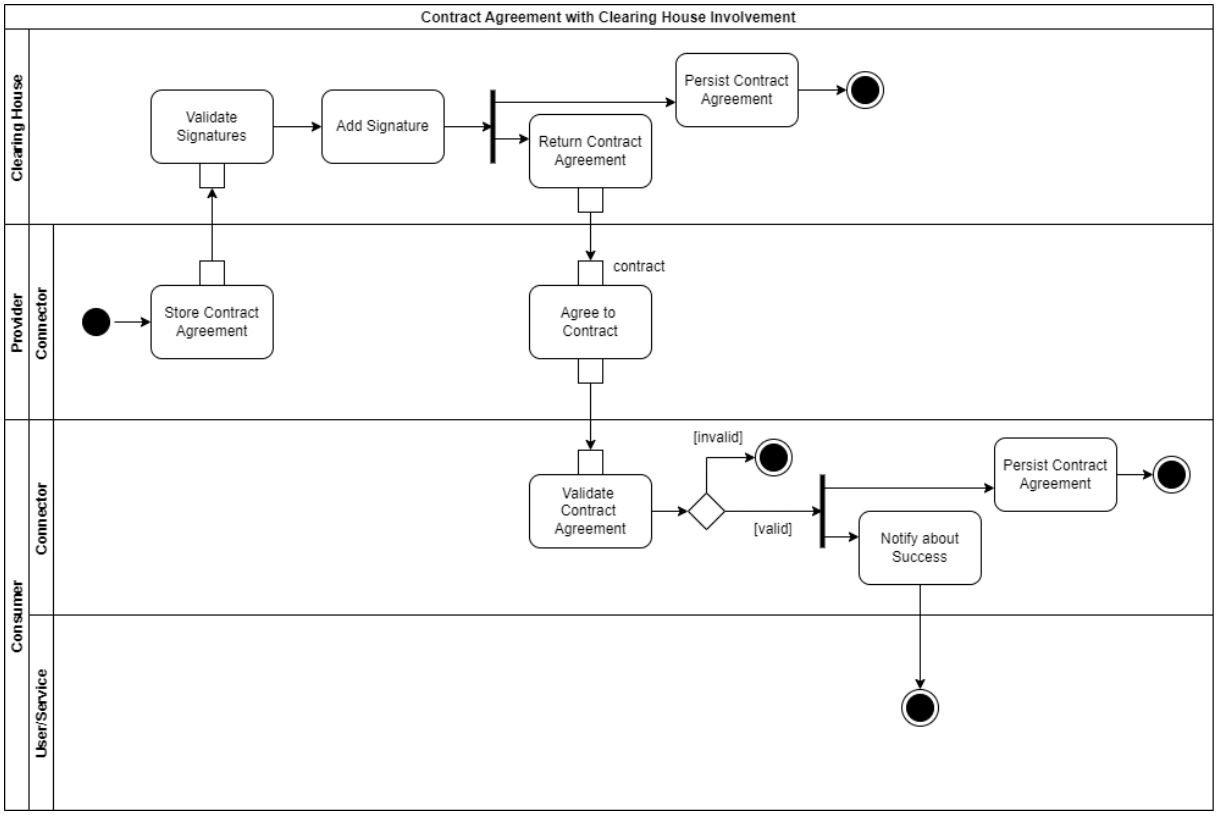

Figure 3 Involvement of Clearing House in contract negotiation process [5]

Fundamentally, the Clearing House can exist as an external centralized service as shown in Figure 4. Activities related to a data exchange (e.g. policy-based decisions) will be emitted to a Data Flow Tracking component deployed in an IDS Connector via event-driven notifications. These event logs will then be forwarded to the centralized Clearing House for storage, which can also be monitored and queried by Connectors through its Privacy Dashboard.

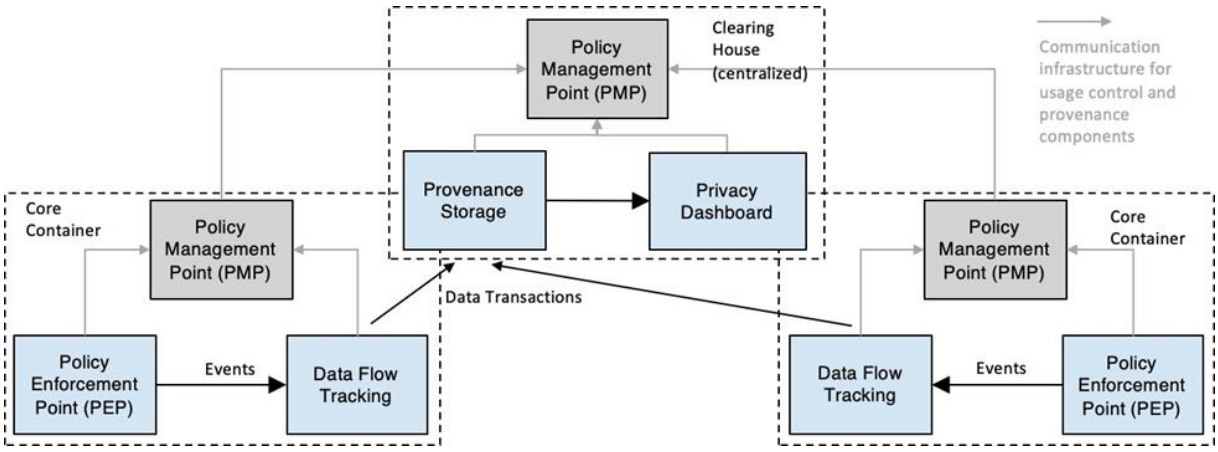

Figure 4 Architecture of a centralized Clearing House [5]

IDS-G (International Data Spaces Global), a set of specifications published by IDSA, provides an API overview for the basic functionalities of a Clearing House. It defines an API to create a process for a specific data exchange with a request comprising a JSON array of additional owners of the process. Similarly, a payload with the model version, issued date, issuer and recipient connectors, sender agent and a security token can be used to log a message for a specific process in the Clearing House. If the process does not exist, Clearing House creates a process before logging its messages. The response contains a signed receipt as a proof that data has been logged in the Clearing House. An owner of the process can also query the messages with a similar payload (additionally with query language and scope), to which the Clearing House responds with a JSON array of logged messages.

However, this centralized model leads to other concerns such as the additional vulnerability of sharing critical data with an external party and the possibility that this central observer may exploit such information for its own financial gain. Alternatively, a federated or a decentralized model can be used for the Clearing House to reduce these risks associated with a centralized observer.

In a decentralized architecture, each participant maintains the logging information separately, including the contract agreements, but linking them together with a correlation ID. A third-party participant such as an auditor can then request for this logging information with a valid authentication (e.g. Verifiable Credential). This approach preserves the sovereignty of participants related to logging information and restrains the messages' observability only to the external trusted parties. Additionally, usage policies can be defined on the logging information to further regulate the actions of third-party services. In fact, these observer actions can also be logged, tracked and monitored, thus enabling a trust relationship in which auditors can be audited by participants.

6.3 Data Sharing Policies

6.3.1.1 Overview

Data sharing can take different forms such as a one-time file transfer, access to an API, subscription to a data stream and even scenarios where data remains at the source and processing algorithms are transferred to use the data. In all these cases, data spaces enable using data across multiple domains and organizations and therefore, it is typical to enforce the restrictions on usage of such shared data. To this end, IDS allows data owners to define policies with a set of rules (e.g. permissions, obligations) that pertain to the processing of the shared data. It further necessitates the participants' actions to be actively monitored and intercepted by control points whenever applicable. Example use cases for such usage restrictions can be listed as below.

- Classified data must not be forwarded to public endpoints.
- Critical data must not be modified by untrusted nodes.
- Shared data must be deleted after a certain period.
- Datasets from competitive participants must not be aggregated.

6.3.1.2 Definition of policies

The usage policies can be defined in self-descriptions of a Connector, which will be referred to and agreed to in advance by the interested participants, specifically during a contract negotiation process. The corresponding rules contained in a policy can be specified at organizational, legal and technical levels, where they will be used interchangeably or complementarily (e.g. using removable storage devices can be prevented through an organizational policy for the employees or technically with the Operating Systems). It is also possible to define its conditions based on the phase, such as before (e.g. integrity check of the component), during (i.e. valid only during business hours) and after (e.g. data should be deleted after usage) the decision-making process.

IDS uses a profile of ODRL (Open Digital Rights Language) to express usage policies in a machine-readable, technology-agnostic and interoperable format. It supports transforming abstract declarative predicates into an operable programming logic that can be implemented and executed in data space components. These policies are further categorised as IDS Policy Classes that define relevant operands, operators, values and data types for specifying such usage restrictions (e.g. time-restricted policy class only accepts values in *xsd:dateTimeStamp* and restricts data usage to the specified time interval).

A key aspect of defining policies is that they should be understood by the participants accurately without any ambiguity. This can be achieved in data spaces via Vocabulary Hubs that contain the semantic models required for the policies. These models can be referenced in a self-description (e.g. Connector, data space) or the specific contract. In

case of mandatory vocabularies, referencing their Vocabulary Hubs can also be a part of the policy definition.

6.3.1.3 Enforcement of policies

Enforcement of the usage policies in IDS may involve multiple services based on the context. For example, a Provider may share data only with the Trusted Connectors certified by a Certification Body that guarantees its software stack and the reliable processing of data.

As the central component in data spaces, an IDS Connector focuses on security and delivers a trusted platform in order to support the usage control. To restrict the data usage, a Connector may use an interceptor pattern with several components. In particular, PEP's (Policy Enforcement Points) can be deployed throughout the route from one participant to another, to enforce the policy-based decisions on data. Such decisions are initially evaluated by a separate component known as PDP (Policy Decision Point). In cases where contextual information (e.g. time, location) is required to evaluate a usage policy, PIPs (Policy Information Points) can be used both internally within a Connector and externally connected to other services (e.g. ParIS). IDS Usage Control language provides ways to address and use this context information. Moreover, a Policy Execution Point (PXP) can be used to implement instructions that must be executed before or after a decision (e.g. when an obligation must be fulfilled to comply with a policy). As similar to PIP, PXP can also be deployed internally (e.g. deleting shared data after processing) or externally (e.g. sending logs to external destination for provenance) based on the use case requirements.

Depending on the transfer type, the actual data exchange/sharing process may take place after a considerable period of time from its corresponding contract negotiation process. Hence, the agreed policies can be revisited again during the actual transfer process to further ensure the validity of the policies.

