CEF2 RailDataFactory

Deliverable 3.4 – Legal and regulatory assessment catalogue

Due date of deliverable: 31/08/2023

Actual submission date: 30/11/2023

Leader/Responsible of this Deliverable: Philipp Neumaier (DB)

Reviewed: **Y**/N

Document status		
Revision	Date	Description
01	25/09/2023	Document template generated
02	13/07/2023	Content transferred from Confluence
03	28/07/2023	First draft complete
04	21/11/2023	Version submitted to advisory board
06	30/11/2023	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	Sensitiv – 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
Philipp Neumaier	DB
Julian Wissmann	DB
Wolfgang Albert	DB
Waseem ul Aslam Peer	DB
Dr. Daniela Simone Kappler	DB
Keith Durczak	DB
Jens Glöckner	DB
Volker Eiselein	DB
Alexander Heine	DB
Bart du Chatinier	NS
Philippe David	SNCF
Patrick Marsch (only editorial effort)	DB

Note of Thanks

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

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 detect 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 RailDataFactory is needed, as a part of the overall 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.

In close sync with related activities listed in Section 1.2, the **CEF2 RailDataFactory** study focuses in particular on the High Speed pan-European Railway Data Factory backbone network and data platforms required to realise the vision of the pan-European RailDataFactory.

In this deliverable of the study, the current bottlenecks of transferring & mutating data within a rail network are studied. The report describes the challenges in data connectivity that are currently present while experimenting and deploying (AI) models within the rail industry. A gap between existing rolling stock and future technological advancements is described and proposals are made how the Rail Data Factory can be supplied with a constant flow of data by participating in a pan-European ecosystem.

ABBREVIATIONS AND ACRONYMS

Abbreviation	Definition
AI	Artificial Intelligence
BDGS	Bundesdatenschutzgesetz (German Federal Data Protection Act)
CISO	Chief Information Security Officer
DSD	Digitale Schiene Deutschland
ERA	European Union Agency for Railways
GDPR	General Data Protection Regulation
IM	Infrastructure Manager
IP	Intellectual Property
ISMS	Information Security Management System
ML	Machine Learning
RAMS	Reliability, Availability, Maintainability and Safety
RU	Railway Undertaking
TTDSG	Telekommunikations-Telemedien Datenschutzgesetz (German Data Protection Act for Telecommunications and Telemedia)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

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. In addition, routing of train could be an accessible and important resource. The challenge is to locate a train based on the given route.), 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, in particular, 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 aforementioned 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].

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 realised.

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 centers 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 has 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 center in Germany [10].

The project aims to investigate a possible Pan-European Railway Data Factory as a prerequisite for the development of a fully digitalised and automated railway system, focusing on the required underlying backbone network and data platform to be built.

1.3 Further Background for this Deliverable

The Pan-European Railway Data Factory is an interconnection of several independent Data Factories via a Pan-European Backbone Network. Each of them consists of one or more Data Center or Cloud resources and can also include Data Loggers in trains (rolling stock) and Touch Points at the trackside. The Touch Points are fetching the data from trains (rolling stock), process, cache and transfer the data via leased line connection, so that the data can be ingested into the storage system in the Data Center.

The aim of the Pan-European Railway Data Factory is to leverage data synergy effects in collecting, storing, managing und processing railway sensor data from trains (rolling stock) and other data sources. The goal is that each member can make use of the data, e.g. for machine learning (ML) training, data analysis, model evaluation, testing and homologation. This shall be ensured by a certified uniform software toolchain, while keeping the data sovereignty of each data provider.

A consortium constituted of participating members manages and governs the Pan-European Railway Data Factory. Thus, data will not be transmitted from one Data Center to another, but a user can access (via multi-tenancy) and process the data locally (e.g., search data, schedule jobs, update metadata, etc.; see D1 [5]).

