CEF2 RailDataFactory

Deliverable 4.2 – Pan-European Railway Data Factory deployment planning and strategy proposal

Due date of deliverable: 31/12/2023

Actual submission date: 22/12/2023

Leader/Responsible of this Deliverable: Philippe DAVID (WP 4 lead) / SNCF

Reviewed: **Y**/N

Document status		
Revision	Date	Description
01	19/11/2023	First draft
02	15/12/2023	Consolidated draft
03	22/12/2023	Version submitted to the project officer

Project funded by the European Health and Digital Executive Agency, HADEA, under Connecting Europe Facilities Digital Grant Agreement 101095272		
Dissemination Level		
PU	Public	X
SEN	Sensitiv – limited under the conditions of the Grant Agreement	

Start date: 01/01/2023 Duration: 12 months

ACKNOWLEDGEMENTS

This project has received funding from the European Health and Digital Executive Agency, HADEA, under Connecting Europe Facilities Digital Grant Agreement 101095272.

REPORT CONTRIBUTORS

Name	Company
Philippe David	SNCF
Philipp Neumaier	DB
Wolfgang Albert	DB
Patrick Marsch	DB

Note of Thanks

We'd like to thank all Advisory board members who have left valuable comments for this deliverable.

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Furthermore, 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 author(s) and project consortium do not take any responsibility for any use of the information contained in this deliverable. The users use the information at their sole risk and liability.

Licensing

This work is licensed under the dual licensing Terms EUPL 1.2 (Commission Implementing Decision (EU) 2017/863 of 18 May 2017) and the terms and condition of the Attributions- ShareAlike 3.0 Unported license or its national version (in particular CC-BY-SA 3.0 DE).

Interne

EXECUTIVE SUMMARY

The document depicts a possible deployment strategy for the Pan-European Data Factory (PEDF), outlining assumptions, hypotheses, and a strategic phased deployment approach.

ABBREVIATIONS, ACRONYMS AND GLOSSARY

Abbreviation	Definition
AI	Artificial Intelligence
ATO	Automatic Train Operation
DF	Data Factory
ETSI	European Telecommunications Standards Institute
European organisation	European organisation from which the Pan European Data Factory consortium want approbation. This approbation may be necessary to get European funding if necessary.
GDPR	General Data Protection Regulation
GoA4	Grade of Automation 4
IM	Infrastructure Manager
PEDF	Pan-European Data Factory
RU	Railway Undertaking
Steering committee	Steering committee or comparable group of members able to take decisions in the PEDF regardless the chosen organization.
UIC	International Union of Railways

Interne

TABLE OF CONTENTS

LIST OF TABLES

Interne

1 INTRODUCTION

The European railway sector is on the verge to the strongest technology leap in its history, with many railway infrastructure managers (IMs) and railway undertakings (RUs) striving toward large degrees of automation in rail operation, and mechanisms to increase the capacity and quality of rail operation.

In particular, various railway companies – both IMs and RUs – and railway suppliers are currently working toward fully automated rail operation (so-called Grade of Automation 4, GoA4), for instance in the context of the Shift2Rail [1] and Europe's Rail [2] programs, in which sophisticated lidar and radar sensors as well as cameras are used to automatically detect and respond to hazards in rail operation, such as objects on the track or passengers in stations in dangerous proximity of the track. Another important use case is high-precision train localization by detecting static infrastructure elements and locating them on a digital map, as for instance covered in the Sensors4Rail project [3]. While the rail system has various properties that render fully automated driving principally easier than, e.g., in the automotive sector (for instance, railway motion is only one-dimensional, scenarios are typically much less complex than automotive scenarios, etc.), key challenges on the way to fully automated driving in the rail sector are that hazardous situations have to be detected much earlier due to long braking distances, and it is very challenging to collect and annotate sufficient amounts of sensor data with sufficient occurrences of relevant incidences to perform the required AI training and to be able to prove that the trained AI meets the safety needs.