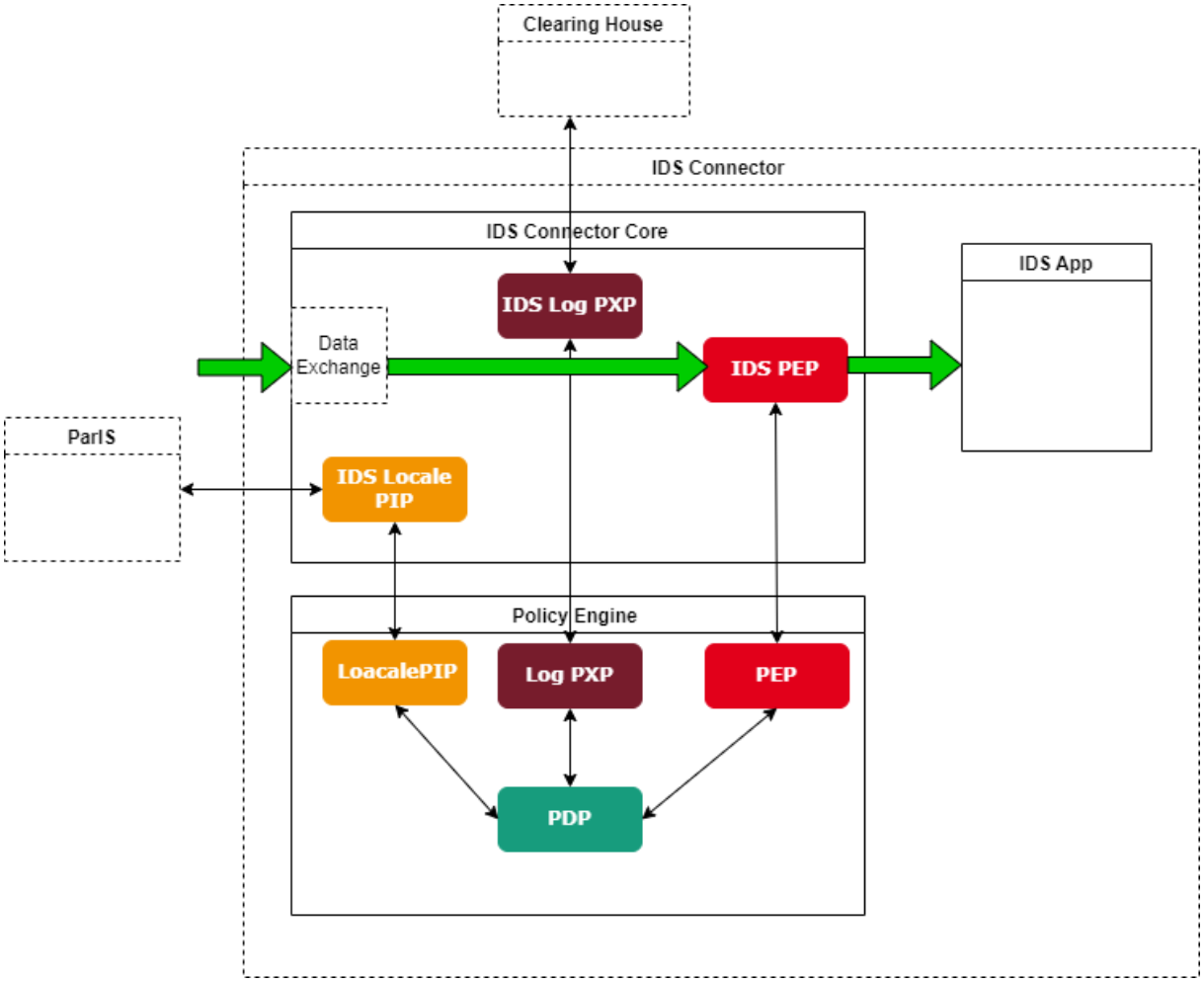

Figure 5 Usage control components of a Connector [5]

In addition to proactive (or preventive) policy enforcement, IDS-RAM 4.0 further specifies on provenance of data that can be passively used for the compliance of usage policies. This functionality can be incorporated as an extension of the PEP component, to transmit events (semantically defined with respect to a data flow model) to an external service such as the Clearing House. Such event logs are stored in data provenance graphs (as a tree data structure for a specific data content) in the Clearing House, which can then be aggregated for clearing, conflict resolution, auditing and billing purposes based on their usage policies. In distributed usage control scenarios such as supply chains, this provenance data can be fed back into the components such as PEP, to support further decision making related to the data content.

6.4 Data standardisation and interoperability

The objective of this chapter is, by gathering the information developed in the operational analysis detailed in Deliverable 29.1 "Technical definition of the standard IVG fir checkpoints and related demonstrators", to establish specific requirements to complete the information of the Conceptual Data Model (CDM), elaborated in Flagship Area 1, Work

Package 30. In Deliverable 29.1, an Operational Analysis was carried out, focusing on the critical operational points of the European railway network.

For each operational point, or use case, a series of operational activities have been identified where the European Railway Checkpoint (ERC) can intervene to digitalize the processes. These activities are detailed below:

Table 1 ERC Operational Use Cases and activities

Location	***Process***	***Description of Process***
UC01 - Cross Border	Technical Wagon Inspections	Technical checks ensure the reliability of the wagons according to GCU, TSI, and TAF TSI regulations. Data on flat wheels, bearing temperature, pantograph damage, wagon dimensions, and locomotive types is gathered and shared.
UC01 - Cross Border	Train Composition Verification	Verification of the train's composition, including train number, locomotive numbers, wagon numbers, and sequence. Ensures readability of labels and signs, as well as compliance with regulations.
UC01 - Cross Border	Brake Test	Brake tests are performed to ensure the train's brake system is functioning properly. A full brake test is required if more than 24 hours have passed or if locomotives are changed at the border.
UC01 - Cross Border	Incident Detection and Hazardous Material Check	Inspectors check for damage, rust, broken parts, and signs of hazardous materials. Ensures all dangerous goods are properly documented, labeled, and safe for transportation.
UC02 - Terminal	Automated Gate Operations	Automated gate operations identify intermodal units upon entry/exit at terminals. Data on wagon numbers, timestamps, train numbers, and positions is captured to manage arrivals and departures.
UC02 - Terminal	Weight and Dimension Checking	Weight and dimensions of intermodal units are verified using weighbridges and dimension scanners to ensure compliance with safety and regulatory requirements.
UC02 – Terminal	Data Logging and Integration with Terminal Systems	Logs and integrates data on intermodal unit movements (entry/exit) into terminal management systems for real-time tracking and inventory updates.
UC03 - Shunting Yard	Train Arrival and Data Acquisition	When a train arrives at a shunting yard, automated data acquisition captures wagon IDs, train numbers, wagon positions, and other critical information for sequencing and reorganization.

UC04 - Workshop	Integration with wear detection systems and wagon tracking	Wear detection systems assess the condition of critical wagon components, including wheels, brakes, and axles, to flag any maintenance needs or anomalies.

Based on the previous Use Cases, in Deliverable 29.1 "Technical definition of the standard IVG fir checkpoints and related demonstrators" an operational analysis was carried out, in which operational activities carried out at each operational point were identified, and associated with these, a series of information relevant to these activities.

Based on the information identified, this chapter will establish a set of formal requirements, which define the standards to be followed when taking this information into account in the Conceptual Data Model (CDM).

6.4.1 System Functionalities

Based on the operational analysis and identified operational activities, this chapter aims to perform an analysis of the system's functionalities, associating the previously identified information with these functionalities. The objective of this analysis is to establish which specific functionality of the European Railway Checkpoint (ERC) system is in charge of generating or consuming such information. ERC system functionalities are outlined in the following table:

Table 2 ERC System Functionalities associated with the identified information

Functionality	Functionality Description	System Function	System Function Description	Data
Rolling Stock Characterization	Describes and monitors the characteristics of railway rolling stock.	Wagon Identification	Identifies individual wagons and tracks their relevant details, including the number assigned to each wagon.	Wagon Numbers
		Container Identification	Identifies the physical characteristics of containers, including the external dimensions of each wagon.	Wagon Outline dimensions
		Composition Determination	Determines the overall composition of a train, including the number of wagons, their arrangement, and related metrics.	Train numbers
				Wagon numbers
				Total number of wagons
				Length of trains
				Position of wagons in the train

Dangerous Goods Characterization	Monitors and evaluates the characterization of dangerous goods being transported.	Placards Detection	Detects placards on wagons indicating the presence of dangerous or hazardous materials, ensuring proper identification.	Information about dangerous goods
		UN Number Identification	Identifies the United Nations (UN) numbers that classify dangerous substances, ensuring safe handling and transport.	Information about dangerous goods
Condition Monitoring	Tracks and evaluates the condition of the rolling stock to ensure operational safety.	Defects Detection	Detects various defects and irregularities in the condition of train components to ensure the safety and functionality of the trains.	Flat wheel damage
				Temperature of bearings
				Pantograph damage
				Status of axles
				Status of brakes
				Status of springs
				Status of the Kingpin or twistlock
				Status of container doors
				Status of the cargo

These functionalities relate to each operational activity as detailed in the following figure.