The accessible data contains perception sensor data, localisation data, and many kinds of meta data. One important data type are annotations (e.g., labelled objects in camera images) which are a prerequisite for ML training. Generating annotations is very costly, and hence another benefit of the Pan-European Railway Data Factory is the joint requirements engineering, creation and usage of these annotations. A small portion of the data and Annotations will be used for a publicly available open data set, which on the one hand advertises the data quality of the Pan-European Railway Data Factory and on the other hand enables smaller companies and universities / institutes to do research in this field.

1.4 Aim and Structure of this Deliverable

The aim of this document is to provide an overview over legal and regulatory factors which have to be considered to achieve technological autonomy in essential digital computing infrastructure to process EU Railway data. The aim is to deliver an initial guidance for the data factories to ensure that the use of sensor, video technology and AI models complies with data protection and security requirements.

It is structured as follows:

- In Chapter 2, two exemplary “data pathways” are shown, and it is indicated which legal fields are likely involved in the different steps. Then, an overview and comparison of applicable European, national and corporate data protection regulations is provided,
- In Chapter 3, certification and homologation aspects are covered, and
- In Chapter 4, this document is concluded.

2 LEGAL FRAMEWORK

The legal framework for data factories in the context of railways across Europe is a complex and evolving landscape. With the rapid digitalisation of the railway industry, data factories play a pivotal role in collecting, processing, and analysing vast amounts of data to enhance safety, efficiency, and passenger experience. European Union regulations such as the General Data Protection Regulation (GDPR) and the European Railway Agency's (ERA) guidelines are fundamental in governing the collection, storage, and sharing of data within the railway sector. These regulations ensure the protection of sensitive and operational data while promoting interoperability among railway operators across borders. Moreover, data ownership and access rights are crucial aspects, requiring clear definitions and agreements among stakeholders. As the railway industry continues its digital transformation journey, a harmonised legal framework that encourages innovation, safeguards privacy, and promotes cross-border cooperation remains essential to the success and sustainability of data factories in the European rail sector.

The following figures show two exemplary “data pathways”. These are European cross-border scenarios involving the recording, transferring and processing of sensor data, and also including the annotation, AI training and model deployment in a train in ten steps.

Along with these ten steps, an assessment of the impact on five topics and legal areas is depicted. These topics are i) Data Protection, ii) Cyber Security, iii) Data Access, iv) Risks & quality control & Certification and v) Intellectual property (IP) and liability.

The influence or importance of these topics or legal areas on each step is represented in four stages, while empty cells means that there is no influence on the topic and “XXX” indicates a very large impact.

