CEF2 Rail Data Factory

D3.3 – Description of cybersecurity vulnerabilities, threat scenarios and usable standards to mitigate associated risks

Due date of deliverable: 31/08/2023

Actual submission date: 28/11/2023

Leader/Responsible of this Deliverable: Bart du Chatinier (WP 3 lead)

Reviewed: **Y**/N

Document status		
Revision	Date	Description
01	09/03/2023	Document template generated
02	25/09/2023	First version, ready for review
03	04/10/2023	Version submitted to advisory board for review
04	3/11/2023	All review comments processed
05	28/11/2023	Version submitted to 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	Sensitive – limited under the conditions of the Grant Agreement	

Start date: 01/01/2023

Duration: 9 months
(note: amendment request for project extension ongoing)

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
Bart du Chatinier	NS
Jan van Gelder	NS
Gertjan Tamis	NS
Vanessa Fong Tin Joen - Baarh	NS
Alexander Heine	DB
Philipp Neumaier	DB
Patrick Marsh (only editorial)	DB

Note of Thanks

We would like to thank our Advisory Board Members for the valuable discussion and their detailed comments on 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

EXECUTIVE SUMMARY

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

In particular in the pursuit of fully automated driving (so-called Grade of Automation 4, GoA4), where sensors and cameras on trains will be used to automatically react to hazards in rail operation, it is commonly understood that an individual railway company or railway vendor would not be able to collect enough sensor data to sufficiently train the artificial intelligence (AI) eventually deployed in the rail system. For this reason, it is commonly assumed that a form of pan-European Rail Data Factory is needed, as an infrastructure and ecosystem that allows various railway players and suppliers to collect and process sensor data, perform simulations, develop AI models, certify models, and ultimately deploy the models in the automated railway system.

This deliverable emphasizes the necessity of using universally accepted standards and agreements regarding data ownership, definitions, and formats among all involved parties. Also mandatory is the need for comprehensive investigation to determine the appropriate information, data, and data exchange standards for effective pan-European interchange. The data factory initiative addresses considerations of data application, benefits and challenges, data security threats and risks associated with sophisticated models, and relevant cybersecurity vulnerabilities in the European train system. It implements a robust and effective data risk management, classification, tagging and labeling, and metadata strategies to process the vast amount of sensor data. Finally, the deliverable emphasises leveraging usable European standards and regulations to mitigate identified risks and enhance the security and efficiency of data exchange within the European railway system.

ABBREVIATIONS AND ACRONYMS

Abbreviation	Definition
AI	Artificial Intelligence
ATO	Automatic Train Operation
CRS	Coordinate Reference System
DDoS	Distributed Denial of Service
EPSG	European Petroleum Survey
FAIR	Findable, Accessible, Interoperable, Reusable
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
GoA4	Grade of Automation 4
HADEA	European Health and Digital Executive Agency
HPC	High-performance computing
GDPR	General Data Protection Regulation
IM	Infrastructure Manager
IMU	Inertial Measuring Unit
ML	Machine Learning
MVB	Multipurpose Vehicle Bus
NCSC	National Cyber Security Center
NS	Nederlandse Spoorwegen
OEM	Original Equipment Manufacturer
RU	Railway Undertaking
RD	Rijks Driehoekstelsel
SOC	Security Operations Center
TCMS	Train Control Management System
VLAN	Virtual Local Area Network

TABLE OF CONTENTS

Acknowledgements 2
Report Contributors 2
Executive Summary 3
Abbreviations and Acronyms 4
Table of Contents 5
List of Figures 6
List of Tables 6
1 Introduction 6
1.1 Aim and Scope of the CEF2 Rail Data Factory Study 7
1.2 Delineation from and Relation to other Works 7
1.3 Aim and Structure of this Deliverable 7
2 Considerations on Data Applications in Train Operation 8
2.1 Challenges related to Data Applications 10
2.2 Data Security Threats in the Pan-European Data Factory 12
2.3 STRIDE-based Cybersecurity Risk Analysis of the Data Factory 13
2.4 A Bowtie Risk Analysis of the Data Factory 18
3 Data Risk Management 21
3.1 Sensor Data is Critical High Volume Data 21
3.2 Data Classification 22
3.3 Data Annotation 23
3.4 Metadata 23
3.5 Difference between Metadata and Data Tagging 24
3.6 Different Data Types 25
4 Usable Standards to mitigate Identified Risks 25
4.1 European Regulations 25
4.2 European Standards 26
5 Conclusions 27
6 References 28

LIST OF FIGURES

Figure 1. Data quality dimensions. 9
Figure 2. Sensor setup example. 10
Figure 3. Examples for annotated sensor data [10]. 11
Figure 4. Data flow from trains to and within the Rail Data Factory. 15
Figure 5. Bowtie Risk Model [17]. 20

LIST OF TABLES

Table 1. STRIDE approach [16]. 14
Table 2. Application of STRIDE to the data flow from trains to and within the Data Factory. 15
Table 3. Exemplary Bowtie Risk Table created for the Data Factory 20

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 artificial intelligence (AI) training and to be able to prove that the trained AI meets the safety and security 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, a 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.

1.1 Aim and Scope of the CEF2 Rail Data Factory Study

The CEF2 Rail Data Factory study focuses exactly on aforementioned vision of a pan-European Rail Data Factory for the joint development of fully automated driving. The study, being co-funded through HADEA, aims to assess the feasibility of a pan-European Rail Data Factory from technical, economical, legal, regulatory, and operational perspectives, and determine key aspects that are required to make a pan-European Rail Data Factory a success. For a better understanding of the study's aim and scope, please see Chapter 1.1 in Deliverable D1 [4].

1.2 Delineation from and Relation to other Works

The notion of Rail Data Factory is also covered in other work, such as the Shift2Rail project TAURO [5] or Europe's Rail Innovation Pillar FP2 R2DATO project [6]. Further, Deutsche Bahn, within the sector initiative "Digitale Schiene Deutschland", has already started setting up some related data center components [7]. For a better understanding of the relationship between the CEF2 RailDataFactory study and these works, please see Chapter 1.2 in Deliverable D1 [4].

1.3 Aim and Structure of this Deliverable

This current document is the deliverable D3.3 of the CEF2 RailDataFactory project, covering the risks and benefits when account cybersecurity and data governance in a pan-European Data Factory Backbone Network. The objectives in WP3 will fit into the broader research question how the rail industry can apply higher levels of automation of rolling stock within an existing rail network.

The remainder of this document is structured as follows:

- In Chapter 2, considerations, improvements, benefits, threats, and risks associated with data applications in train operation and in the pan-European Rail Data Factory are listed. Further, security risk assessments of the Rail Data Factory are applied;
- In Chapter 3, data risk management is covered;
- In Chapter 4, (European) standards and regulations are listed that may help to mitigate and address the aforementioned challenges;
- In Chapter 5, conclusions are provided.

2 CONSIDERATIONS ON DATA APPLICATIONS IN TRAIN OPERATION

In the following, the term **"Data Application"** refers to a software application or system that is designed to handle, process, and utilize data for specific purposes. These applications are developed to collect, store, analyses, and present data in a meaningful way, often to support decision-making, provide insights, automate processes, or deliver numerous services. Data applications in trains, particularly those that utilize information technology and data analytics, can bring numerous benefits, including improved safety, efficiency, and passenger experience. Nevertheless, they bring about specific risks and challenges. The potential risks linked to applying data in train operations are:

Cybersecurity Vulnerabilities: As data applications in trains become more interconnected and reliant on the Internet and communication networks, they become vulnerable to cyber threats and thus represent the first point of attack for compromising the data within pan-European Data Factory. Hackers may attempt to gain unauthorized access to train systems, leading to potential disruptions, safety risks, or even remote control of train operations. Ensuring robust cybersecurity measures, such as encryption, firewalls, and regular security audits, is crucial to safeguarding these applications and usage of data inside in a pan-European Data Factory setup (see deliverable D2.2 [8] for more details).