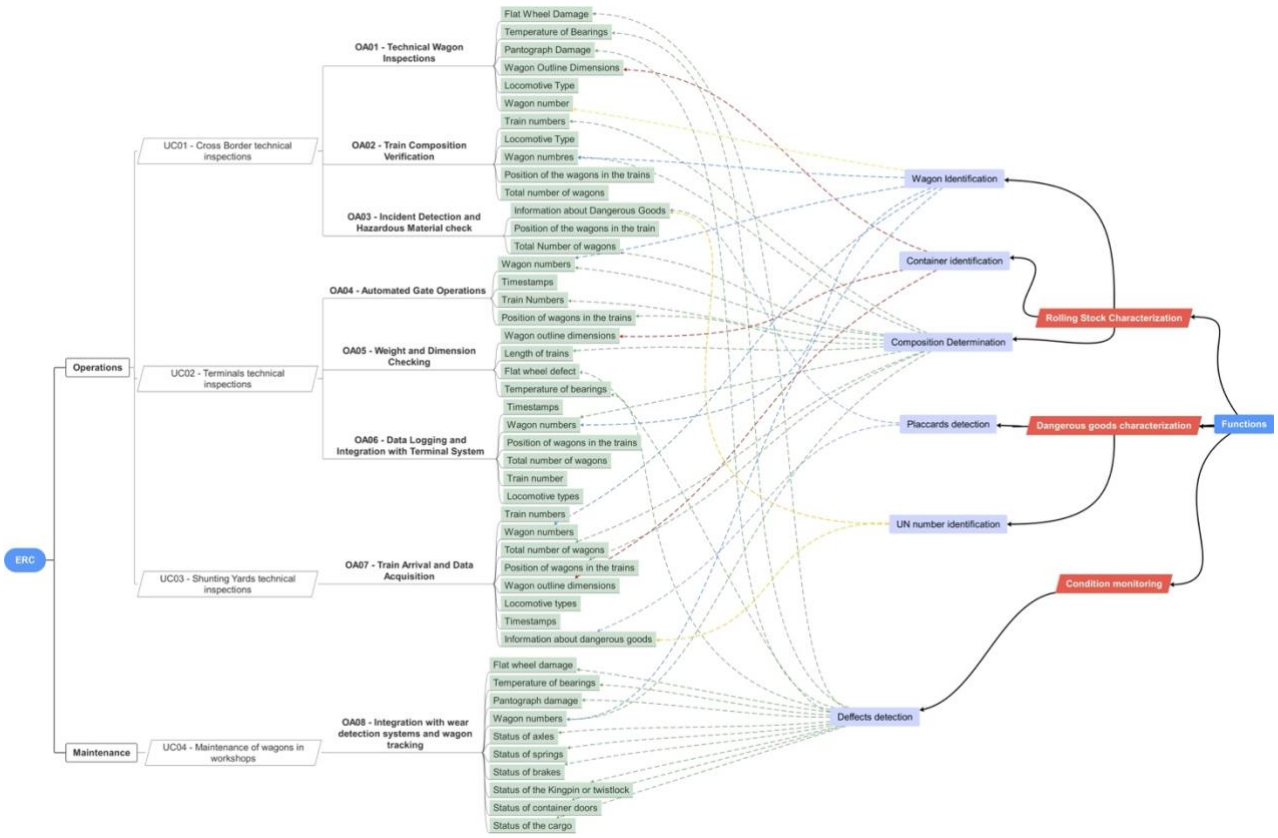

Figure 6 Operational Capabilities and System Functions relation

6.4.2 Data Standardization and Interoperability requirements

In this section, based on the system analysis carried out before, a set of requirements are established in order to include all the relevant information for the ERC processes to the CDM. These requirements outline:

- **Semantics of the information that needs to be included:** Details the meaning of the information, which is essential to verify whether this information has already been defined in related initiatives or other standard Data Models.
- **Data format:** Which indicates how this information must be formatted when used by the systems or entities that takes place in the processes.
- **Associated standards:** This field aims to identify which initiatives already define the information considered. In case any initiative defines it, it must be considered whether to use the definition or not. From this analysis, the following standards has been identified:
- **TAF TSI (Telematics Applications for Freight Technical Specifications for Interoperability):** TAF TSI is an EU regulation aimed at improving the interoperability of the European railway system. It focuses on freight telematics, including messaging and data exchange between different railway actors such as infrastructure managers and railway undertakings. It defines standards for train numbering, data format, and the operational train number.

- **EVN (European Vehicle Numbering Standard):** The European Vehicle Numbering system is a standardized method for identifying railway vehicles across the EU. Each vehicle is assigned a 12-digit number, ensuring uniformity and facilitating easier tracking and management of vehicles in the European rail network.
- **RID (Regulations concerning the International Carriage of Dangerous Goods by Rail)**: The RID standard is part of the Convention concerning International Carriage by Rail (COTIF). It regulates the transportation of hazardous goods by rail, ensuring safety by defining specific labelling, documentation, and operational requirements.
- **GCU (General Contract of Use for Wagons)**: GCU provides a standardized contract for the use of wagons across different European rail operators. It defines technical and operational requirements related to the condition, maintenance, and damage of wagons. It also provides the framework for fault detection and operational safety requirements.
- **ILU/BIC Code Standard (Intermodal Loading Unit/Bureau International des Containers)**: This standard provides a globally recognized system for identifying intermodal loading units, such as containers. The code consists of four letters and seven digits, ensuring consistency and ease of tracking across different transportation modes.

D25.1 Section 8.2.5.16: This section refers to a part of the Shift2Rail (S2R) program, specifically the document D25.1, which outlines research and innovations in rail transport, particularly for the freight and logistics sector. Section 8.2.5.16 focuses on complex measurements such as wheel profile data for condition-based monitoring.

Table 3 Requirements associated to Train entity

1. Train					
Requirement Code	***Requirement Name***	***Requirement Description***	***Data Format***	***Associated Standards***	***Criticality***
Req001	Train Number				Mandatory
Req001.1	Train Number in Cross Borders	In Cross Borders operations, it is mandatory to identify the train. Train identification cannot be supplied by the ERC system -is not physically identified in the train-. However, the system must consume this	String	ERA-TD-105: TAF TSI — ANNEX D.2: APPENDIX F — TAF TSI DATA AND MESSAGE MODEL https://www.era.europa.eu/system/files/2022-11/era_technical_document_taf_d_2_appendix_f.pdf	Mandatory

		information from existing systems, such as TMS. TAF/TSI defines the Operational Train Numer as the unique identifier of a train for traffic management purposes by the Dispatcher, GSMR services, etc.			
Req001.2	Train Number in Non-Cross Border Locations	Defines the identification of the train number outside cross-border operations. If the location allows to consume this information from external systems, such as TMS, the ERC system must use the defined standard. Otherwise, the system must generate a synthetic identifier, following the requirements	String	None	Optional
Req002	Total Number of Vehicles	Captures the total number of vehicles in the train.	Integer	None	Mandatory
Req003	Train Length	Length of the train in meters.	Floating-point number	None	Mandatory
Req004	Train Weight	Weight of the train in kilograms.	Floating-point number	None	Mandatory

Table 4 Requirements associated to Checkpoint entity

2. Checkpoint					
Requirement Code	***Requirement Name***	***Requirement Description***	***Data Format***	***Associated Standards***	***Criticality***
Req005	Checkpoint Identification	Identifies the checkpoint where data is collected. This identifier must be an enumeration of different aspects regarding the location.	enum		Mandatory
Req005.1	Checkpoint ID	Unique ID of the checkpoint.	String	None	Optional
Req005.2	Checkpoint Coordinates	Geolocation of the checkpoint (latitude, longitude, and altitude).	String	None	Mandatory
Req005.3	Checkpoint Location	Identifier of the checkpoint's location (Cross Border, Terminal, Shunting Yards, Workshops or Main Line.). Depending on this location, the ERC system will have different functionalities associated to the Use Cases covered	String	None	Mandatory

Table 5 Requirements associated to Timestamps

3. Timestamp Information					
Requirement Code	***Requirement Name***	***Requirement Description***	***Data Format***	***Associated Standards***	***Criticality***
Req006	Timestamps				Mandatory
Req006.1	Timestamp (Train)	Timestamp when the first vehicle was detected at the checkpoint.	String	None	Mandatory
Req006.2	Timestamp (Vehicle)	Timestamp when the vehicle was registered at the checkpoint.	String	None	Mandatory

Table 6 Requirements associated with Vehicle entity

4. Vehicle					
Requiremen	***Requirement***	***Requirement***	***Data***	***Associated Standards***	***Criticality***

t Code	*Name*	*Description*	*Format*		
Req007	Vehicle Number	12-digit vehicle ID adhering to the European vehicle numbering standard.	String	EU Directive (EU) 2016/797 Annex 6 definition of European Vehicle Number (EVN) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0797	Mandatory
Req008	Position of the wagon in the train	Index of the vehicle in the train. Integer index defined in TAF TSI appendix F https://www.era.europa.eu/system/files/2022-11/era_technical_document_taf_d_2_appendix_f.pdfhttps://www.era.europa.eu/system/files/2022-11/era_technical_document_taf_d_2_appendix_f.pdf	Integer	ERA-TD-105 Annex D.2 Appendix F https://www.era.europa.eu/system/files/2022-11/era_technical_document_taf_d_2_appendix_f.pdfhttps://www.era.europa.eu/system/files/2022-11/era_technical_document_taf_d_2_appendix_f.pdf	Mandatory
Req009	Locomotive Type	Type of locomotive (electric, diesel, or dual).	Enum	EU Directive (EU) 2016/797 Annex 6 definition of European Vehicle Number (EVN) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0797	Mandatory
Req010	Wagon Payload Weight	Weight of the wagon payload in kilograms.	Floating-point number	None	Optional
Req011	Wagon Outline Dimensions	Dimensions of the wagon's load, indicating if the load is outside the allowed gauge.	To be decided	None	Optional
Req012	Out-of-Gauge Load	Indicates if the wagon's load is outside the allowed gauge.	Boolean	None	Optional
Req013	Dangerous Goods (Wagon)	Hazard identification as described in RID specification	List of String	RID (Regulation concerning the International Carriage of Dangerous Goods by Rail) https://otif.org/en/?page_id=1105	Mandatory
Req014	UN numbers (Wagon)	Four digits code representing the hazardous material on a wagon or container, as described in RID specification	String	RID (Regulation concerning the International Carriage of Dangerous Goods by Rail) https://otif.org/en/?page_id=1105	Mandatory
Req014	Perishable Goods (Wagon)	Information about perishable goods.	String	None	Optional
Req015	Vehicle Damages	Information about damages to the wagon.			Optional
Req015.1	Irregularity Code	A unique numerical identifier representing	String	Appendix 9 of the GCU (2023)	Optional

		specific wagon irregularities as defined in Annex 1 of the GCU (2023). In case Irregularity code described in Appendix 9 of the GCU (2023) is not identified, this field is empty.		https://gcubureau.org/wp-content/uploads/Contract/2021/20210101_A09_EN.pdf	
Req0152	Irregularity Position			-	Optional

Table 7 Requirements associated to ILU entity

5. ILU (Intermodal Loading Unit)					
Requirement Code	***Requirement Name***	***Requirement Description***	***Data Format***	***Associated Standards***	***Criticality***
Req016	ILU Code	Four-letter and seven-digit Identifier for Intermodal Loading Units (ILUs) according to UIRR (2011) standard.	String	UIRR EN 13044 standard https://www.ilu-code.eu/en/standards	Mandatory
Req017	ILU Position in wagon	Position of the ILU on the wagon.	Integer	None	Mandatory
Req018	ILU Type	Type of ILU (container, semi-trailer, etc.).	Enum	None	Mandatory
Req019	Dangerous Goods (ILU)	Hazard identification as described in RID specification	List of String	RID (Regulation concerning the International Carriage of Dangerous Goods by Rail) https://otif.org/en/?page_id=1105	Mandatory
Req020	UN number (ILU)	Four digits code representing the hazardous material on a wagon or container, as described in RID specification	String	RID (Regulation concerning the International Carriage of Dangerous Goods by Rail) https://otif.org/en/?page_id=1105	Mandatory
Req021	Perishable Goods (ILU)	Information about perishable goods related to ILUs.	String	None	Optional
Req022	ILU Damages	Information about damages to ILUs	List of enums	None	Optional

Table 8 Requirements associated to specific components

6. Component-Specific Requirements				
Requirement Code	***Requirement Name***	***Requirement Description***	***Data Format***	***Associated Standards***
Req023	Wheel Damages	Information about damages in wheels.		