For this, it is expected that single railway suppliers, IMs and RUs will not be able by themselves to collect and annotate sufficient amounts of sensor data for AI training purposes – but instead, an European data platform and ecosystem is required into which railway stakeholders (suppliers, Ims, Rus, railway undertakings, safety authorities, and others) can feed, process and extract sensor data, as well as simulate artificial sensor data, and through which the stakeholders can jointly develop and assess the AI models needed for fully automated driving.

Cross-border data exchange is crucial for railway undertakings, even if nationally different requirements exist. Through an improved use of technology, for example transfer learning or self-supervision learning with existing data, these national requirements can be partially resolved, and a significant acceleration can be achieved. As an example, transfer learning is a machine learning (ML) technique in which knowledge learned from one task is reused to improve performance on a related task. Among other things, cross-border data exchange enables seamless coordination of the development of fully automated driving and interoperability between different national railway networks and ensures efficient and smooth cross-border operations. The EU Directive (EU) 2016/797 [4] on the interoperability of the rail system provides guidelines and rules to promote such data exchange and ensures a standardised and effective approach across Europe.

1.1 AIM AND SCOPE OF THE CEF2 RAILDATAFACTORY STUDY

The CEF2 RailDataFactory study focuses exactly on vision of a pan-European RailDataFactory for the joint development of fully automated driving. The study, being co-funded through HADEA, aims to assess the feasibility of a pan-European RailDataFactory from technical, economical, legal, regulatory and operational perspectives, and determine key aspects that are required to make a pan-European RailDataFactory a success. For a better understanding of the studys aim and scope, please see Chapter 1.1 in Deliverable 1 [5].

Interne

1.2 Delineation from and Relation to other Works

The Shift2Rail project **TAURO** [6] also looks into the development of fully automated rail operation, for instance focusing on developing

- a common database for AI training;
- a certification concept for the artificial sense when applied to safety related functions;
- track digital maps with the integration of visual landmarks and radar signatures to support enhanced positioning and autonomous operation;
- environment perception technologies (e.g., artificial vision).

The difference of the CEF2 RailDataFactory project is that this puts special emphasis on the **pan-European Railway Data Factory backbone network and data platform** (located on the infrastructure side, but used for sensor data collected through both onboard and infrastructure side sensors) required for the Data Factory, and also investigates **commercial, legal and operational aspects** that have to be addressed to ensure that the vision of the pan-European RailDataFactory can be realized.

DB Netz AG and the German Centre for Rail Traffic Research (DZSF) have released OSDaR23, the first publicly available multi-sensor data set for the rail sector [7][8]. The data set is aimed at training AI models for fully automated driving and route monitoring in the railway industry. It includes sensor data from various cameras, infrared cameras, LiDARs, radars, and other sensors, recorded in different environments and operating situations, and annotated with labels for different objects and situations. The data set will be utilized in the Data Factory of Digitale Schiene Deutschland to train AI software for environment perception, and more annotated multi-sensor data sets will be created in the future.

The Europe's Rail Innovation Pillar **FP2 R2DATO project** [9], overall focusing on the further development of automated rail operations, also has a work package dedicated to the pan-European RailDataFactory. Here, however, the main focus is on creating first implementations of individual data centres and toolchains as required for specific other activities and demonstrators in the FP2 R2DATO project, and on developing an **Open Data Set**. A strong alignment between the CEF2 RailDataFactory study and the FP2 R2DATO pan-European RailDataFactory activities is ensured through an alignment on use cases and operational scenarios, though the actual focus of the projects is then different.

EU-wide research programs are being carried out on Flagship Project 2: “Digital & Automated up to Autonomous Train Operations” and in this context the European perspective is discussed. In addition, each country and each railway infrastructure provider have its own programs, where there is usually also an exchange within the Innovation and System Pillar in the R2DATO. The participants in this study also work in these bodies and try to reflect the European picture. Within the sector initiative “Digitale Schiene Deutschland”, Deutsche Bahn already started to set up some components of the data centre in Germany [10].

1.3 Aim and Structure of this Deliverable

This current document is the deliverable D 4.2 from WP4 of the CEF 2 RailDataFactory project.