The infrastructure is also vulnerable to a specific type of manipulation called "data manipulation". This is an attack that tries to modify the input data of a machine learning (ML) model, so that the model produces a desired output. Adversarial attacks are an example for deceiving or manipulating an artificial intelligence (AI) system by exploiting its weaknesses or limitations. An attacker may modify the infrastructure in an imperceptible way to humans, e.g. adding a certain noise pattern that causes the AI system to misclassify the recorded image. Such attacks can have serious consequences for AI systems used for security or critical decision-making. In distinction to the vulnerabilities of the infrastructure, there is also “data poisoning”, an attack that tries to corrupt the training of ML models so that the model will produce wrong or biased predictions.

Data Privacy Concerns: Data applications in trains often collect and process sensitive information, such as passenger data, travel patterns, and payment details. With additional, highly accurate sensors and cameras at the train front, personally identifiable data and the recording of travel patterns are of concern. Mishandling or unauthorized access to this data can lead to privacy breaches and identity theft. Data protection regulations, such as the General Data Protection Regulation (GDPR) [9] in Europe, must be adhered to strictly to protect passengers' personal information.

Reliability and Redundancy Issues: Data applications are reliant on stable and resilient communication networks. Any disruptions or outages in connectivity could impact the functioning of these applications, affecting train operations, passenger services, and safety. Implementing redundant systems and backup plans is essential to minimize downtime.

Data Accuracy and Quality: Data-driven decisions and actions are only as good as the quality of the data being used. Inaccurate or incomplete data can lead to incorrect insights, potentially resulting in operational inefficiencies or safety risks. Continuous data validation and maintenance procedures are necessary to ensure the accuracy and reliability of the data. Figure 1 shows the six dimensions of data quality considered in NS.

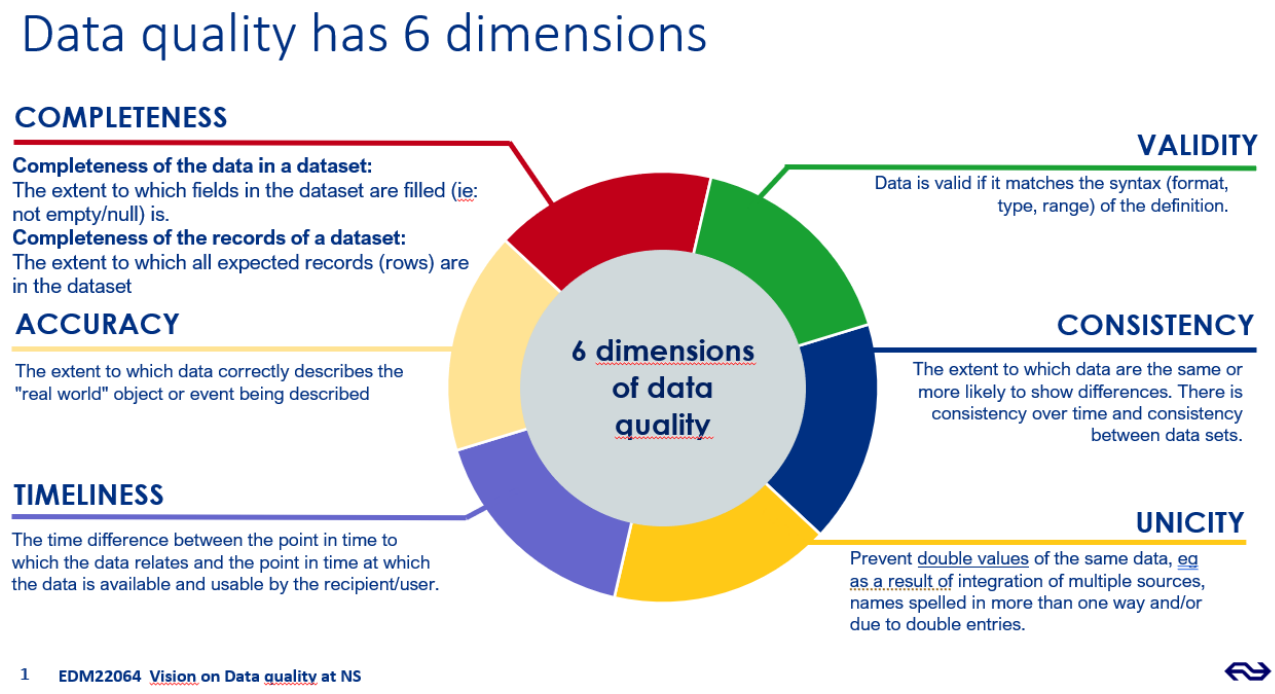

Figure 1. Data quality dimensions.

Dependency on Technology: Increasing reliance on data applications means a higher dependency on technology. Technical failures, software bugs, security issues, or hardware malfunctions can disrupt train operations and require quick resolution to minimize the impact on passengers and safety.

As a consequence, pending solutions to solve technical failures could lead to trains unsuitable for service or decreased lifecycle when they cannot be solved.

Regulatory Compliance: Data applications in trains must comply with various industry regulations and standards to ensure safety and interoperability. Failure to meet these requirements can lead to legal issues, fines, or operational limitations.

Copy Protection and Copy Control: Copy protection in data security is the process of protecting files and folders from being copied without proper authorization to any device in the same network. Unauthorized copying of data can lead to data leak, exposure, or even a breach. File copy protection ensures the safety of data at rest and in use. It should be noted that copy protection of intellectual property or assets is another component of data security that must not be confused with copy protection of business-critical data in general.

To mitigate these risks, train operators, technology providers, and regulators must collaborate to develop comprehensive risk management strategies, adopt industry best practices, and stay up-to-date with the latest advancements in cybersecurity and data protection measures. More on this topic can be found in RailDataFactory Deliverable D1 [4] and Deliverable D2.2 [8].

2.1 CHALLENGES RELATED TO DATA APPLICATIONS

The Data Factory shall allow present and future objects of the real world, including their properties, relationships, and behavior, to be mapped and simulated, in the form of an open data model that is accessible and usable to a broad pan-European array of stakeholders from railway infrastructure managers to railway undertakings, vendors, safety and security authorities, and academic partners.

Examples of annotated sensor data

For illustration purposes, Figure 2 shows an NS SNG train equipped with additional sensors, and Figure 3 shows examples of annotated sensor data.