Req023.1	Wheel ID	Identifier for the wheels.	String	Appendix 9 of the GCU (2023) https://gcubureau.org/wp-content/uploads/Contract/2021/20210101_A09_EN.pdf
Req023.2	Wheel Contact Force	Peak wheel-rail contact force in kN.	Floating-point number	Appendix 9 of the GCU (2023), section 5.4.4.3.1. https://gcubureau.org/wp-content/uploads/Contract/2021/20210101_A09_EN.pdf
Req023.3	Wheel Flatness	Indicates if there is a flatness issue with the wheel.	Boolean	Appendix 9 of the GCU (2023) https://gcubureau.org/wp-content/uploads/Contract/2021/20210101_A09_EN.pdf
Req023.4	Wheel Profile	11 measurements for the wheel profile.	To be decided	D25.1 section 8.2.5.16
Req023.5	Wheel Degradation	Represents degradation level of the wheel.	To be decided	None
Req023.6	Wheel Overheating	Indicates overheating of the wheel.	Boolean	None
Req029	Bearing Temperature	Temperature of the wheel bearings.	Floating-point number	Appendix 9 of the GCU (2023), section 1.8.3. https://gcubureau.org/wp-content/uploads/Contract/2021/20210101_A09_EN.pdf
Req030	Spring ID	Identifier for the spring.	String	None
Req031	Spring Vibration Frequency	Frequency of spring vibration in Hz.	Floating-point number	None
Req032	Spring Cracks	Indicates if there are cracks in the spring.	Boolean	None
Req033	Brake ID	Identifier for the brake.	String	None
Req034	Brake Status	Status of the brake (locked/unlocked).	Boolean	None
Req035	Brake Pad Thickness	Thickness of the brake pad.	Floating-point number	None
Req036	Axle ID	Identifier for the axle.	String	None
Req037	Axle Status	Status of the axle (normal/abnormal).	Boolean	None
Req038	Cargo Temperature	Temperature of the cargo.	Floating-point number	None
Req039	Cargo Relative Humidity	Relative humidity of the cargo.	Floating-point number	None
Req040	Cargo Pressure	Pressure of the cargo (unit to be defined).	Floating-point number	None
Req041	Kingpin	Status of the kingpin	Boolean	None

	Status	(locked/unlocked).		
Req042	Twistlock Status	Status of the twistlock (locked/unlocked).	Boolean	None
Req043	Tarpaulin Cover Status	Status of the tarpaulin cover (damaged/undamaged).	Boolean	None
Req044	Floor Status	Status of the floor (damaged/undamaged).	Boolean	None
Req045	Pantograph Damages	Information about damages to the pantograph.	List of enums	None

6.4.3 Data Standardization and Interoperability Specification

Following on from the requirements of the previous chapter, this section will present, by way of conclusion, a specification of the information in which an associated initiative has been identified that can define it.

Req001.1 - Train Number Specification according to TAF TSI for Cross Borders Operations

The train number is defined in the context of TAF TSI (Technical Specifications for Interoperability for Telematics Applications for Freight. Here, the element ***OperationalTrainNumber*** is defined as the primary identifier used by the infrastructure manager and railway undertaking for coordinating train paths and movements, specially across international borders.

OperationalTrainNumber is used by several Sub-Systems of the TAF-TSI specification, detailed in the following relational diagram.

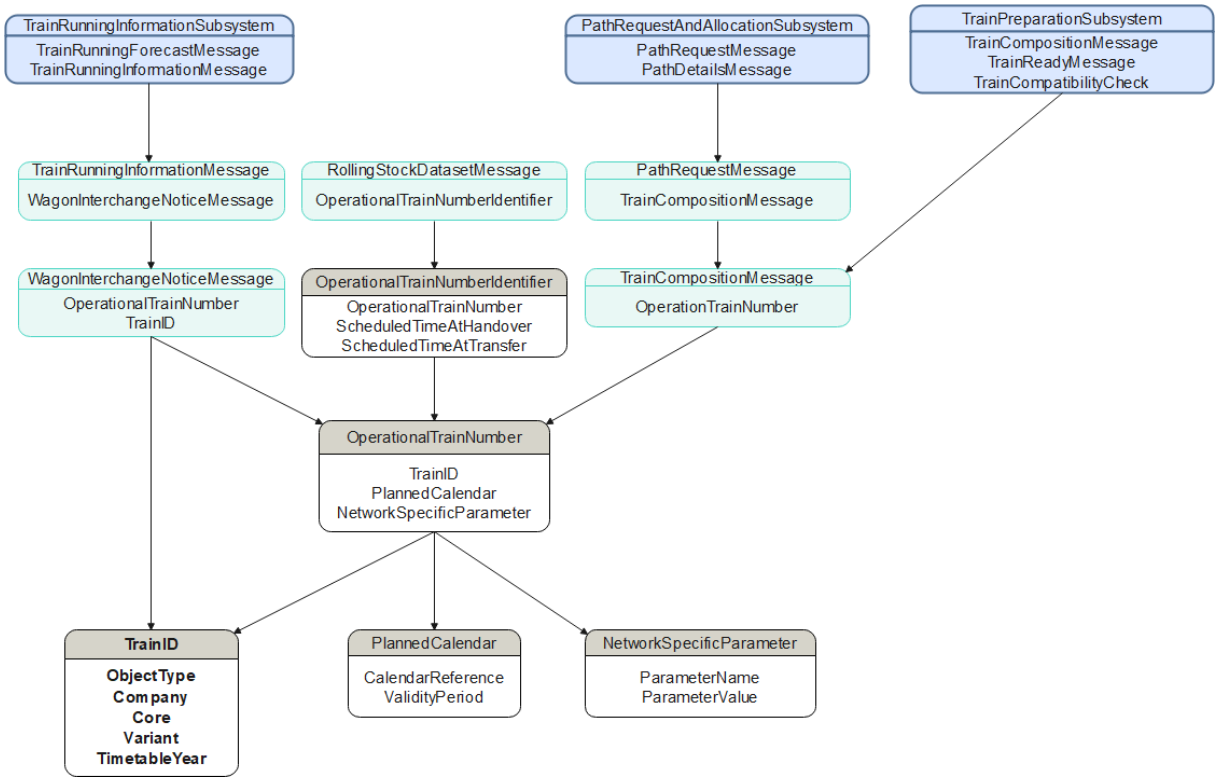

Figure 7 ERA-TD-105: TAF TSI — ANNEX D.2: APPENDIX F — TAF TSI DATA AND MESSAGE MODEL - Operational Train Number relations.

Based on *ERA-TD-105: TAF TSI — ANNEX D.2: APPENDIX F — TAF TSI DATA AND MESSAGE MODEL* and the requirements defined in Work Package 29, *TrainID* element contains the required information to identify a train in the context of European Railway Checkpoint Use Cases.

The message shall be included in the *Conceptual Data Model* according to the following specification:

Table 9 Train Number specification according to TAF TSI standard.

Req001.1 - Train Number according to TAF TSI		
Specification	*ERA-TD-105 Annex D.2 Appendix F definition of OperationalTrainNumber might be used in operations where the consistent identification and monitoring of train is mandatory. Speciffically, in cross borders oprations the identification of trains is required.*	
Information Required		
TrainID	*Unique identifier of the train*	

Field Name	Description	Type
ObjectType	*Specifies the type of the object (e.g., train, path).*	*String*
Company	*Identifies the company operating the train.*	*String*
Core	*Core identification part of the train number.*	*String*
Variant	*Identifies any specific variants of the train.*	*String (Optional)*
TimetableYear	*Indicates the timetable year for which the train number is valid.*	*Integer (4 digits)*

Req007 – VehicleNumber specification according to the European Numbering Vehicle Standard

The vehicle number, referred to as the European Vehicle Number (EVN), is a 12-digit numeric identification code used to uniquely identify each railway vehicle within the European rail system.

Based on *EU Directive (EU) 2016/797 Annex 6 "definition of European Vehicle Number (EVN)"* and Work Package 29 Requirements, *EVN* element contains the required information to unambiguously identify every railway vehicle in the context of the European Railway Checkpoint use cases.

The message shall be included in the *Conceptual Data Model* according to the following specification:

Table 10 EuropeanVehicleNumber specification according to EU 2016/797.

Req007 - EuropeanVehicleNumber according to EU Directive (EU) 2016/797 Annex 6		
Specification	*EU Directive (EU) 2016/797 Annex 6 definition of European Vehicle Number (EVN)* can be used to ensure interoperability and compliance with safety standards across different national rail systems.	
Information Required		

EVN	*Unique identifier of each railway vehicle within the European Rail System*	
Field Name	Description	Type
Interoperability Capability and Vehicle Type	*Identifies the vehicle type and its interoperability capability.*	*Integer (2 digits)*
Country Code	*Represents the country where the vehicle is registered.*	*Integer (2 digits)*
Technical Characteristics	*Describes technical characteristics of the vehicle.*	*Integer (4 digits)*
Serial Number	*A unique serial number assigned to the vehicle.*	*Integer (3 digits)*
Check Digit	*A digit calculated from the previous digits to ensure validity.*	*Integer (1 digit)*

Req008 – Position of the wagon in the train according to TAF TSI

The position of a wagon in a train is defined within the TAF TSI framework as part of the TrainCompositionMessage. Specifically, this message includes information about the order of wagons within the train, their technical details, and their association with the OperationalTrainNumber. The composition of the train must be communicated between the responsible railway undertakings (RUs) and infrastructure managers (IMs) and updated whenever changes occur to the train's makeup.

The Wagon Position is specified through the *WagonSequenceNumber* element, which indicates the position of each wagon within the train. This sequence is part of the overall train composition and is used to organize the wagons in the correct operational order.

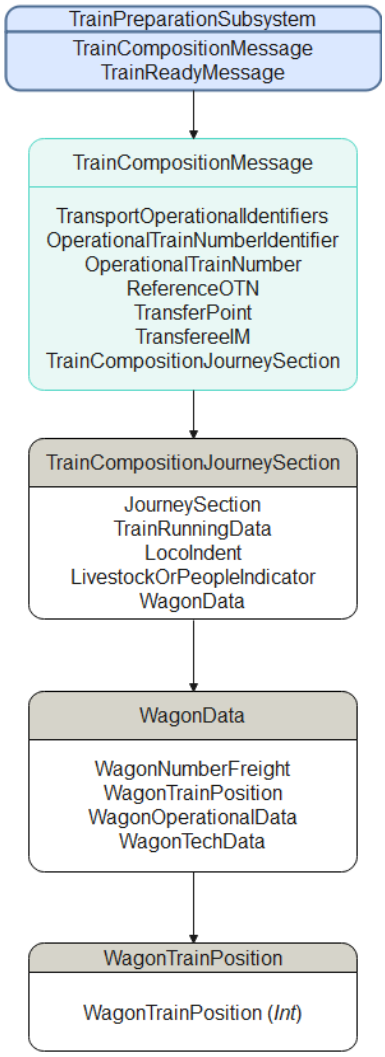

Figure 8 ERA-TD-105: TAF TSI — ANNEX D.2: APPENDIX F — TAF TSI DATA AND MESSAGE MODEL - Operational WagonTrainPosition

Based on *ERA-TD-105: TAF TSI — ANNEX D.2: APPENDIX F — TAF TSI DATA AND MESSAGE MODEL* and the requirements defined in Work Package 29, *WagonTrainPosition* element contains the required information to identify the position of a wagon in a certain train in the context of European Railway Checkpoint Use Cases.