This document aims at providing information about the deployment strategy and the possible methods to meet the target of a Pan-European Data Factory (PEDF). It complements Deliverable D 4.1 [11] that has already provided considerations on, e.g., different levels of membership and steps to be pursued for adding individual members to the PEDF. The remainder of the document is structured as follows:

- Chapter 2 elaborates on the notion of "national" Data Factories and formulates various hypotheses on the PEDF that are essential for the subsequently defined deployment strategies;
- Chapter 3 provides a possible deployment strategy, along the notion of potential interface and toolchain strategies;
- Chapter 4 lists risks related to these strategies incl. their possible mitigation;
- Chapter 5 provides a possible phased deployment plan, and finally,
- Chapter 6 concludes the deliverable.

2 ASSUMPTIONS BEHIND THE PROPOSED DEPLOYMENT STRATEGY

In this chapter, we first elaborate on some key assumptions that have served as a basis for the PEDF deployment strategy and the possible phased deployment listed in the subsequent chapters.

2.1 NOTION OF "NATIONAL" DATA FACTORIES

In the discussion on the PEDF, both within the CEF2 RailDataFactory study and beyond, there is often the underlying assumption that the PEDF basically constitutes an interconnection of "national" Data Factories. This somewhat implies (or at least could be interpreted in the direction) that in each country, there is one single Data Factory (DF), or at least one harmonised flavour of Data Factory, for instance driven by one major player, such as the main railway infrastructure manager.

However, in practice, there may also be multiple Data Factories within one country, driven by different players with different interests, and hence with different scope, timeline and properties. More specifically, it is envisioned that Data Factories may also be owned and operated by train manufacturers and railway suppliers or railway undertakings, all obviously with very different needs and perspectives.

In the following, we hence want to shortly elaborate on the motivation why different entities could own and operate their own Data Factory, and the different perspectives these may take.

Case of the train manufacturer or railway supplier

It is quite realistic that train manufacturers or railway suppliers will establish their own Data Factories (in some cases, this is already happening), as they need these to develop the AI-based products incl. models that they ultimately sell to their customers. As already mentioned in the introduction, train manufacturers and railway suppliers would likely not be able to collect enough sensor data for training purposes themselves, hence they will have the motivation to join the Pan-European Data Factory to gain access to significantly more sensor data. Also, it may be beneficial for them to use validation approaches that are standardised on a European level. At the same time, they may want

Interne

to use tool chains highly related to their own specific AI approach and to proprietary know-how, which they may not want to share with others, and particularly not with competitors.

Case of the railway undertakings

Many different scenarios are possible that lead to the situation that a railway undertaking owns its own Data Factory. One scenario involves validating and testing algorithms and AI models used in autonomous trains and working with certification organisations. In this respect, it is important to mention that the deployment of autonomous trains may vary from country to country. We assume that retrofitting the existing fleet to introduce GoA4 functions is considered technically and economically unfavourable. Therefore, the deployment of autonomous trains across Europe will probably be related to fleet renewal, and in consequence happen over a long period of time. If a level of automation of railway operation is chosen taking into account technical and human factors, then this will also depend on the local (regional) transport authority. Such decision may often be rather politically driven, and not only by rational technical and economic aspects. We further assume that there may be large differences in complexity in the technical implementation of GoA4 operations in train fleets, for example when comparing high-speed train fleets with regional or freight trains. It can therefore be assumed that different levels of automation from GoA1 to GoA4 will be used depending on the region and fleet type, and on political and economic arguments. In consequence, each railway undertaking may desire a specific extent of Data Factory, tailored to the very specific level of automation in rail operation it is aiming at, and the very specific requirements involved.

Case of the infrastructure manager

In the case of DB, it is currently the infrastructure manager DB Netz AG driving the development of a first larger Data Factory in Germany. It is doing so in the assumption that individual suppliers or railway undertakings would likely not be able to establish GoA4 development including the required Data Factory infrastructure all by themselves, and that a multi-tenancy and supplier-agnostic sector ecosystem has to be initiated to facilitate this. In this respect, the infrastructure manager (at least in countries where there is one main infrastructure manager) is in a good position to do so, as its own infrastructure in the form of data centres, connectivity backbone, etc., is typically already a good starting point for the infrastructure required for the notion if a “national” Data Factory. Also, the infrastructure manager sets the network access conditions for the railway undertakings and may anyway tend to use these to enforce a maximally harmonised approach to GoA4 within its rail network. If in a country the (main) infrastructure manager is the driving force behind the country’s main Data Factory, it can be expected that this will aim at a large degree of harmonisation across railway undertakings and suppliers.