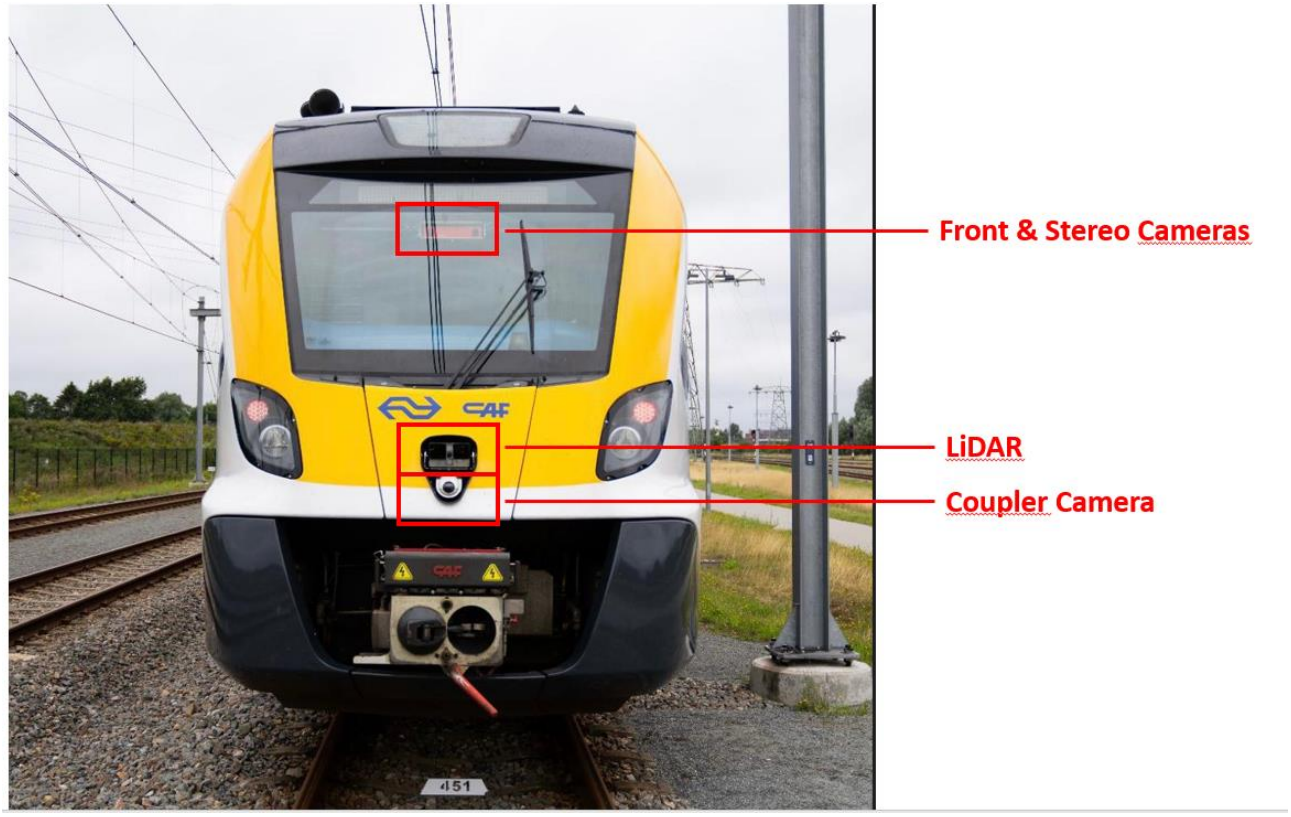

Figure 2. Sensor setup example.

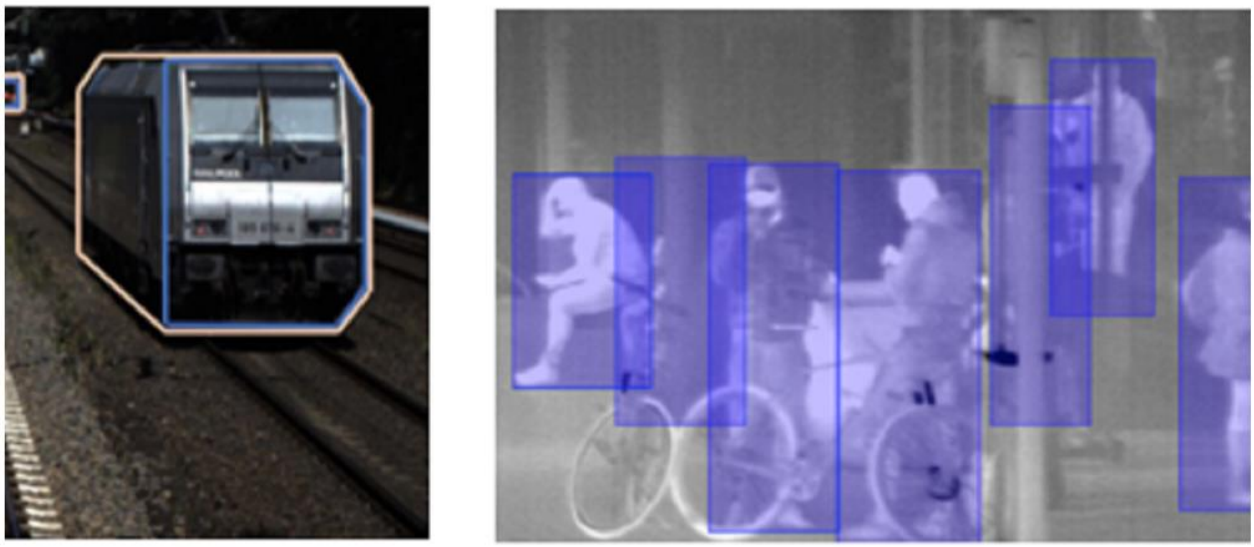

Figure 3. Examples for annotated sensor data [10].

The following three challenges were identified and are discussed within this deliverable:

Integration Challenges: Trains often have complex and diverse systems, and integrating data applications into existing infrastructure can be challenging. Compatibility issues and interoperability concerns may arise when trying to connect new data applications with legacy systems.

Since its beginning in early 2018, the GO FAIR community has been working towards implementations of the FAIR Guiding Principles [11]. This collective effort has resulted in a three-point framework that formulates the essential steps towards the goal of a global Internet of FAIR Data and Services where data are **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**) for machines. Especially interesting for the data factory is the 'I' – Interoperable [11][12].

Data usually needs to be integrated with other data. In addition, the data needs to interoperate with applications or workflows for analysis, machine learning, storage, and processing.

Data in transit or data in motion includes all data that is shared or transmitted within any network or outside through the Internet. A few examples include files shared with coworkers, data uploaded to cloud applications, and data sent to business associates. Data in transit is most vulnerable as it gets exposed to high security threats like eavesdropping attacks, ransomware attacks, and data theft. Additionally random corruption of data can happen in transit which requires additional precautions to be taken.

Currently, there is no mandatory EU industry standard for sensor and diagnostic data. Thus, each manufacturer uses their own data and data exchange formats. This makes it even more complex to get data on and off the train. For a pan-European data interchange to work, information, data and data exchange standards are needed. Which standards to use requires further investigation.

A good start would be an EU glossary of objects that can be used to create a conceptual information model. Next step is a logical information model. This logical information model should be the base for the data exchange model.

Costs and Return on Investment: Implementing and maintaining data applications plus data transport can involve significant costs. Ensuring that the benefits and returns on investment justify these expenses is crucial for the long-term viability of these projects [13].

2.2 DATA SECURITY THREATS IN THE PAN-EUROPEAN DATA FACTORY

The previous paragraph gave an outline on risks and challenges when applying data communication on trains. This paragraph gives a short impression of known threats for operators and/or infrastructure managers that are applicable for rolling stock participating in a pan-European Data Factory Backbone Network.

1. **Data manipulation**: Manipulated object detection models for Automatic Train Operation (ATO) could fail to identify obstacles or detect nonexistent objects, leading to a heightened risk of collisions with other trains, vehicles, pedestrians, or objects on the tracks.
2. **Adversarial attacks**: Adversarial attacks in AI attempt to deceive or manipulate an AI system by exploiting its weaknesses or limitations. For example, an attacker may add a small amount of noise or distortion to an image that is imperceptible to humans but causes the AI system to misclassify the image. Such attacks can have serious consequences for AI systems used for security or critical decision-making. Therefore, it is important to develop robust and secure AI algorithms (see "Secure architecture Design" in the next section) and to ensure cybersecurity.
3. **System Malfunction and Failure**: Malicious data can lead to software bugs, crashes, or system malfunctions, resulting in temporary or prolonged service outages and delays. Data and Privacy Breach: Cybersecurity breaches can lead to unauthorized access to sensitive data, intellectual property, and operational data. This may result in privacy violations and data theft or ransom.
4. **Financial Losses**: Cybersecurity breaches can lead to costly disruptions and system downtime.
5. **Reputation Damage**: Public trust in rail services may be severely impacted if cybersecurity breaches result in breach of privacy and service interruptions.
6. **Supply Chain Disruptions**: Attacks on rail infrastructure or third-party vendors can disrupt the supply chain, leading to unusable sensor data.