The message shall be included in the *Conceptual Data Model* according to the following specification:

Table 11 WagonPosition specification according to TAF TSI standard.

Req08 - WagonPosition according to TAF TSI

Specification	*ERA-TD-105 Annex D.2 Appendix F definition of WagonTrainPosition shall be used in operations where it is mandatory to identify the position of a certain wagon in the train. This position is defined as a sequential number starting with the first wagon at the front of train as N°1.*	
Information Required		
WagonTrainPosition (Simple)	*Identifies the position of a wagon within a train.*	*Integer*

Req009 – Locomotive Type specification according to the European Numbering Vehicle Standard

LocomotiveType element is defined in the European Vehicle Number (EVN) description in *EU Directive (EU) 2016/797 Annex 6 "definition of European Vehicle Number (EVN)"*.

The message shall be included in the *Conceptual Data Model* according to the following specification:

Table 12 VehicleType specification according to EU 2016/797.

Req009 - VehicleType according to EU Directive (EU) 2016/797 Annex 6		
Specification	*EU Directive (EU) 2016/797 Annex 6 definition of European Vehicle Number (EVN)* includes the definition of the vehicle type in the second character of the code.	
Information Required		
Locomotive type	*Defined by the second digit of the EVN*	
Field Name	Description	Type
Vehicle Type	*Identifies the vehicle type*	*Integer*

Req013 – Dangerous Goods (Wagon) specification according to RID (Regulation concerning the International Carriage of Dangerous Goods by Rail)

The Hazard identification for dangerous goods wagons shall comply with the RID specification as described in the International Carriage of Dangerous Goods by Rail (RID). Each Dangerous Goods element is defined according to a class, as stated in https://otif.org/fileadmin/new/3-Reference-Text/3B-RID/RID_2023_e_30_June_2023.pdf.

The message shall be included in the Conceptual Data Model according to the following specification:

Table 13 : Dangerous Goods specification according to RID standard.

Req013 - Dangerous Goods according to RID		
Specification	*RID (Regulation concerning the International Carriage of Dangerous Goods by Rail) classifies Dangerous Goods among 9 different classes according to https://otif.org/fileadmin/new/3-Reference-Text/3B-RID/RID_2023_e_30_June_2023.pdf. This classification shall be considered by the ERC system.*	
Information Required		
Dangerous Goods Class (Simple)	*Identifies the class to which the dangerous goods belongs*	*String*

Specification also applicable to Req19 – Dangerous Goods (ILU).

Req014 – Dangerous Goods (Wagon) UN Number according to RID (Regulation concerning the International Carriage of Dangerous Goods by Rail)

The UN number specification for dangerous goods wagons shall comply with the RID specification, as outlined in the International Carriage of Dangerous Goods by Rail (RID) regulations (in https://otif.org/fileadmin/new/3-Reference-Text/3B-RID/RID_2023_e_30_June_2023.pdf). Each dangerous goods element is defined by a unique UN number, represented in orange placards located in vehicles (wagon and load), facilitating identification and compliance checks as specified in the RID documentation.

The message shall be included in the Conceptual Data Model according to the following specification:

Table 14 Dangerous Goods (UN) specification according to RID standard.

Req014 - Dangerous Goods (UN number) according to RID	
Specification	RID defines a UN number as a unique identifier for specific substances or articles considered dangerous for transport, according to https://otif.org/fileadmin/new/3-Reference-Text/3B-RID/RID_2023_e_30_June_2023.pdf. This number aligns with the global UN system and shall be utilized

	in the ERC system for standardized identification, tracking, and reporting.	
Information Required		
UN number (Simple)	*A unique, four-digit identifier assigned to a specific substance or article in the RID classification*	*String (4 digits)*

Specification also applicable to Req20 – UN number (ILU).

Req015.1 – Irregularity Code Specification according to GCU (2023)

The Irregularity Code for identifying operational issues with rail wagons shall comply with the standards outlined in Appendix 9 of the GCU (2023) (https://gcubureau.org/wp-content/uploads/Contract/2021/20210101_A09_EN.pdf). Each code represents a specific irregularity, allowing for systematic reporting and management of wagon issues during transit This information is critical in ERC associated use cases where the inspection of the rolling stock is mandatory.

Table 15 Irregularity Code specification according to GCU standard.

Req015.1 - Irregularity Code according to GCU Appendix 9		
Specification	Appendix 9 of the GCU (2023) (https://gcubureau.org/wp-content/uploads/Contract/2021/20210101_A09_EN.pdf) provides a detailed list of irregularity codes to standardize the identification and reporting of wagon irregularities. The ERC system must incorporate these codes in a structured format to align with GCU specifications and to support efficient data sharing across the railway network.	
Information Required		
Irregularity	*Identification of a damage or irregularity associated to a class described in Appendix 9 of the GCU (2023)*	
Field Name	Description	Type

Irregularity Code	A unique numerical identifier representing specific wagon irregularities as defined in Annex 1 of the GCU (2023). In case Irregularity code described in Appendix 9 of the GCU (2023) is not identified, this field is empty.	*String (3 digits)*
Irregularity Description	*Descriptive text associated with each Irregularity code. In case Irregularity code described in Appendix 9 of the GCU (2023) is not identified, this field must contain a string describing the irregularity.*	*String*

Req016 – ILU Code Specification According to EN 13044

The ILU (Intermodal Loading Unit) Code is a standardized identification for non-ISO containers, swap-bodies, and semi-trailers used in European combined transport.

Based on *EN 13044 standards* and Work Package 29 Requirements, *ILU code* element contains the required information to unambiguously identify every Intermodal Loading Unit in the context of the European Railway Checkpoint use cases.

The message shall be included in the *Conceptual Data Model* according to the following specification:

Table 16 ILU specification according to UIRR standard.

Req016 - ILU code according to UIRR (2011)		
Specification	The ILU Code is defined according to the EN 13044 standard for identifying intermodal loading units within European transport. The structure comprises an Owner Key (three letters), a Product Group Key (one letter), a Registration Number (six digits), and a Check Digit (one digit). For full details on the ILU Code format, refer to EN 13044.	
Information Required		
ILU code (Simple)	*The full ILU Code composed of the Owner Key, Product Group Key, Registration Number, and Check Digit according to EN 13044 standard definitions*	*String*

Req023 – Wheel Damage Specification According to GCU Standard

Wheel damage on railway wagons shall be identified, recorded, and classified following the guidelines provided by the GCU (General Contract of Use for Wagons) standard. Each type of wheel damage is documented with specific attributes to ensure consistent assessment and reporting.

The message shall be included in the *Conceptual Data Model* according to the following specification:

Table 17 Wheel Damages specification according to GCU standard.

Req023 - Wheel Damages according to GCU Standard		
Specification	GCU standard defines wheel damages assessment criteria for railway wagons, including identification, contact force, and flatness indicators	
Information Required		
Wheel damage	*Identification of a wheel damage associated to a criteria described in GCU (2023) standard.*	
Field Name	Description	Type
Wheel ID	*Unique identifier for each wheel on the wagon.*	*String*
Wheel Contact Force	*Peak wheel-rail contact force, measured in kN, for monitoring purposes.*	*Floating-point number*
Wheel Flatness	*Boolean indicator for detecting wheel flatness issues.*	*Boolean*

7 Strategies to enhance Data Management

7.1 Strategies to enhance data quality

As described in section 5.3, achieving high data quality is important for the improvement of railway freight. In this chapter, we will discuss how data governance in general and data quality management in particular can contribute to this goal.

7.1.1 Data Governance

Data governance in railway transport is a crucial framework to ensure the effective management, quality, and security of data across the entire transportation system. As the railway industry increasingly relies on digital technologies for operations, schedule management, and customer service, establishing robust data governance policies becomes essential to enhance decision making, improve safety, and optimize resource allocation.

One of the core components of data governance in this context is data quality management. This involves systematic processes for data collection, validation, and maintenance to ensure accuracy and reliability. In railway transport, where safety and operational efficiency are paramount, high-quality data facilitates better train scheduling, predictive maintenance of rolling stock, and real-time monitoring of infrastructure health. By implementing strict data quality standards, railway operators can significantly reduce the risks associated with data errors, ultimately leading to improved service delivery and customer satisfaction. Data quality management is further discussed in section 7.1.2.

Another important aspect of data governance in railway transport relates to compliance and regulatory requirements. Railways must adhere to a myriad of regulations concerning safety, data privacy, and environmental protection. Effective data governance ensures that data practices meet these legal requirements while also aligning with industry best practices. This can include the establishment of clear data ownership, documentation processes, and access controls to protect sensitive information linked to rail operations, employee performance, and customer data.

Fostering a culture of accountability and transparency within organizations is vital for successful data governance in railway transport. This includes engaging various stakeholders—such as data stewards, engineers, and business analysts—in governance initiatives, promoting data literacy, and ensuring that everyone understands their roles in managing data. By cultivating this collaborative environment, railway organizations can enhance their overall data management practices, improve operational resilience, and drive innovation through data-driven decisions.

7.1.1.1 Addressing Challenges

The railway transport sector faces unique challenges in data governance due to the vast amount and diversity of data generated from various sources. This includes operational data from train control systems, customer data from ticketing systems, and maintenance data from various assets. One significant challenge is data silos, where data is isolated within departments, leading to inefficiencies and missed opportunities for actionable insights. Effective data governance helps to break down these silos by establishing standardised protocols for data sharing and integration across departments, enabling a holistic view of operations.

Additionally, the complexity of railway networks can create difficulties in data accuracy. Train schedules, for instance, can be affected by factors such as weather conditions or infrastructure issues, making real-time data updates essential. A robust data governance framework can include automated data quality checks and real-time monitoring systems, allowing for swift corrections and adjustments that enhance operational reliability.

7.1.1.2 Technological Integration

Technological advancements play a crucial role in strengthening data governance in railway operations. The adoption of IoT (Internet of Things) devices across trains and tracks generates continuous streams of data that can be leveraged for predictive analytics. For example, within WP29 related to checkpoints, this technology is applied in the onboard wagon monitoring system. Additionally, sensors on rolling stock can monitor wear and tear, allowing for predictive maintenance that minimises downtime and costs. However, to fully capitalise on this data, railway operators must implement stringent data governance policies to ensure data is accurately collected, securely stored, and properly analysed.