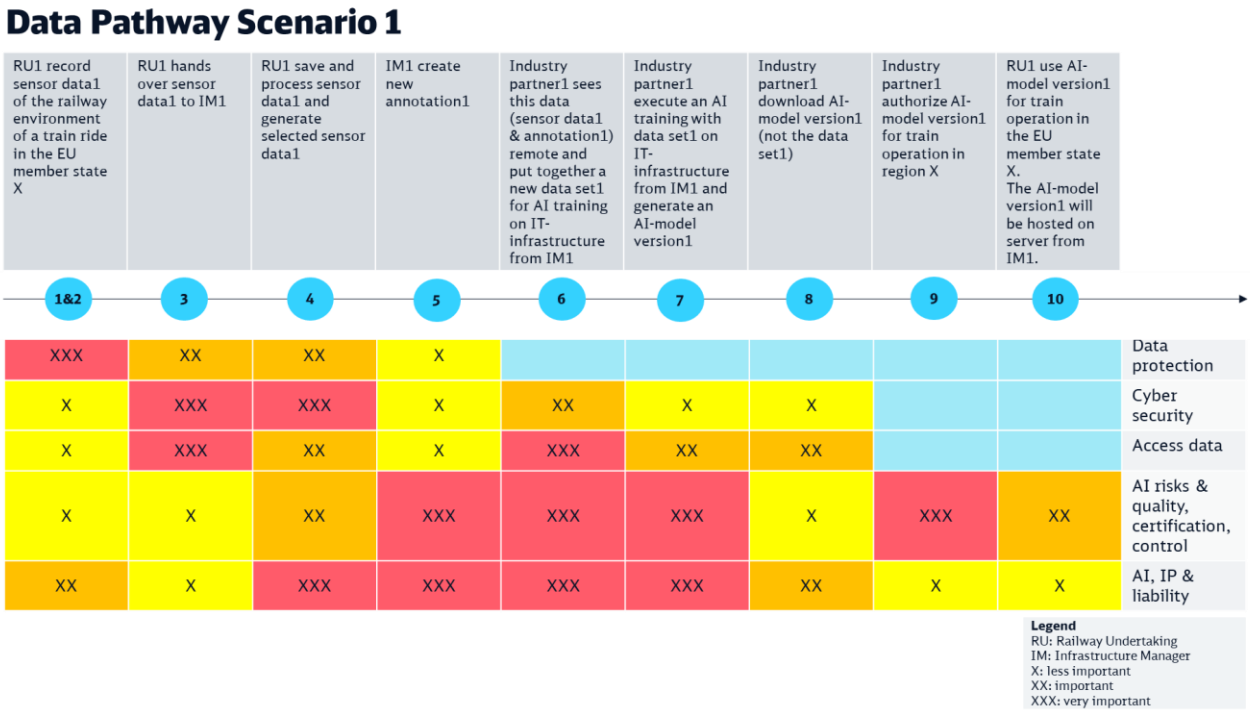

Figure 1. First data scenario: From data recording to AI model deployment in the train.