These threats above can be caused by many types of vulnerabilities. It is important for the rail industry to have up-to-date policies to minimise the risk of trains or connected infrastructure being attacked. Multiple useful references are available how this can be done, such as [14]. It's important to note that not only software related attacks (such as DDoS and ransomware) but also physical attacks are to be considered. In modern days, trains are moveable data centers, and an undetected intruder can access and tamper train systems directly on the asset itself. To minimize the potential impacts of cybersecurity breaches, the rail industry should implement robust cybersecurity measures. Valuable recommendations are listed in [15], but in summary a short synopsis of mitigation strategies is written below.

- **Secure architecture Design:** Train systems and the Data TouchPoint should be designed with security in mind, incorporating encryption, secure authentication, and access controls. It is recommended to consider cybersecurity design in the tender for new rolling stock. Additionally, a periodic review with the supplier or pen testing can show how robust the train is for recent developments within the cybersecurity industry.
- **Regular software updates:** It's the responsibility of manufacturer and operator to keep train and Data TouchPoint software and systems up to date with the latest security patches to address known vulnerabilities. A documented configuration management system where

both software and hardware components are registered, including a management of change process, will help ensuring the latest train configuration and to detect possible deviations.

- **Network segmentation:** Together with the manufacturer of the train and the supplier of the component of the Data TouchPoint, it is important to segregate critical control systems from non-essential networks to limit the spread of cyber threats. By using VLANs and firewalls, the architecture can help reducing the probability of cyberattacks via the network or via physical ports in the trains or on the interfaces of the TouchPoint.
- **Intrusion detection and prevention:** When technically possible, it is recommended to implement intrusion detection and prevention systems to detect and block suspicious activities. This capability is dependent on company policies and active monitoring, such as the presence of a Security Operating Centre (SOC) with Security Information and Event Management.
- **Security testing:** The importance of conducting thorough security testing, including penetration testing, to identify and address potential vulnerabilities on a yearly basis cannot be stressed enough. The rail industry and the infrastructure managers should assume a constant development of new breaches and vulnerabilities, forcing active risk management on trains and infrastructure components. One could also consider regular security audits by external parties.
- **Employee training:** For both manufacturers and operators, employees should be trained to recognise and report potential security threats. This also requires the existence of security protocols and allowing anyone working with trains to easily notify suspicions or threats.
- **Collaboration and information sharing:** Finally, the collaboration with industry stakeholders, other operators and EU working groups are a cornerstone for sharing information on emerging threats and to share best practices.

2.3 STRIDE-BASED CYBERSECURITY RISK ANALYSIS OF THE DATA FACTORY

Since a lot of the data within the Rail Data Factory is collected from various onboard systems and sensors, it is important that this data can be considered reliable and trustworthy. This has been described extensively in deliverable D2.2 [8]. Mitigating measures, such as encryption and digital signatures, protect data from tampering and unauthorized modifications during transmission from trains to the Rail Data Factory. This ensures that the data retains its integrity and authenticity, enabling accurate analysis and decision-making. Additionally, improved protocols provide a secure environment for transferring data from trains over the Data TouchPoint to the Rail Data Factory. Preventing unauthorized access and eavesdropping, maintaining the confidentiality of sensitive data. In addition, it helps with the compliance with data privacy regulations, such as the GDPR [9] in Europe.

By implementing best practices to safeguard data integrity, protect against cyber threats, and ensure data privacy, the rail industry can rely on accurate, high-quality data for informed decision-making, improved operational efficiency, and enhanced safety standards. One of these frameworks that can be used is STRIDE [16]. This is a mnemonic for a set of threats – Spoofing, Tampering, Repudiation, Information disclosure, DDoS and Elevation of privilege as described in Table 1.

Table 1. STRIDE approach [16].

	Type of Threat	What Was Violated?	How Was It Voilated?
S	Spoofing	Authentication	Impersonating something or someone known and trusted
T	Tampering	Integrity	Modifying data on disk, memory, network, etc.
R	Repudiation	Non-repudiation	Claiming not to be responsible for an action
I	Information Disclosure	Confidentiality	Providing information to someone who is not authorized
D	Denial of Service	Availability	Denying or obstructing access to resources required to provide service
E	Elevation of Privilege	Authorization	Allowing access to someone without proper authorization

The application of STRIDE on the entire flow from trains to and within the Rail Data Factory shows the usefulness of the framework. Figure 4 gives an extended view of the data chain for risk assessment purposes. The letters and arrows are indicating the possible points of attack. The data chain starts with the sensors on board of the train (B) creating large data files from the supplier software (A). The data is offloaded through a wireless connection (C) to Touch Points (D). At the Touch Points, data computing takes place and the reduced data files are sent through landlines (F) to HPC & Storage (G). Data may also be artificially generated through simulation (I). HPC & storage sites are interconnected (H). Software suppliers create and deliver software for the sensors (A) and the Touch Points (E).

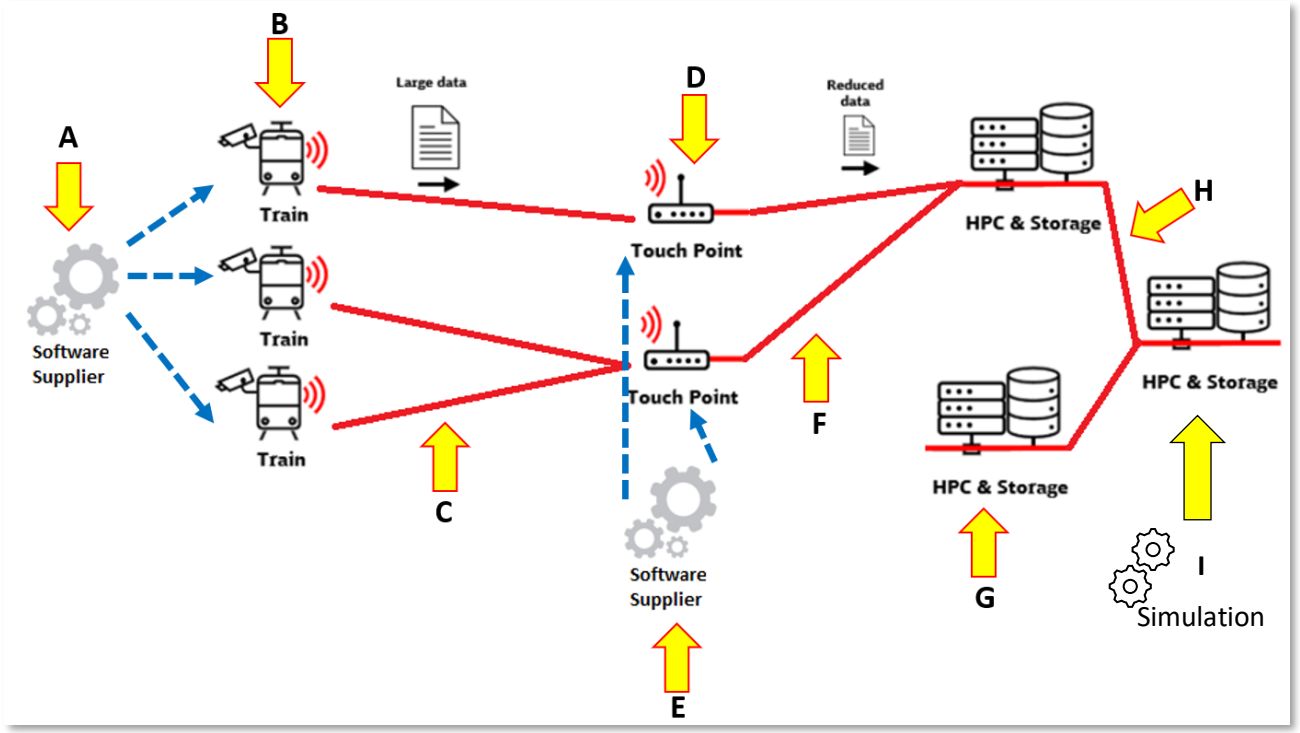

Figure 4. Data flow from trains to and within the Rail Data Factory.