Moreover, the integration of advanced analytics and machine learning solutions can elevate the capabilities of data governance. Predictive analytics can provide insights into passenger trends, operational inefficiencies, and risk management, assisting in strategic planning. However, these technologies require high-quality input data, underscoring the importance of well-defined data governance principles.

7.1.1.3 Strategic Benefits

Implementing effective data governance can lead to several strategic advantages for railway transport organisations. First, it enhances operational efficiency by enabling better decision making through access to accurate and timely data. By leveraging data for real-time operational adjustments—such as re-routing trains or optimising schedules—railway companies can reduce delays and improve service reliability.

Safety is another critical area where data governance could have a transformative impact. A well-governed data environment supports adherence to safety regulations and helps in conducting thorough investigations of incidents by providing a reliable data trail. Enhanced data tracking and reporting can improve compliance with industry standards and regulations, thus safeguarding public safety and minimising liability.

Furthermore, enhanced data governance fosters better customer experience through improved service personalisation. With accurate customer data, railway operators can tailor marketing efforts, provide timely information on services. This can lead to increased customer loyalty and satisfaction.

In summary, data governance in railway transport requires a multifaceted approach that addresses challenges such as data silos, accuracy, and compliance while leveraging technological advancements for strategic benefits. A strong data governance framework ultimately leads to more efficient operations, improved safety, and an enhanced customer experience, positioning railway operators to thrive in a competitive landscape.

7.1.1.4 Summary

In summary, here are some key aspects:

- **Data Quality and Integrity**: Ensuring that data collected from various sources, such as ticketing systems, train schedules, and maintenance records, is accurate and reliable. This involves regular audits and validation processes.
- **Data Accessibility**: Establishing protocols that allow relevant stakeholders, including transport operators, regulatory bodies, and third-party service providers, to access necessary data while ensuring that sensitive information is protected.
- **Compliance and Regulation**: Adhering to legal and regulatory requirements, such as those set by the European Union Agency for Railways (ERA) and national regulatory bodies. This includes data protection laws like the General Data Protection Regulation (GDPR), which affects how personal data is handled.
- **Data Security**: Implementing measures to protect data from breaches and unauthorised access. This is particularly important in combating cyber threats that can impact transportation systems.
- **Interoperability**: Promoting standards and protocols that allow different railway operators and infrastructure managers across Europe to share data and systems seamlessly. This is crucial for enhancing operational efficiency and passenger experience, especially in a cross-border context.
- **Data Stewardship**: Designating roles and responsibilities for managing data across various departments within a railway organisation. This includes data custodians who oversee data governance practices and ensure compliance.

- **Innovation and Analytics**: Leveraging data analytics tools to draw insights from operational data, which leads to improved decision-making, predictive maintenance, and enhanced customer service.
- **Stakeholder Engagement**: Involving relevant stakeholders, such as local authorities, transport operators, and customers, in the data governance process to ensure that the data serves the needs of all parties involved.

By establishing robust data governance frameworks, railway operators in Europe can enhance the efficiency and reliability of their services, ultimately leading to a better experience for customers and an overall improvement in rail transport operations.

7.1.2 Data quality management

As mentioned in section 7.1.1, an important part of data governance is data quality management, that is, processes to ensure that data are of high quality. Below, we will discuss some of the aspects of data quality listed in section 5.3 and how they can be achieved, and examples from this project will be given.

Some basic aspects of data quality are correctness (absence of errors), completeness and consistency. These characteristics are often hard to verify without an independent source of information (which is known to be of very high quality) with which to compare the data, while it is generally easy to check if a certain piece of data follows a standard (for example, using the check digit to ensure that a vehicle number is valid according to the standard for European vehicle numbers). Within Europe's Rail, the conceptual data model (discussed in chapter 8) will support data consistency. How strictly requirements on correctness, completeness and consistency must be enforced may depend on the intended use of the data.

Ensuring high data quality is a combined responsibility of the systems producing the data and the systems transmitting and receiving them. If a piece of data is suspected or found to be invalid, incorrect, incomplete or inconsistent, the question arises how this should be handled. Possible strategies are (a) to allow the data to be stored and possibly transmitted, but to log or flag that an error has been detected, (b) to try to adjust or complete the data manually or automatically, (c) to discard only the problematic piece of data (for example, an invalid vehicle number), or (d) to discard a larger amount of data (for example, to discard all information about a vehicle passing a checkpoint if its vehicle number is found to be invalid).

As an example of how these strategies for data quality management can be applied, consider the data collected by checkpoints. For some of these data (for example, vehicle numbers and information about dangerous goods), it is often possible to make comparisons to data from other sources. However, there is no general guarantee that the data provided by these other sources are of higher quality. Indeed, an important reason

for developing checkpoints is that there is a need for improvement of the data quality in relation to the current situation. Therefore, it is not advisable to use data from another source to automatically adjust or complete data from checkpoints just because there is a disagreement between the sources. However, if a piece of data provided by the other source is valid according to a standard while its counterpart from a checkpoint is not, it is at least reasonable to consider the piece of data from the other source to be more likely. Also, the other strategies for handling invalid, incorrect, incomplete or inconsistent data listed above are applicable to data from checkpoints.

If there is no other source of information with which to do comparisons, it might be hard to distinguish between incorrect data and actual anomalies. For example, an unusually high temperature may be due to an incorrect reading but may also indicate that there actually is some problem with a vehicle. Therefore, it is wise not to discard such data without checking them against reality.

Furthermore, even though checking data collected by checkpoints against data from other sources might be problematic in an operative setting, it is a very useful method during system development. If a disagreement is found there, it is an indication that the data from the checkpoint must be manually checked against the corresponding images, to find out if the image quality or the image analysis must be improved (and how), or if it for some reason (such as characters being covered by graffiti) is impossible to expect better data quality.

Another important aspect of data quality is timeliness, that is, that data should be made available without unnecessary delays, and at the very least before they are needed by consuming systems. This requires that all systems that produce or transmit data are reasonably fast and have a high uptime. In particular, data that are intended to describe a current situation must be made available quickly and be updated as soon as the situation changes. For example, one key input for the prediction of arrival times in WP28 is an up-to-date operational timetable. Any changes to the planned route, time and so on must be communicated to the connected systems in order for the prediction to reference this up-to-date information.

7.2 Strategies to enhance data availability

7.2.1 Data storage solutions

Although the selection of storage solutions is independent from the topics developed in the other WPs, and the data sharing process itself shall be carried out technically through the use of data connectors (by following IDSA recommendations), it is also important to review transversal aspects related to data storage solutions. Data storage is relevant for both the data provider (where the data generated in the source is stored) and also for the data consumer (where the data received is stored). Hence, the data storage technology itself does not need to be mandatorily harmonized as there might be several equally valid alternatives. However, it is considered relevant to highlight some critical aspects related

to data storage, that if not taken into account, could prevent the entities from exchanging data (e.g. due to data loss). With this objective, first an overview of storage solutions and methods are provided. Then, critical aspects like safety and security of storage systems and Data Protection and Backup Plans are reviewed, as good practices that should be applied as strategies to enhance data availability.

Introduction to Storage Solutions

Data storage solutions refer to methods or systems designed to store and access data in an electronic format that can be processed by machines. The main goal of these solutions is to provide convenient and reliable data storage and Access.

Data storage methods can be categorized into several key types: Direct Attached Storage (DAS), Network Attached Storage (NAS), Hybrid. Each of these methods offers distinct advantages, drawbacks, requirements, technologies, etc.

An overview of the possible storage systems that can be used in the developments of the other Seamless WPs in the FP5-TRANS4M-R project is given below.

Storage Methods

a. **Direct Attached Storage (DAS):**

- Definition: Direct Attached Storage, also known as Local Storage, refers to storage infrastructure physically located within an organization's facilities. It is typically in a nearby area and directly connected to the machine accessing it. This storage is generally accessible to only a single machine, the connected one.
- Requirements: Local servers with physical hardware such as Hard Disk Drives (HDDs), Solid State Drives (SSDs) or Network-Attached Storage (NAS) devices and Storage Area Networks (SAN). This leads on a significant upfront investment in hardware, dedicated IT teams for maintenance, physical space, and regular software updates.
- Availability: Availability is limited by hardware reliability and physical access. As well, data is only available if the hardware is operative (does not ensure continuous access to data). Users are usually individuals or small businesses. It is a good option for organizations with many security requirements and sensible data.
- Advantages: Full control over data, better for organizations with stringent data residency or regulatory requirements. DAS is more affordable than NAS.
- Drawbacks: Scalability issues (can struggle, slow down, or fail when the workload grows if infrastructure or algorithms are not optimized), high capital expenditures (CapEx), sharing difficulties and limited availability.

b. **Network Storage:**

- Definition: Network Storage, also known as Cloud storage, refers to storage which data is stored on remote servers and accessed over the internet. This solution allows multiple machines to share storage over a network and centralize data, which makes it better for data sharing and collaboration. There are two common network-based storage: Network Attached Storage (NAS), which is usually a single device; and Storage Area Network (SAN), which can be a network of multiple devices (not necessarily of the same type).
- Requirements: Stable internet connection, subscription to a cloud service provider (Amazon Web Services (AWS) S3, Google Cloud Storage, Microsoft Azure Blob Storage, etc.), and suitable security and safety protocols for data access.
- Availability: It has high availability with options for redundancy (which means having multiple copies of data or extra hardware components. This allows the system to continue functioning without data loss in case one part fails) and failover (which means a secondary system will assume the functions of the primary system if it becomes unavailable. This is a backup operational mode). Cloud providers often guarantee 99.9% or higher uptime, meaning minimal disruptions. Users are usually individuals, businesses and enterprises.
- Advantages: Highly scalable, pay-as-you-go pricing model, global availability, easy integration with cloud-native services and easy to share data and collaborate.
- Drawbacks: Limited control over physical storage location, potential latency issues and reliance on third-party providers. NAS is less affordable dan DAS.

c. **Hybrid Storage:**

- Definition: A mix of both Direct Attached Storage and Network Storage, enabling organizations to balance control and scalability by having locally saved sensible data and cloud backups and other data.
- Requirements: Ability to manage cloud and local storage infrastructure simultaneously, additional overhead in data synchronization and Hybrid Cloud solutions (Azure Arc, AWS Outposts, etc.)
- Availability: Balances the benefits of both cloud and local storage and provides flexibility in data access and storage management. Users are usually businesses needing local and remote access.
- Advantages: Flexible storage options, with sensitive data stored locally and other data in the cloud for scalability.
- Drawbacks: Complexity in management, potential integration issues and cost associated with maintaining dual environments.