In conclusion, it is unlikely to expect that there will be one uniform notion of Data Factory or even a single “national” Data Factory in one country. There may obviously be certain national specifics related to legal or regulatory constraints, but it is expected that there will still be a strong heterogeneity in the different Data Factories within one country, driven by the different perspectives and business case of the aforementioned types of potential owners, as highlighted before.

Interne

2.2 Hypotheses for the PEDF Deployment

The previous considerations (and further independent considerations) lead to the following hypotheses about the PEDF that provide the basis for the strategy outlined in the subsequent chapters:

Hypothesis 1: There is no “national” Data Factory, but many individual Data Factories

The Data Factories constituting the Pan-European Data Factories are likely not managed nationally, but can belong to different stakeholders, as elaborated in the previous section. This is the reason why the term "national” Data Factory", as often used in the discussion on this topic, is misleading and it’s better to speak about individual Data Factories.

Hypothesis 2: Each Data Factory as its own business model

Each individual Data Factory can and should have its own business model, and may encompass individual tools to meet the individual requirements of the owners and operators. However, it should be pointed out that the synergy effects within the framework of PEDF can leverage further business potential and, in the long term, increase cost efficiency in meeting individual requirements.

Hypothesis 3: The PEDF is a federation of individual Data Factories

Regardless of the legal representation, elaborated in Deliverable D 4.1 [11], the PEDF is an association of several entities owning and operating individual Data Factories. The respective benefits of each possible expansion stage of the PEDF are described in Chapter 0.

Hypothesis 4: There is no need for real-time data transfer between train and ground

This means that in the case of sensor data or other data with a large data size, no real-time data transmission is required, but instead the data is uploaded or downloaded during maintenance. This could for instance be based on the usage of the data Touch Points introduced in Deliverable D 1 [5], which are able to transfer, store and pre-process large amounts of data to and from the train over short distances. In the special case of cross-border or supra-regional train services, data (e.g., trained AI models) or software components that are specific to individual regions must also be transmitted at suitable locations. For this purpose, also Touch Points could be used, unless the amounts of data are so limited that synchronisation can happen en route, for instance through FRMCS.

Hypothesis 5: The use of the PEDF is not geographically limited to the location of its members

We assume that once a software component has been validated by the relevant European safety authorities for use on board a train on European territory, this component may be used on European trains. This applies regardless of whether or not there is an individual Data Factory in the European country in question. This applies in particular to non-safety-related software components or those that fulfil the necessary European specifications.

We further assume that special data or software components with or without safety reference may be required to cover certain local-specific functions in the train.

The lack of a local data provider, for example for the production of locally specific ML-based perception functions, may pose a problem with regard to the developability of such a function. Of course, this does not imply that a potential member cannot use the PEDF without an own Data Factory.

Hypothesis 6: No technical specification will come from the PEDF for the perception systems to the train manufacturers

The PEDF defines requirements with regard to data quality and data formats and can make some recommendations on the technical specifications of the various perception systems, but will not prescribe the latter specifications.

3 POSSIBLE DEPLOYMENT STRATEGY

Based on the findings and hypotheses in Chapter 2, the different speed of the expansion of GoA4 automated train driving and the local and political circumstances, there will be parallel activities with technical requirements among the potential members of the PEDF, to which the actual expansion and development of individual (national) Data Factories will be aligned. As the merger of the individual Data Factories into a PEDF requires coordination activities, it is first necessary to make a strategic distinction between the long-term development of the PEDF and the short and medium-term needs that the individual Data Factories serve. These short-term needs are, for example, (R&D) projects of the individual potential members that require individual Data Factories as a technical solution.