In Table 2, the risks per element are described using the aforementioned STRIDE method. For each part, the unwanted event with some clarification and the level of probability and impact are set out.

This is a high-level approach; in a later stadium a risk analysis based on IEC 62443-3-3 should be executed on at least the software supplier (A) and Train (sensors) (B) and, probably as well on the Touch Point (D) and Touch Point Software Supplier (E).

Table 2. Application of STRIDE to the data flow from trains to and within the Data Factory.

Location	STRIDE	Unwanted event	Clarification	Probability (L-M-H)	Impact (L-M-H)
B	Tampering	Perception data is poisoned on the train, thereby making data sent to the HPC & Storage unusable.	Software on the train is manipulated by an attacker.	L	L
A	Tampering	An attacker performs a supply chain attack, thereby poisoning all train series.	Software delivered by the software supplier is manipulated, resulting in unusable data from the train series to the HPC & Storage.	L	M
A	Tampering	The supplier delivers a defective software update, resulting in malfunctioning perception systems.	Software delivered by the software supplier is manipulated, resulting in unusable data from the train series to the HPC & Storage.	L	M

Location	STRIDE	Unwanted event	Clarification	Probability (L-M-H)	Impact (L-M-H)
E	Tampering	Perception data is poisoned on the Touch Point, thereby making data sent to the HPC & Storage unusable.	Software on the Touch Point is manipulated by an attacker, resulting in unusable data	L	M
G	Tampering	Perception data is poisoned on the HPC & storage, thereby making data sent to the HPC & Storage unusable.	Software on the HPC & storage is manipulated by an attacker, resulting in unusable data	L	H
D	Spoofing	A system or person impersonates itself as another train, thereby injecting false data to the HPC & storage.	Though this threat, the use of misleading images could be shared with the HPC & storage.	M	L
D	Spoofing	A system or person impersonates itself as another Touch Point, and thereby injecting false data to the HPC & storage.	Though this threat, the use of misleading images could be shared with the HPC & storage.	L	M
H	Spoofing	A system or person impersonates itself as another HPC & storage, and thereby injecting false data to other HPC & storages	Though this threat, the use of misleading images could be shared with the HPC & storage.	L	H
I	Spoofing & Tampering in Simulation Toolchain	A system or person impersonates itself as an entity with elevated rights and tampers unauthorized with the data in the HPC & storages	Data is the HPC & Storage is tampered	L	H
BCDGH	Repudiation	Due to the lack of cyber security logging in the train / Touch Point / HPC & Storage, it is not clear what happened to the equipment in the train during a cyber security incident.	If there is no logging, it is difficult to reconstruct what happened during a cyber security incident.	L	M
BCDGH	Repudiation	Due to the lack of cyber security	If there is no logging, it is difficult to reconstruct what	M	M

Location	STRIDE	Unwanted event	Clarification	Probability (L-M-H)	Impact (L-M-H)
		logging in the touchpoint, it is not clear what happened to the equipment in the touchpoint during a cyber security incident.	happened during a cyber security incident.		
BCDGH	Repudiation	Due to the lack of cyber security logging in the HPC & Storage, it is not clear what happened to the equipment in the HPC & Storage during a cyber security incident.	If there is no logging, it is difficult to reconstruct what happened during a cyber security incident.	H	H
BCDGH	Information disclosure	The data sent by the train / Touch Point / HPC & Storage for ATO / TCMS is viewed by unauthorized persons or computers.	This will influence the image of the stakeholder and probably the privacy of the data subjects (i.e. camera images recorded by the cameras on the train),	L	L
BCDGH	Information disclosure	The data sent by the touchpoint for ATO / TCMS is viewed by unauthorized persons or computers.	This will influence the image of the RUs and probably the privacy of the data subjects (i.e. camera images recorded by the cameras on the train),	L	L
BCDGH	Information disclosure	The data sent by the HPC & Storage for ATO / TCMS is viewed by unauthorized persons or computers.	This will influence the image of the RUs and probably the privacy of the data subjects (i.e. camera images recorded by the cameras on the train),	L	H
D	Escalation of privileges	An attacker gains admin access to perception systems on the train and manipulates the data on the perception system.	Touch Point data gets manipulated.	L	L
D	Escalation of privileges	An attacker gains admin access to a Touch Point and manipulates the data on the perception system.	HPC & storage data gets manipulated.	L	M

Location	STRIDE	Unwanted event	Clarification	Probability (L-M-H)	Impact (L-M-H)
G	Escalation of privileges	An attacker gains admin access to the HPC & storage and manipulates the data on other HPC & storages.	HPC & storage data gets manipulated.	L	H
C	Denial of Service	The connection between the train and the Touch Point is being disrupted for example by jamming the signal, power outage, or denial of service on the Touch Point itself.	No sensor data from the train can be received by the touchpoint and sent to the HPC & Storage	L	M
F	Denial of Service	The connection between the Touch Point and the HPC & storage is being disrupted for example by jamming the signal, power outage, or denial of service on the Touch Point itself.	No reduced data can be sent from the Touch Point to the HPC & Storage	L	M
H	Denial of Service	The connection between the HPC storage and the HPC & storage is being disrupted for example by jamming the signal, power outage, or denial of service on the HPC & Storage itself.	No data can be shared between different HPC & Storage locations	L	M

2.4 A BOWTIE RISK ANALYSIS OF THE DATA FACTORY

This paragraph will apply two of vulnerabilities described above in the Bowtie Risk Model [17]. It will show how this model can be applied to translate vulnerabilities into threat scenarios. The first vulnerability is the lack of network traffic encryption in rolling stock. Modern trains utilise various communication protocols to transfer data between onboard systems and wayside infrastructure, such as Multipurpose Vehicle Bus (MVB) or Ethernet protocols. Unfortunately, these communication protocols lack built-in encryption mechanisms. As a result, sensitive data transmitted over the network, such as control commands, train status information, and passenger data, are susceptible to interception and tampering by malicious actors.

In the area of cybersecurity, it is essential to maintain the basic objectives of availability, confidentiality and integrity. For the rail sector, it is crucial to complement these objectives with authenticity, accountability, non-repudiation and data protection, as stated in the past section.

The collection, storage, transmission, processing and presentation of data is associated with specific risks and potential impacts, which we will explain with examples for better understanding:

- **Unauthorised Access**: Hackers could gain unauthorised access to the train's systems, potentially controlling train subsystems and possibly even operational activities, leading to safety hazards and operational disruptions. This is dependent on mitigating measures taken by RUs and OEMs.
- **Data Interception**: Sensitive information exchanged between trains and wayside systems, including passenger data and maintenance records, can be intercepted, leading to privacy breaches and data theft.
- **Data Manipulation**: Unencrypted data can be altered by malicious actors, leading to false commands or misleading information, which may result in incorrect decisions by train operators or faulty data uploaded to the Rail Data Factory.

The second group of examples concerns the limited capability of data quality checks in rolling stock. Unlike protocols used in critical industries like aviation, which employ stringent error-checking procedures, protocols in rolling stock are more lenient when it comes to data validation. This might lead to:

- **Data Integrity Issues**: Without proper data quality checks, corrupted or incomplete data may go unnoticed, leading to unreliable information being used for (critical) train operations.
- **System Instability**: Malformed data packets or errors in communication may cause system malfunctions resulting in delays, defective components, and potential safety hazards.
- **Cyber Attacks**: Cyber attackers could exploit the absence of data quality checks to inject malicious data into the network, without the train or operator being able to detect the impact of this injection.