The following point highlights the critical aspects where good practices are recommended.

Safety and Security of Storage Systems

Data safety and security are critical in modern storage solutions. They are the protection against unauthorized access, breaches, and accidental loss of data.

a. **Encryption**

- At-Rest Encryption: Data stored on drives is encrypted. This prevents unauthorized accesses.
- In-Transit Encryption: The use of protocols like TLS/SSL for encrypting data during data transfer between users or between data centers.
- Technologies: AES-256 (Advanced Encryption Standard), SSL/TLS for web-based access.

b. **Access Control and managements**

- Role-Based Access Control (RBAC): Permissions are assigned to limit data access based on the role of the user (admin, user, read-only…).
- Multi-Factor Authentication (MFA): Adds an additional layer of security by requiring multiple forms of verification for access.
- Auditing logs: Keeping records of accesses and changes to data for accountability.

c. **Firewalls and Intrusion Detection**

Technologies: Advanced firewalls (protect unauthorized access), network segmentation and intrusion detection systems to monitor and block suspicious activities. Cloud providers offer built-in firewalls and monitoring tools (AWS GuardDuty, Azure Security Center) and Anti-Malware (software which detects and prevents malware infections).

Data Protection and Backup Plans

Robust backup and recovery plans are essential. In case of system failures, disasters, or cyber-attacks, they ensure data resilience.

a. **Backup Strategies and recovery**

- Full backup: A complete copy of all data is created at schedule intervals.
- Incremental backup: The backup is made up solely of the data that has changed since the last backup. This reduces backup time and storage requirements.

- Differential backup: Unlike the previous option, this backup is composed of the data that has changed since the last full backup was saved. This offers a balance between speed and data completeness.
- Technologies: Cloud-native backups (AWS Backup, Google Cloud Snapshots…), external drive systems or dedicated backup software (Veeam…).

b. **Backup solutions**

- On-site Backups: Physical backups are stored locally.
- Off-site Backups: Backups are stored in a different location (in the cloud or in another server) to protect against disasters.

c. **Disaster Recovery Plans (DRP)**

- Definition: Strategies to quickly recover and restore access to data in case of catastrophic failures (hardware failures or cyber-attacks).
- Recovery Time Objective (RTO): Defines how quickly systems must be restored after a failure.
- Recovery Point Objective (RPO): Defines the maximum tolerable amount of data loss (in terms of time) that a business can endure.
- Technologies: Off-site backups, cross-region replication (for example, AWS S3 Cross-Region Replication), and automated failover systems.

d. **Data Redundancy**

- Definition: The process of keeping multiple copies of data to ensure its availability and integrity. For best results, there should be three copies of the important files: the primary data plus two backups (preferably one backed up remotely and offsite).
- Technologies: RAID (Redundant Array of Independent Disks), geo-redundant storage (GRS), multi-cloud backups, or regional replication in cloud environments.

7.2.2 Data Integration Platforms for Situational Awareness

To facilitate the emergence of situational awareness building on multiple data sources, a data integration platform operated by a trusted party becomes critical. Such data sharing platform is to be seen as a data broker with data being provided from diverse sources, with possible capabilities to associate different data elements to each other to allow for consumers of data to access meaningful holistic interpretations as aggregations of the combination of the data as well as individual messages.

In local, regional, and global transports, data that needs to be shared are typically concentrated around events during the transport process; it could be a ship arriving at a

port (Lind et al., 2019), an airplane arriving at an airport, or cargo that is transhipped at a logistic centre. A flow of events with associated data makes up a so-called event stream.

Of outmost concern is to channel data associated to plans, estimates, and progress associated to events in the transport sector. This also requires standardized message format as the transport sector is operating within a self-organizing ecosystem (Watson et al., 2021) which requires a commonly agreed way of interacting.

One project initiative building upon the principles of a federated network of platforms advocated for by digital transport logistic forum (DTLF.eu) is Federated (www.federatedplatforms.eu). Within this initiative several use cases for supply chain visibility and infrastructure optimization relying on data sharing have been explored. Some of these use cases involved train operations. Each data sharing environment is based on that a group of actors reach an agreement on which data to share, when it should be shared, and in which format the data should be shared and consumed. As there are multiple actors involved providing data streams from multiple sources, these data streams need to be integrated, why a data integration platform is needed. When dealing with e.g. end2end supply chains there will be multiple data sharing environments involved. And to get a holistic view of the supply chain data needs to be shared between these different environments as to be facilitated by a federated network of platforms. The concerns in the transport sector for data sharing is also a domain specific interest for the efforts within the European data space initiatives.

Such data sharing has since 2019 been demonstrated within concepts for Collaborative Decision Making (CDM) adapted to the railway sector (Lind et al., 2022).[1] (Note that this meaning of "CDM" has nothing to do with the Conceptual Data Model discussed elsewhere in this deliverable.) Emerging from the passenger-centric rail related xCDM-initiative, StationCDM, and the cargo-centric rail related xCDM-initiative, YardCDM, RailwayCDM (R-CDM) has been coined. RailwayCDM is a concept for digital collaboration that aims to contribute to the optimisation of stations and yards as transport nodes, as parts of the larger transport system. Such a transport node can also cooperate with other nodes, which enables multi-modal coordination and synchronisation. Coordinated execution of railway yards, stations and other transport nodes with rail capabilities in the transport system is also important for contributing to the UN Agenda 2030 sustainability goals and for the strategy of moving freight transport from road to sea and rail.

The xCDM concepts are based on a common situational awareness being shared among the involved actors as a basis for better planning capabilities and as an improved decision-making basis for enhanced coordination of resources and infrastructure as well as increased information transparency. Improved situational awareness contributes towards the need to optimise the use of existing infrastructure because of ever-increasing freight volumes, which also requires a higher coordination ability among actors in the rail transport system but also to other nodes, regardless of the mode of transport. Data

sharing platforms are instrumental to be able to capitalize on an ever-growing accessibility to data streams.

The overall goal of the xCDM concepts is to enhance the predictability of departure and arrival times from/to the stations and yards to enhance punctuality in the rail transport system.

8 Conceptual Data Model

8.1 Implementation Plan

This chapter proposes a detailed implementation plan, indicating the coordination between the different work packages involved in the requisitioning and standardisation process. The implementation plan follows the flow outlined in Figure 9. This workflow will be eventually implemented for other use cases which involves data exchange by applying the CDM approach.

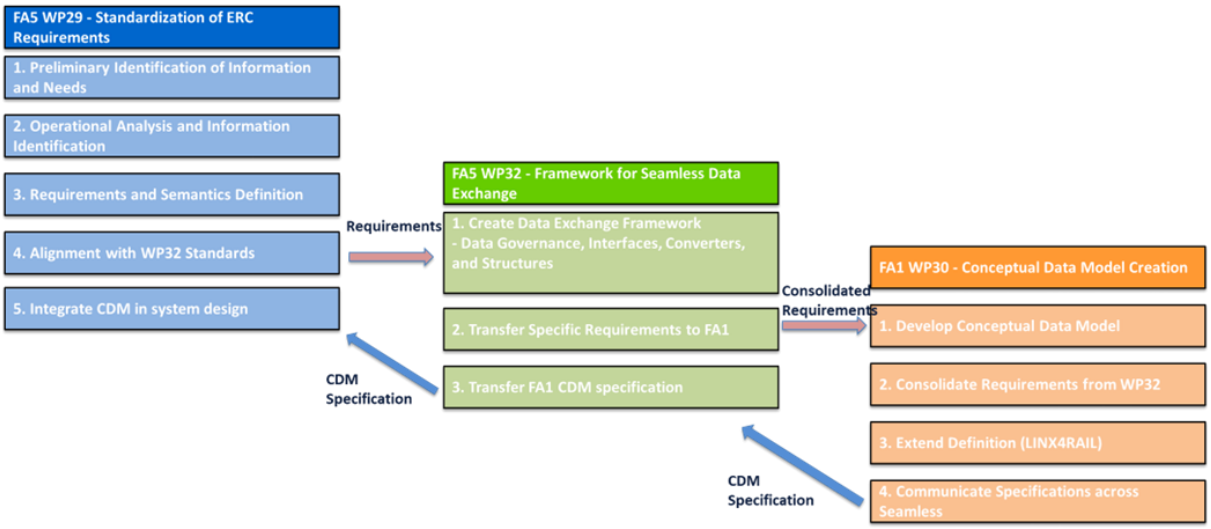

Figure 9 CDM implementation plan workflow.

8.1.1 FA5 WP29 - Standardization of ERC Requirements

Work Package 29, sub task 29.1.3 *"Conceptual Data Model"* aims to establish the specific requirements from the European Railway Checkpoint point of view for consolidating a Common Data Model. These requirements focus solely on the identification of needs, definition of those needs and identification of initiatives that can already meet them and can be integrated into the CDM.

In order to obtain this information, WP29 follows the workflow below:

- **Preliminary identification of information and needs by operational points**. Based on the work carried out in D25.1, which specifies various activities carried out at operational points, and which are of interest to digitise, this stage aims to identify the information that is essential or of interest in the development of these activities.
- **Operational Analysis and information identification.** The efforts undertaken in this sub-task are largely focused on organising the information from the previous point. To this end, this task implements an operational analysis, which tries to ensure that all identified use cases, activities, operations and needs of the different stakeholders are covered. This allows to clearly detail what information has to be

considered in each operational point, covering all cases completely and speeding up the design and requisitioning process. For example, the operational point 'Cross Border' identifies three operational activities, 'technical inspection', 'train composition check', and 'incident and dangerous goods detection'. In all of these activities, the operational analysis identifies the inability of the train to detect incidents and dangerous goods.

- **Requirements and semantics definition.** Work Package 29 communicates with work package 32 to transfer its requirements and to be able to define a data exchange in Seamless. In this sense, WP29 defines the previously identified information, giving a semantic meaning to this information and identifying initiatives that possibly define it.
- **Alignment with WP32 Standards**: Once CDM specification is obtained from Flagship Area 1, through the interface with Work Package 32, this Work Package integrates the feedback and aligns the conceptual model with WP32 standardization efforts for cohesion across frameworks.

8.1.2 FA5 WP32 - Framework for Seamless Data Exchange

- **Create a Data Exchange Framework**: Develop a robust framework to facilitate seamless data exchange across ERC systems, incorporating:
- **Data Governance**: Establish policies for data management and usage.
- **Data Interfaces**: Define standard data interfaces for interaction across systems.
- **Data Converters**: Implement converters to standardize diverse data formats.
- **Common Data Structures**: Use standardized data structures to ensure consistency.
- **Transfer Specific Requirements to FA1**: Work Package 32 is in charge of channelling specific requirements, associated with the standardisation of information for the completion of the CDM, to FA1 for further refinement and integration. For the definition of these requirements, the work package starts from the information transmitted by WP29 and the identified initiatives. WP32 establishes a system analysis to verify whether this information covers all the functionalities of the systems involved in the data exchange and analyses in depth the initiatives identified to check the feasibility of their use in the definition of data within the CDM.