The long-term strategy for establishing the PEDF should therefore initially consist of carrying out the necessary coordination and standardisation in the areas of data quality, data formats, data and system interfaces, interconnection and connectivity via a pan-European high-speed backbone, as well as the common uniform toolchain. This long-term strategy also requires the creation of an organisation that is responsible for the governance of the topics described before, as already elaborated in D 4.1 [11]. This requires the creation of a legal entity or legal person that will form the PEDF and which will be governed by a certain European and/or national law. This organisation described here should be a European, independent and non-profit standardisation body for data, data interfaces and the PEDF and should also promote the standardisation of a common, uniform tool chain. For example, it could be modelled similar to ETSI, the European Telecommunications Standards Institute [12].

In the short term, this means that the individual national and possibly international users of a Data Factory must strive to find individual technical and legal solutions that meet the requirements of these (R&D) projects. From a technical perspective, this means that the users must develop or obtain their own solutions. From a legal perspective, this means that the respective contractual obligations must be mapped in the individual projects, for example as a network of contracts in which the individual rights and obligations of the stakeholders involved are mapped. This applies, for example, to individual cooperation agreements, regulations on intellectual property (IP) and data ownership, as well as regulations on liability, etc.

The medium-term strategy must therefore reflect the transition to the PEDF, which is to be seen as a merger or integration of the individual (national) Data Factories. This means that the ongoing individual activities and developments relating to the individual Data Factories must be aligned with the standardisation requirements of the PEDF and that the individual Data Factories must be adapted step by step so that they comply with the standardisation requirements of the PEDF.

Interne

In our opinion, a potential future member should have two ways of participating in the PEDF, namely by participating in the so-called **interface-pillar** or the **toolchain-pillar**. The interface-pillar focuses on the greatest possible freedom for future members and specifies interfaces for data exchange, in particular for sensor data, annotations and machine learning models. In addition to interface standardisation as in the interface-pillar, the toolchain-pillar in addition focuses on the creation of a common and uniform tool chain. Although this form of participation restricts the freedom in the development and expansion of individual Data Factories, it offers enormous synergy potential through complete interoperability of tools and data in the areas of data acquisition, data processing, data quality assurance, data transfer and access, as well as simulation and the training and evaluation of ML models.

Each strategy has its pros and cons and may gather a different number of members, as further elaborated in the following:

The mid- and long-term strategy within the interface-pillar

We have already noticed the divergences between companies working on obstacle detection systems for automated train operation (GoA4) on the architecture, the sensors used, the filtering of data, the setup of the sensors, the algorithms used, just to quote the most evident features. This indicates that we must be careful in the utility of the services we want to provide in the PEDF. This strategy is addressing this diversity by trying to relax the constraints on the members.

In this strategy, a great freedom is given to each member to use its own parameters (with the term “parameters” to be understood in a broad sense).

This strategy limits the discussion to the way to make the exchange possible at all the possible levels:

- data format;
- data organisation;
- data quality;
- annotation quality;
- model organisation;
- anonymisation and conformity to the privacy rules;
- sensor description;
- acquisition rules (triggering, frequency, etc.).

This discussion will give the guarantee that the exchange of data among the PEDF members is well facilitated.

The mid- and long-term strategy of the toolchain-pillar

Here, the idea is to harmonise the complete tool chain which gives an intrinsic versatility to the PEDF. The specifications in this strategy are more complete regarding not only the data but also the tools and the interfaces between the tools. The drawback is its rigidity and not necessarily adaptability to some specific Data Factory configuration.

Interne

4 RISKS AND THEIR POSSIBLE MITIGATION

In Table 1, risks related to the aforementioned strategic considerations are listed, together with possible mitigation measures.

Table 1. Risks related to the PEDF deployment incl. possible mitigation measures.