In the previous section, examples were given for cybersecurity vulnerabilities when applying data transfer in trains and its infrastructure. This paragraph gives an example of the Bowtie Risk Model [17] where cybersecurity vulnerabilities are translated into threat scenarios. The model is a visual representation to help analyse and manage risks. It visualises potential hazards, their causes, and the measures in place to prevent or mitigate them. The model consists of three main components:

1. **Hazard**: This is represented on the left side of the bowtie and signifies the event or hazard that could lead to a risk.
2. **Threats and Preventions**: These are measures put in place to prevent the threat from occurring or to reduce its likelihood. They are depicted on the left side of the bowtie, connecting the threat to the central knot.
3. **Mitigations and Consequences**: On the right side of the bowtie, these represent measures to mitigate the consequences if the threat materializes. They connect the threat to the central knot as well.

The central knot of the bowtie represents the actual risk event. It's the point where the threat connects to both preventative controls (left side) and mitigative controls (right side). The right side of the bowtie mirrors the preventative controls and shows how actions are taken if the hazard leads to an incident. When a model is filled it provides a visualisation of potential risks, their causes, and the safeguards in place. Table 3 shows an example how this analysis can be used for the vulnerabilities and risks described in this deliverable.

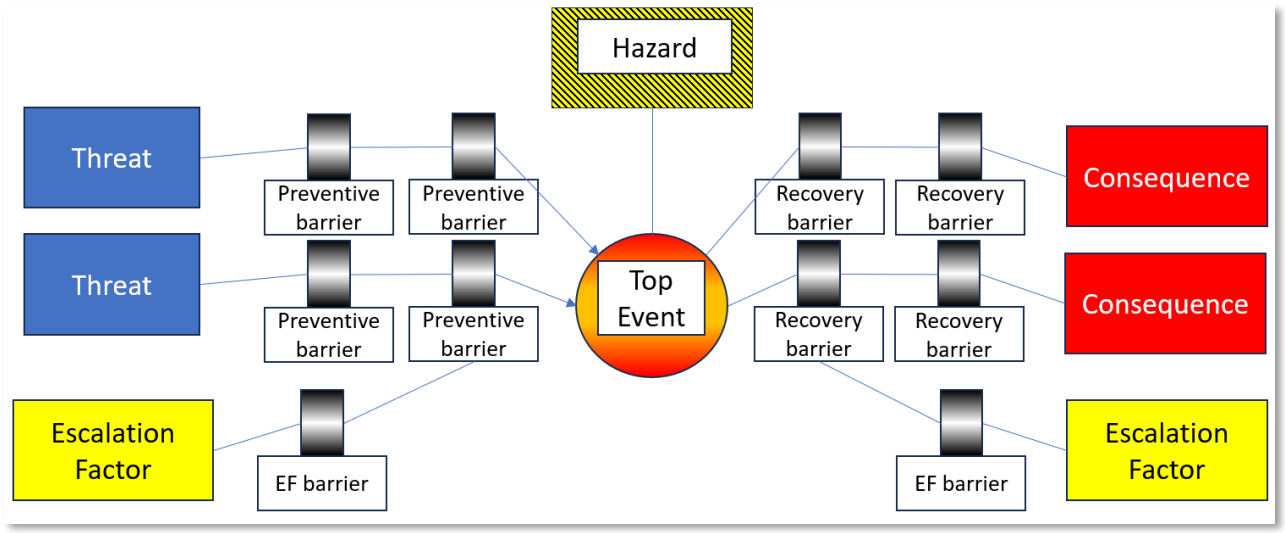

Figure 5. Bowtie Risk Model [17].

Table 3. Exemplary Bowtie Risk Table created for the Data Factory.

Hazard Rationale	*Incorrect data leads to incorrect correlations leading to incorrect operational decisions.* Successful supply chain attack on software development company, developing software for Touch Points
Top Event	Hacked software designated for use in Touch Points
Threat	Software development toolchain is hacked and wrong code is injected
Barrier	Only validated software development tools are used
Barrier	Developed software libraries are hashed and hash is (automatically) checked and validated
Threat	Developers make use of unsupported tooling
Barrier	Unsupported tooling is blocked
Barrier	Train software developers and raise their level of security awareness
Threat	File injection after security testing
Barrier	Block write access (CRUD) to files after testing
Barrier	Use protected connections
Escalation Factor	Software development company hires third party to do part of the job
EF barrier	Only supported tooling is allowed and other tooling is blocked
Consequence	Bad software installed in Touch Points leading to unusable data from trainsets
Recovery barrier	Check the hash before software is installed on Touch Points

3 DATA RISK MANAGEMENT

3.1 SENSOR DATA IS CRITICAL HIGH VOLUME DATA

Statistics on the collected data from a previous project “Sensors4Rail” [18] show a total of over 500 hours of time-synchronous multi-sensor data and functional data of the Sensors4Rail system were recorded from the trial operation in the Hamburg S-Bahn network. This corresponds to over 450 Terabytes of data. This includes the raw data from 14 sensors installed at the front of the vehicle (cameras, lidars, radars) as well as the course of the track detected by the system, detected landmarks, people, trains and potential obstacles in and at the edge of the clearance profile at any given time. The recording period covered a little less than a year (from May 2022 to the end of March 2023). The Sensors4Rail system was monitored remotely by the experts of the project team. The data extraction was done manually in the project and was only feasible with a high resource effort. However, based on the specifications outlined in D2, there is a growing need for an automated extraction process within the data touchpoint architecture of this project. The following data risk management considerations are seen across the entire process chain, from the vehicle to the Data Touchpoint to the data center.

Critical high volume sensor data presents several risks and challenges, ranging from privacy concerns to data quality issues:

- **Privacy and Security Risks:**
  - Data Breaches: As sensor data stores a vast amount of personal information, it becomes an attractive target for hackers and cybercriminals. A data breach can result in the exposure of sensitive information.
  - Privacy Violations: Collecting and analyzing large datasets may inadvertently reveal personal or sensitive information about individuals, potentially violating their privacy.
- **Data Quality and Accuracy:**
  - Garbage In, Garbage Out (GIGO): Poor-quality data can lead to inaccurate insights and decisions. Data systems for engineering and operations have to deal with noisy, incomplete, or inconsistent data, which can result in flawed analysis.
  - Data Bias: Biased data can lead to biased results, reinforcing existing prejudices or leading to unfair decisions. This is particularly important in areas like machine learning and AI.
- **Compliance and Legal Risks:**
  - Regulatory Compliance: Many industries have strict regulations governing data handling and storage (e.g. GDPR). Failing to comply with these regulations can result in significant legal and financial penalties.
  - Data Ownership: Determining who owns the data in a pan-European data ecosystem can be complex, leading to potential disputes and legal issues.
- **Ethical Concerns:**
  - Surveillance and Tracking: The extensive collection and analysis of data can be seen as invasive, raising concerns about mass surveillance and tracking individuals without their consent.

- Ethical Use: Using data for purposes that harm individuals or society can raise ethical questions. For example, using data to manipulate public opinion or discriminate against certain groups.

- **Data Governance Challenges:**
  - Data Governance: Managing and governing high volume sensor data can be complex. Without proper governance, organizations may struggle to maintain data quality, ensure data security, and enforce policies.
  - Data Silos: Data may be scattered across different departments and systems, making it difficult to access and integrate for analysis.
- **Scalability and Infrastructure Risks:**
  - Scalability: As data volume grows, infrastructure and processing power requirements increase. Organisations need to invest in robust infrastructure to handle high volume data effectively.
  - Technical Challenges: Data process technologies are evolving rapidly. Keeping up with the latest tools and techniques can be challenging for organisations.
- **Cost Overruns:**
  - Infrastructure Costs: Building and maintaining the necessary infrastructure for high volume and critical data can be expensive, and organisations may not see a return on investment if not managed efficiently.
  - Data Storage Costs: Storing large volumes of data can lead to significant ongoing expenses.
- **Analysis Paralysis:**
  - Information Overload: Having access to vast amounts of data can lead to information overload. Organisations may struggle to extract meaningful insights from the data deluge.
- **Vendor Lock-In:**
  - Dependence on Vendors: Organizations that rely heavily on third-party data solutions may become locked into specific vendors, limiting flexibility, and potentially leading to cost and compatibility issues.