8.1.3 FA1 WP30 - Conceptual Data Model Creation

Accordingly, Work Package 30 within Flagship Area 1 is responsible for collecting all elaborated requirements, consolidating a Conceptual Data Model and transmitting this specification for use.

- **Develop a Conceptual Data Model**: Initiate the creation of a foundational conceptual data model to support standardized data requirements.
- **Consolidation of requirements from WP32.**
- **Extend the Definition**: Utilize the existing LINX4RAIL framework to expand and refine the conceptual data model.
- **Communicating the specification to all work packages within Seamless**.

9 Defining Data Exchange for Use Case

9.1 Collection of use cases and the data sharing aspects from these use cases.

As part of the scope in task 32.2, it was defined the activity "collection of the use cases from WP26 to WP31", to examine the areas in which data exchange is required.

In the first step, the list of declared Use Cases (see picture below) in Seamless for WP26-WP27-WP28-WP29-WP31 was reviewed and analyzed from the perspective of which use cases needed alignment with WP32 for data exchange process.

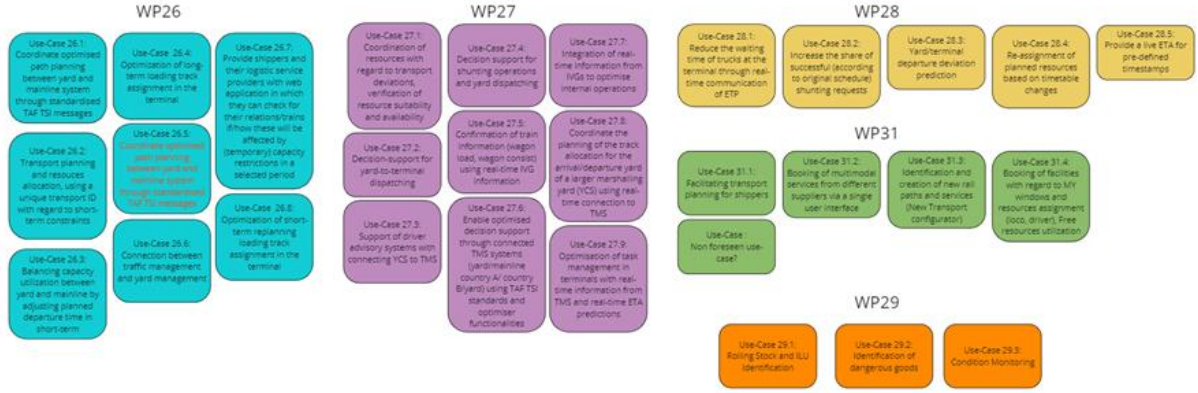

Figure 10 List of Use cases in FP5, Seamless Operations

For that purpose, it has been created a template to collect the basic necessary information by Use Case. The collection was divided into two phases:

1. Basic general information about the Use Case and its participants:

General Information

Name of the UC	
Description of the UC	
Demonstrator related	
WP involved	
Leader WP	
Leader UC	
Participants of the UC	

2. Technical information about systems, technologies, data, CDM, etc.

Use-Case Description

Needs	
Systems	
Technologies	

Data-related Information

Please provide information here and on the following tables below if you already know what kind of data exchange there will be for the use-case, e.g. who is the data provider and who receives the data, what kind of data is transferred etc. If you do not have all information, leave some fields empty.

Data Provider	***Data Owner***
1)	
Data Set Description	
Data Set Format	
Used Data Model	
Allowed data Usage	**Offered Transmission Mode**
☐ *Read* ☐ *Write* ☐ *Redistribute*	☐ *Once* ☐ *Push* ☐ *Pull* ☐ *Interval*

Data Consumer	**Data Owner**
Needed Data Model	
Intended Data Usage	**Needed Transmission Mode**
☐ *Read* ☐ *Write* ☐ *Redistribute*	☐ *Once* ☐ *Push* ☐ *Pull*

Service-related Information (optional)

Name	**Description**
[Name of the Service]	*[Describe the Service]*
☐ *Service implemented?* ☐ *Need dev-resources* ☐ *Needs orchestration?* ☐ *Certified Service?*	☐ Service dockerized? ☐ Service provides Interface? ☐ Service Consumer identified? ☐ SLA necessary?
Output Data Format	
Description of Output Data	

CDM

List of data	**Definition of the concept**
[Name of the data]	*[Describe the data and the range of values, scale, alarm levels, relationships, or the related regulation to define it]*

Data exchange concept

Data storage	*[Where the data will be storage]*
Frequency of exchange	*[How often the data should be exchange]*
Any other identified concept	*[How often the data should be exchange]*

During phase one, so far, only basic information has been collected for 16 Use Cases out of the 29 Use Cases, which are the ones which need alignment with WP32.
Phase two will be developed at the same time as the UCs, since they need to be advanced in order to obtain the necessary information for this task.
Demonstrations, as well as the progress of the rest of subtasks from Task 32.2, will be included in the next deliverable.

9.2 Demonstration plan - use cases

The 29 use cases will be showcased and demonstrated across European Rail Corridors (WP33 and WP34). These corridors will be supported by relevant enabler use cases, ultimately combining into a comprehensive corridor framework. The ScanMed corridor will serve as the primary showcase, with additional use cases from other corridors demonstrating solutions applicable to a generic corridor setup. For network management, collaboration with Destination 1 (FA1) and potentially RNE is essential. WP33 and WP34 will showcase and test seamless operations concepts, utilizing techniques developed in earlier work packages. The demonstrations' results will be evaluated, particularly in terms of data sharing and exchange. This task will summarize the lessons learned from Task 32.2 and those demonstrated in WP33 and WP34, such as interfaces and converters. The results from these demonstrations will also be included in the subsequent deliverable.

10 Conclusion

This deliverable, D32.2, provides recommendations to help reduce the technical and administrative challenges that hinder data sharing in the FP5-TRANS4M-R project, especially in the context of multimodal freight transport. By standardizing data formats and implementing advanced management tools, the recommendations will significantly enhance operational efficiency. Stakeholders can expect faster data exchanges, reduced errors, and improved decision-making capabilities, leading to smoother and more reliable multimodal freight operations. The main goal was to create a framework that ensures smooth data exchange and improves the quality and availability of data among different stakeholders, supporting the objectives of Flagship Project FP5.

In this document, we identified several key issues, such as data standardization, system compatibility, and data security, as well as the need to comply with regulations. Through discussions with stakeholders and an analysis of various use cases, we have outlined practical steps to overcome these issues. These include adopting new data management tools, strengthening data security measures, and aligning data formats to make sharing between different systems easier.

Although the recommendations are a strong step forward, some limitations remain. For example, the adoption of new technologies might take time, and not all stakeholders may be at the same level of digital readiness. Despite these challenges, the proposed actions will lead to better data availability, improved cooperation, and more efficient operations in the rail freight sector.

This section has provided the general principles for data sharing and data structure, by highlighting the main technological enablers that help to increase the trust of the data exchange process and provide a step forward in the interoperability of the data exchange. It has provided the detailed description regarding some of the core functionalities of the IDSA-based data spaces (as the Rail Data Space which is being deployed in FP1-MOTIONAL project and that is expected to be used within seamless use cases for data exchange), in particular related to: identity and access management, data auditing and monitoring and data sharing policies.

Next steps regarding the general principles include two action areas. First of all, to understand the process of developing and deploying an IDSA-based data connector compliant with Rail Data Space, which needs to be carried out by every company exchanging (providing or consuming) data in the seamless use cases. Second, to operationalize the process of applying the best practices in the data standardization and interoperability by defining and applying procedures to use the CDM approach for data definition and exchange.

There are many storage solutions and many characteristics to take into consideration. A

good storage solution must balance performance, availability, security, and cost-effectiveness. The selection of these solutions is independent of the system or development that exists in Seamless. Local storage solutions offer control while cloud storage solutions offer scalability and flexibility. Hybrid solutions balance control and scalability and are increasingly popular. Regardless of the chosen storage solution, security and safeness must be kept in mind. For these critical aspects, the application of good practices and possible harmonization is recommended in order to enhance data availability.

Ways to improve data quality are to check if data follow standards (for example, that vehicle numbers have a correct check digit) and to compare data to corresponding data from other sources. Data that are found to be incorrect, incomplete or inconsistent can be handled in different ways. Possible strategies are to log or flag the detection of an error, to do corrections and to discard smaller or larger amounts of data.

The data standardization and interoperability process developed in this deliverable aims to concisely specify how information related to the European Railway Checkpoint system should be handled within the CDM implementation carried out in FP1-MOTIONAL. This specification is based on a prior operational and system analysis, which concludes with an in-depth analysis of open initiatives and standards, such as TAF TSI, through which a specification has been determined to define how the detailed information must be interpreted and implemented in the CDM.

In conclusion, this deliverable offers a clear path to improving data sharing and quality in the FP5-TRANS4M-R project. The next phase will involve practical demonstrations to validate the proposed solutions. The conclusions and results of these demonstrations will be essential components of the next deliverable, ensuring the recommendations for specific use cases

11 References

1. Lind M., Ward R., Bergmann M., Haraldson S., Zerem A. (2019) Digitalizing the port call process, UNCTAD Transport and Trade Facilitation Series No. 13, UNCTAD (https://unctad.org/publication/digitalizing-port-call-process)
2. Lind M., Haraldson S., Lind K., Bergstrand J., Roos A., Lundgren M. (2022) The Role of Democratic Platforms in Transport System Innovation, 14th ITS European Congress, Toulouse, France, 30 May-1 June 2022 (https://maritimeinformatics.org/wp-content/uploads/2022/07/The-Role-of-Democratic-Platforms-in-Transport-System-Innovation.pdf)
4. Watson R. T., Lind M., Delmeire N., Liesa F. (2021), Shipping: A Self-Organising Ecosystem, in M. Lind, M. Michaelides, R. Ward, R. T. Watson (Ed.), Maritime informatics. Heidelberg: Springer.
5. https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4 , accessed on 15.10.2024
6. https://www.youtube.com/watch?v=R-DEkD7qkI4 , accessed on 15.10.2024

12 Appendix

12.1 Questions for the Stakeholders Survey

- **What systems are you involved with in the project TRANS4M-R?**
- **What kind of systems (for example: Traffic Management System) are your systems connected to (receiving and sending data to/from)?**
- **What are the issues and challenges (technical issues, confidentiality issues etc.) regarding exchanging or receiving data with/from other systems?**
- **What, in your opinion, are possible solutions to reduce the current problems and barriers for data exchange?**
-