Risk	Description	Impact	Possible mitigation measures
Lack of exchange of data	Activities between the DFs are limited because of lack of interest in the data of the other members. It could be because of the differences in the sensors used or an absence of diversity between the data coming from different DFs	Minor impact If each Data Factory is supported by its own business model.	Discussion between members to get more complementarity data which are fruitful to share between members.
Network failure	The bandwidth between individual DFs collapses and cannot provide the desired performance needed by the members.	Minor impact at short term, as the PEDF is not directly used for real-time rail operation, but in the longer a term this is a potential showstopper for the PEDF.	Early invest in the pan-European backbone network, already in the phase where individual DFs are operating rather independently.
Loss of member	A member having a significant role in the PEDF is leaving the PEDF.	Minor impact if appropriate measures are taken in advance (see right column). Maybe an impact on available data or compute capabilities.	The member may leave all its right on the existing data, models, etc. the usage rights to the data have been legally clarified in advance when the member joins PEDF or contributes the data to PEDF in the event of the member leaving. The computing services provided may be taken over by another member etc.

Technical evolution of a member providing data	A member is modifying its technical specification and the generated data is no longer compatible with that of the other members	Major impact: Could be a problem if not anticipated by members	The strategy of each member must be tuned correctly, and it is up to the members to agree with the others to avoid dependency toward data provider.
Modification of data offer or flow	A member is modifying its data offer and modifying the data flow (for instance offering less or more data or compute resources)	Minor impact: Taking apart the modification in the member-to-member relation, it may impact the needs in terms of bandwidth in the network supporting the PEDF	Each member may consider that a modification in the offer or demand has an impact on the network and that providing a service which could not be accessed by the others may not be a good modification. It has the implication that anticipation in the modification may come with anticipation in the network modification

5 POSSIBLE PHASED DEPLOYMENT PLAN

This section outlines a possible phased deployment plan for the PEDF, a key initiative aligned with the European Union's goals for digital transformation and innovation in the railway sector.

Phase 1: Initial Analysis and Framework Development

- **Needs Assessment**: An exhaustive assessment to understand existing data management practices within the European railway sector, identifying key areas for improvement;
- **Framework Formulation**: Development of an overarching framework, addressing technical, operational, and regulatory dimensions, ensuring compliance with EU standards and regulations.

Phase 2: Pilot Implementation and Iterative Development

- **Pilot Projects**: Execution of targeted pilot projects across selected regions to evaluate practical viability and operational impact;
- **Feedback Integration**: Incorporation of feedback from pilot phases to refine the system, ensuring alignment with user needs and operational realities.

Interne

Phase 3: Infrastructure Development and Standardisation

- **Network Infrastructure**: Establishment of a high-speed backbone network, crucial for seamless data transfer and compliance with EU digital infrastructure goals;
- **Data Centers**: Deployment of strategically located data centres, both centralised and decentralised, equipped to handle large-scale data processing.

Phase 4: Stakeholder Engagement and Collaborative Frameworks

- **Inclusive Involvement**: Active participation of diverse stakeholders, including railway operators, technology providers, and policy makers, fostering a collaborative approach in line with EU principles of stakeholder engagement;
- **Harmonisation Efforts**: Efforts to harmonise operational procedures, ensuring interoperability and standardization across member states.

Phase 5: Training, Capacity Building, and Compliance

- **Skill Development**: Comprehensive training initiatives to enhance user competencies and familiarity with the Data Factory system;
- **Regulatory Adherence**: Ensuring strict adherence to EU regulations, particularly the General Data Protection Regulation (GDPR), reinforcing the commitment to data protection and privacy.

Phase 6: Gradual Rollout and Scalable Expansion

- **Controlled Deployment**: A phased rollout strategy to manage scalability and ensure a controlled expansion of the Data Factory's capabilities;
- **Adaptive System Design**: Designing the system to be inherently adaptable, allowing for future technological integrations and sectoral evolutions.

Phase 7: Performance Monitoring and Continuous Enhancement

- **Impact Assessment**: Ongoing monitoring of the system's performance against predefined KPIs, ensuring alignment with EU objectives for digital innovation in transportation;
- **Dynamic Evolution**: Commitment to continuous improvement, reflecting the dynamic nature of technological advancements and sector-specific needs.