To mitigate these risks, the pan-European Data Factory community must adopt robust data governance practices, prioritise data security and privacy, adhere to relevant regulations, and continually assess and adapt their high-volume sensor data strategies to evolving challenges and opportunities. One recommendation is to build a standardised framework to agree on the previous points as a framework among RUs and IMs.

3.2 DATA CLASSIFICATION

A large amount of data is expected that needs to be stored, processed, and retrieved, and consequentially data classification, data tagging, and metadata will be important. Data classification is the process by which data is categorised based on various parameters such as sensitivity and vulnerability. Data classification in IT security is vital to ensure that critical data is protected with appropriate levels of security.

Sensitive data classification is one of the primary requirements of the GDPR and other compliance standards. Numerous regulatory bodies mandate that sensitive personal data be protected against accidental loss, destruction, and damage. This can be done effectively only if this data is identified and classified appropriately.

Data classification levels

The various levels in which data is classified depends on the organisation and how it wishes to handle its data. The most common classification scheme consists of three levels:

Public:

Data classified as public is freely disclosed and does not have any access controls in place.

Private or internal:

Private data has minimal security restrictions in place and is intended for internal use within the organisation.

Restricted:

Files classified as restricted are also known as sensitive files and consist of highly sensitive internal data. Stringent access controls are put in place to ensure that these files are secure.

3.3 DATA ANNOTATION

Data annotation is the process of assigning a label to a piece of data, such as an image, website, or video. The tags associated are often metadata that indicate the author's name, date created, department, file format, or some other defining detail. These tags distinguish a data set from other data within an environment, making it easy to search for.

Why is data annotation important?

Data annotation provides an identity to your data by associating it with metadata. In an organisation, an employee ID serves the purpose of providing a unique identity to its employees. Similarly, in a football match, a seat number indicates the location of where you'll be seated in a stadium.

3.4 METADATA

Metadata tags are very important. Metadata describes the data itself, but also describes the age and quality of the data. Metadata tags are used to find the data (The ‘F’ of FAIR [11]) and used to determine whether the data can be used in the data factory. E.g. data that is too old or lacks a needed precision should not be used. Metadata tags can also be used to decide if and when data should be archived.

When metadata is defined it is highly recommended to use an already existing standard for metadata.

When dealing with camera imagery, especially stereo imagery with depth data, there are several metadata standards and formats. Below standards commonly used for sensor data:

- **EXIF (Exchangeable Image File Format)**: EXIF is a standard for storing metadata in image files. It includes information about the camera settings (such as exposure, aperture, and focal length), date and time of capture, and other technical details. While EXIF is primarily designed for 2D images, it can still be used to store basic information for stereo images.
- **XMP (Extensible Metadata Platform)**: XMP, developed by Adobe, extends the capabilities of EXIF. It allows for more extensive metadata to be embedded in image files. XMP can be

used to include additional information about the stereo imagery, such as depth data sources, camera calibration, and capture conditions.

- **LAS / LiDAR Metadata**: If you are dealing with stereo imagery that includes LiDAR data for depth information, the LAS format is commonly used for storing LiDAR point cloud data. The LAS metadata header contains information about point cloud attributes, sensor parameters, and data acquisition.
- **ISPRS Metadata Standards**: The International Society for Photogrammetry and Remote Sensing (ISPRS) has established standards for metadata in photogrammetry and remote sensing. These standards cover various aspects of image acquisition, including camera parameters, sensor specifications, and image orientation.

When dealing with stereo imagery and depth data, the metadata should ideally cover aspects such as camera calibration, image synchronisation, baseline distance (for stereo), depth map resolution, depth accuracy, and sensor specifications. Depending on your use case, you might need to combine multiple standards or design a custom metadata schema to capture all the relevant information.

When dealing with geospatial data there are already metadata standards and formats defined:

Geospatial Metadata Standards: this is a type of metadata applicable to geospatial data usually stored, maintained, and used in a Geographic Information System (GIS). The international metadata standard for geographic data is ISO 19115:2014 [19].

3.5 Difference between Metadata and Data Tagging

Metadata and data tagging are related concepts, but they serve slightly different purposes and have distinct characteristics:

Metadata:

Metadata refers to information about data. It provides context and additional details about a piece of data, making it easier to understand, manage, and organise. This type of data is characterised by low volume and the description of various data attributes, such as its source, creation date, format, author, location, and more. It is typically used for data management, data discovery, and data governance. It helps users locate and use data effectively. Metadata is often stored separately from the actual data it describes and is structured in a standardised way, making it machine-readable and searchable.

Data Tagging:

Data tagging, also known as data labelling or tagging data, involves the process of attaching labels or tags to specific data points or objects within a dataset. These labels or tags are used to classify or categorise the data, often for the purpose of training machine learning models. Data tagging is commonly used in supervised machine learning to create labelled datasets for tasks like image classification, text sentiment analysis, or object detection. The tags applied during data tagging may represent categories, classes, or attributes that are relevant to the specific task.

In summary, metadata provides general information and context about data, making it easier to manage and discover, while data tagging is a specific process of labeling data points within a dataset to create training data for machine learning models. While they serve different purposes, both metadata and data tagging contribute to effective data management and utilisation.

3.6 DIFFERENT DATA TYPES

There are numerous image and video formats. Image and video data is always large in file size, and therefore the Rail Data Factory should prevent data conversion when this type of data is used. Thus, the Rail Data Factory should standardise on a few image and video formats and fully support those images. Which data formats to use needs further investigation.

In the case audio, radar or lidar data is used for analysis in the Rail Data Factory the same consideration as for video and image applies.

Standard for Geospatial data

There is already an EU standard for Geospatial Data of Railway Networks [20]. This standard was designed to be able to report in a European standard but can be used as a starting point for railway network data.

Important when geospatial data is used from an EU perspective is that the rail industry agrees upon using the same Coordinate Reference System (CRS), and on a uniform way of transforming local data to the EU standard. This transformation should result in an EU wide and uniform coordinate precision. All coordinate systems are described by EPSG. EPSG stands for European Petroleum Survey Group and is an organisation that maintains a geodetic parameter database with standard codes [21].

E.g.: The Dutch standard currently is RD (Rijks Driehoekstelsel) which consists of an X and an Y Axis. When coordinates are transformed from RD to WGS84 or ETRS89 the Dutch transformation standard RDNAPTRANS is mandatory (in NL).

4 USABLE STANDARDS TO MITIGATE IDENTIFIED RISKS

In the following, European regulations and standards are listed that may help to mitigate the risks identified in the previous chapters.

4.1 EUROPEAN REGULATIONS

WBNI: As a rail service provider, NS has been designated as a vital organization, to which the WBNI applies [22]. The WBNI entails a number of rights and obligations that NS, including the ATO program, must comply with. For example, the NS is entitled to assistance from the National Cyber Security Center (NCSC) in taking measures to guarantee or restore the continuity of the services. In addition, NS is obliged to immediately report incidents with significant consequences for the continuity of the service provided by NS to the NCSC and must take appropriate measures to prevent incidents.