The proposed deployment strategy for the Pan-European Data Factory is designed to align with the European Union's strategic objectives of enhancing digital infrastructure, fostering innovation, and improving operational efficiency in the railway sector. This multi-phased approach ensures a meticulous and inclusive process, paving the way for a transformative impact on Europe's railway systems.

6 CONCLUSIONS

The document has presented a possible deployment strategy for the Pan-European Data Factory (PEDF), outlining assumptions, hypotheses, and a strategic plan. It critically evaluates the notion of

Interne

"national" Data Factories, acknowledging the diverse perspectives and interests of potential owners such as train manufacturers, railway suppliers, railway undertakings, and infrastructure managers.

One key conclusion drawn is the hypothesis that there won't be a uniform notion of a "national" Data Factory; instead, multiple individual Data Factories will exist, each driven by different stakeholders with distinct business cases. This realisation emphasises the necessity for a pan-European Data Factory that accommodates this heterogeneity while facilitating collaboration and standardisation across various entities.

It is further concluded that there are likely two main deployment strategies or two main pillars, an interface-pillar and a toolchain-pillar. Identified risks appear to be mitigatable if certain measures are taken.

Ultimately, a phased deployment strategy is proposed, encompassing the phases:

- Initial analysis and framework development;
- Pilot implementation and iterative development;
- Infrastructure development and standardisation;
- Stakeholder engagement and collaborative frameworks;
- Training, capacity building, and compliance;
- Gradual rollout and scalable expansion;
- Performance monitoring and continuous enhancement.

Each phase is meticulously designed to address specific needs, including technical, operational, regulatory, and collaborative aspects. Emphasis is placed on stakeholder engagement, harmonization, skill development, regulatory adherence, controlled deployment, and continuous improvement.

The strategy acknowledges potential risks, such as data exchange challenges, network failure, member departure, and technical evolutions affecting data compatibility. It proposes solutions like fostering discussions among members, anticipating technical modifications, and ensuring legal agreements to mitigate potential impacts.

Overall, the proposed deployment strategy aligns with the European Union's objectives for digital transformation, innovation, and improved efficiency in the railway sector. The phased approach ensures a comprehensive and inclusive process, setting the stage for a transformative impact on Europe's railway systems through the establishment of a Pan-European Data Factory

REFERENCES

[1] Shift2Rail program, see https://rail-research.europa.eu/about-shift2rail/

[2] Europe's Rail program, see https://projects.rail-research.europa.eu/

[3] Sensors4Rail project, see "Sensors4Rail tests sensor-based perception systems in rail operations for the first time," Digitale Schiene Deutschland, 2021. [Online]. Available: https://digitale-schiene-deutschland.de/en/Sensors4Rail

[4] DIRECTIVE (EU) 2016/797 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, see https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016L0797

[5] CEF2 RailDataFactory Deliverable 1, "Data Factory Concept, Use Cases and Requirements", Version 1.1, May 2023. [Online]. Available: https://digitale-schiene-deutschland.de/Downloads/2023-04-24_RailDataFactory_CEFII_Deliverable1_published.pdf

[6] Shift2Rail TAURO project, Horizon 2020 GA 101014984, see https://projects.shift2rail.org/s2r_ipx_n.aspx?p=tauro

[7] P. Neumaier, "First freely available multi-sensor data set for machine learning for the development of fully automated driving: OSDaR23", 2023. [Online]. Available: https://digitale-schiene-deutschland.de/en/news/OSDaR23-multi-sensor-data-set-for-machine-learning

[8] Open Sensor Data for Rail 2023, 2023. [Online]. Available: https://data.fid-move.de/dataset/osdar23

[9] R2DATO project, see https://projects.rail-research.europa.eu/eurail-fp2/

[10] P. Neumaier, "Data Factory - "Data Production" for the training of AI software," Digitale Schiene Deutschland, 2022. [Online]. Available: https://digitale-schiene-deutschland.de/news/en/Data-Factory

[11] CEF2 RailDataFactory, D 4.1 – "Deployment activities description for a pan-European Railway Data Factory", 2023. [Online]. Available: https://digitale-schiene-deutschland.de/news/en/Data-Factory

[12] ETSI, see https://www.etsi.org/