IT-SIG 2.0 (IT-Sicherheitsgesetz 2.0): The Dutch WBNI Act and the German IT Security Act 2.0 are in some ways comparable, as both acts are designed to increase the security of network and information systems and are a response to the EU NIS Directive (Directive on measures to ensure a high common level of security of network and information systems). However, the IT Security Act 2.0 is specific to Germany and extends the regulations of the first IT Security Act. It contains extended obligations for operators of critical infrastructures (KRITIS) and for certain providers of digital services as well as new regulations for manufacturers and providers of IT products and services. It also expands the powers of the Federal Office for Information Security (BSI).

NIS-2 directive: The NIS2 directive [23] is an extension of the already applicable WBNI legislation. Although the NIS2 directive is not yet in force at the time of writing, a draft legislative text is currently available in which the most important changes can be summarized in outline. These changes have an impact on NS and the ATO program in the sense that stricter requirements are set for subjects such as risk management, cyber incident management, cyber governance & supply chain risk.

BSI-KritisV: The BSI Critical Infrastructure Ordinance – BSI-KritisV is a German regulation concerning critical infrastructures (KRITIS). KRITIS are institutions and organizations of significant importance for the community, the failure or impairment of which would lead to sustained supply shortages, significant disruptions to public safety, or other dramatic consequences. The BSI-KritisV establishes requirements for operators of critical infrastructures to enhance the IT security of these facilities. This regulation falls under the Federal Office for Information Security (BSI) in Germany. Rail infrastructure providers and railway undertakings fall under "Section 8 Transport and Traffic Sector" (§ 8 Sektor Transport und Verkehr) [24] this means that they must meet certain security standards and implement measures to protect their network and information systems from cyber attacks. This also includes that they have to report security incidents and cooperate with the Federal Office for Information Security (BSI).

4.2 EUROPEAN STANDARDS

IEC 62443-3-2: In creating the ATO cybersecurity reference architecture, the 62443-3-2 standard was used as a guideline for designing a cyber security architecture for the ATO target solution. In addition, this standard was used in shaping the ATO risk analysis and cyber security requirements.

TS 50701: To give more structure to how risks are identified and cyber security requirements are created, the TS 50701 standardization was used.

ISO / IEC 27001: ISO 27001 specifies the requirements for the introduction, implementation, monitoring and improvement of an information security management system. The aim is to protect the confidentiality, integrity and availability of information by identifying risks and implementing appropriate security measures. The standard includes a set of security controls as well as best practice recommendations and enables organizations to demonstrate that they provide information security according to internationally recognized standards through ISO 27001 certification. For the railway provider, DB Netz AG, internal security and IT operational rules are based on ISO 27001, among other regulations like BSI-KritisV.

5 CONCLUSIONS

Sensor data is a critical component in training AI systems, particularly in achieving level GoA4 for fully automated driving within rail systems. Addressing the challenge that individual railway companies face in gathering sufficient sensor data, a proposal for a collaborative pan-European Rail Data Factory has been put forth. This initiative aims to enable and facilitate data collection, processing, simulations, AI model development, certification, and deployment across a unified European automated railway system.

Key Conclusions & Strategic Recommendations:

1. **Universal Standards & Agreements**: Develop and adopt universally accepted standards and agreements concerning data ownership, definitions, and formats among all stakeholders to ensure seamless integration and cooperation.
2. **Comprehensive Investigation & Standardization**: Conduct a thorough investigation to identify and standardize the essential information, data, and data exchange protocols, fostering effective pan-European interchange and collaboration.
3. **Data Security & Risk Management**: Address and mitigate data security threats, risks, and vulnerabilities associated with sophisticated models and large-scale data applications in the European train system. Implement robust data risk management, classification, tagging, and metadata strategies to effectively manage the extensive volume of sensor data, acknowledging its characterization as Critical High Volume Data.
4. **Regulatory Compliance & Risk Mitigation**: Leverage applicable European standards and regulations proactively to mitigate identified risks, enhance data exchange security, and boost the overall efficiency of the European railway system.

The key conclusion emphasizes the necessity of using universally accepted standards and agreements regarding data ownership, definitions, and formats among all involved parties. Also mandatory is the need for comprehensive investigation to determine the appropriate information, data, and data exchange standards for effective pan-European interchange. The RailDataFactory initiative addresses considerations of data application, benefits and challenges, data security threats and risks associated with sophisticated models, and relevant cybersecurity vulnerabilities in the European train system. It implements a robust and effective data risk management, classification, tagging and labeling, and metadata strategies to process the vast amount of sensor data. Finally, the initiative emphasizes leveraging usable European standards and regulations to mitigate identified risks and enhance the security and efficiency of data exchange within the European railway system.

Strategic Vision: This initiative envisions fostering a collaborative and secure data-driven environment, aiming to revolutionize and future-proof the European railway system through technological advancements and unified efforts. By addressing the challenges and harnessing the potential of high-quality, public available and standardized data, we can significantly enhance operational efficiency, safety, and innovation in railway automation across Europe.

The successful implementation of the Pan-European Rail Data Factory is contingent upon strategic leadership, cross-border collaboration, and a shared commitment to innovation and excellence. This will not only elevate the European railway system to new heights but also position it as a global leader in railway technology and automation.

6 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] CEF2 RailDataFactoryDeliverable 1, "Data Factory Concept, Use Cases and Requirements", Version 1.1, May 2023. [Online]. Available: https://digitale-schiene-deutschland.de/Downloads/2023-04-24_Rail Data Factory_CEFII_Deliverable1_published.pdf

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

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

[7] 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

[8] CEF2 RailDataFactory D2.2, "Technical specifications and available solutions for Identity Access Management (IAM), Data Management and Transfer and Cyber-Security", see https://digitale-schiene-deutschland.de/en/news/Pan-European-Railway-Data-Factory

[9] General Data Protection Regulation (2016), see https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

[10] 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

[11] FAIR Guiding Principles for scientific data management and stewardship, see https://www.go-fair.org/fair-principles/

[12] How to GO FAIR, see https://www.go-fair.org/how-to-go-fair/

[13] CEF2 RailDataFactory D3.2, "Evaluating the potential of the Rail Data Factory: A Business Case for Implementing Edge Computing and an open data infrastructure in the European Railway Industry", see https://digitale-schiene-deutschland.de/en/news/Pan-European-Railway-Data-Factory

[14] Dimitri van Zantvliet & Joseph Mager, Rail's need for cyber-resilience and digital solutions solutions in the face of rising cyber-threats (2022), Globalrailreview.com, see https://www.globalrailwayreview.com/article/136588/rails-need-for-cyber-resilience-cyber-threats/

[15] ENISA. Railway cybersecurity – Good practices in cyber risk management (2021), see https://www.enisa.europa.eu/publications/railway-cybersecurity-good-practices-in-cyber-risk-management

[16] Praerit Garg and Loren Kohnfelder, "The Threats To Our Products. More information about STRIDE", see https://learn.microsoft.com/en-us/azure/security/develop/threat-modeling-tool

[17] De Ruijter, A., & Guldenmund, F., "The bowtie method: A review", 2016, Safety science, 88, pages 211-218.

[18] Sensors4Rail, see https://digitale-schiene-deutschland.de/en/news/Sensors4Rail-A-successful-project-comes-to-an-end

[19] ISO 19115:2014, see https://www.iso.org/standard/53798.html

[20] EU standard for Geospatial Data of Railway Networks, INSPIRE, see https://inspire.ec.europa.eu/applicationschema/tn-ra

[21] Geodetic parameter database with standard codes, see https://epsg.io/?q=

[22] WBNI, see https://www.nctv.nl/onderwerpen/wet-beveiliging-netwerk--en-informatiesystemen/voor-wie-geldt-de-wbni/vitale-aanbieders

[23] NIS-2 directive, see https://www.enisa.europa.eu/topics/nis-directive

[24] Verordnung zur Bestimmung Kritischer Infrastrukturen nach dem BSI-Gesetz (BSI-Kritisverordnung - BSI-KritisV), see https://www.gesetze-im-internet.de/bsi-kritisv/BJNR095800016.html