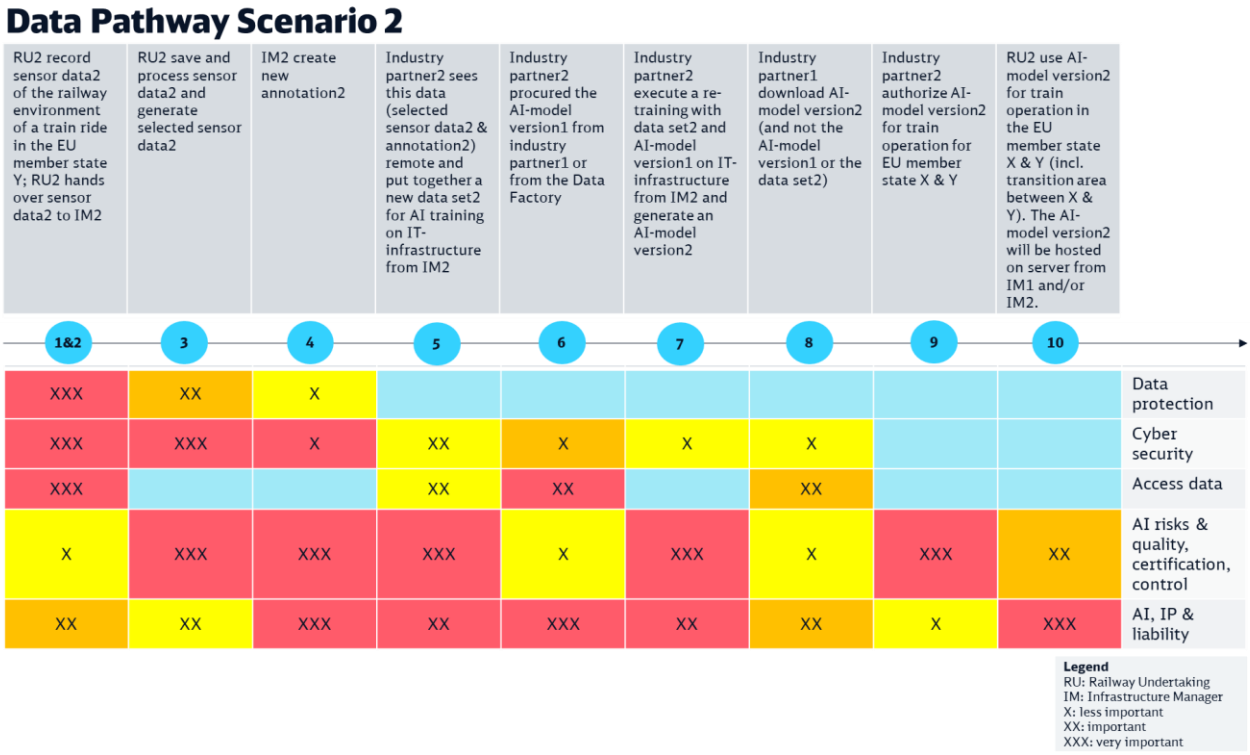

Figure 2. Second data scenario: From data recording to AI model deployment in the train.

Due to the diversity and complexity of these topics, we focus on data protection in this deliverable.

2.1 Overview of European Data Protection Regulations

Various stakeholders (such as IM's and industrial partners) develop and operate data factories in different EU member states. These data factories in different EU member states are to be connected to form a pan-European Data Factory. All parties involved can access the data centers of the other participants. In addition, third parties should be able to access the data centers. In order to ensure the data sovereignty of each data owner, there should be storage locations for data that is not shared ("private data") and externally accessible areas ("shared data"), as also stated in Deliverable D1 [5].

To take into account the aforementioned requirements for data processing and especially the data sharing with various data factories based in the EU, the focus in the creation of the legal framework was on laws, regulations and directives in context of:

- Privacy;
- Data integrity and security;
- Legal aspects of data sharing;
- Legal aspects in the context of AI;
- Railway-specific case law regarding data sharing.

The legal framework to target operations has several layers within. On top, there are European policies, below national policies or corporate policies like DB policies. We have incorporated a

consolidation of data protection policies, considering both European-wide perspectives and country-specific regulations, as part of the legal and regulatory assessment, as shown in Figure 3.

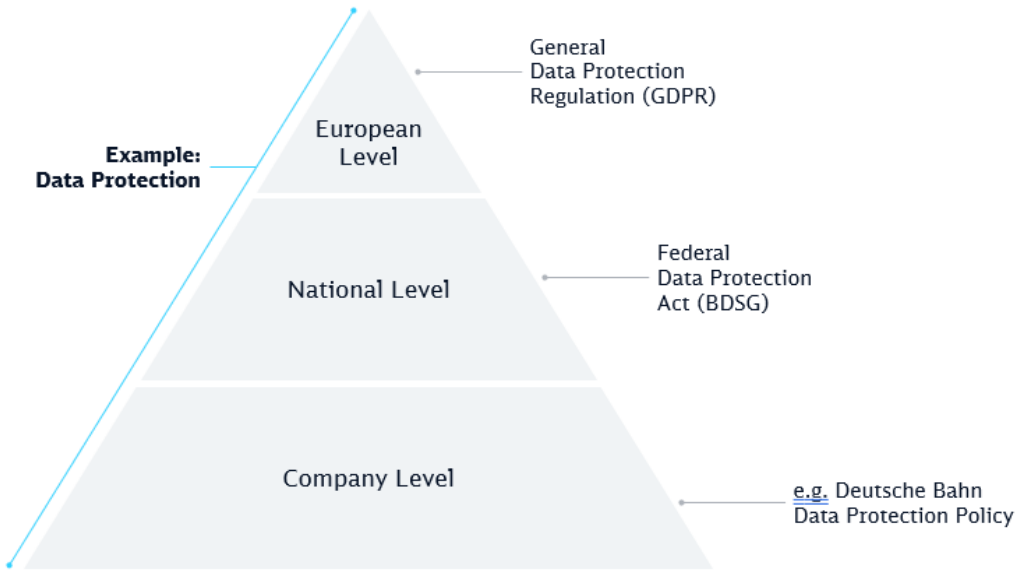

Figure 3. Norm pyramid following a data protection example.

The following overview includes the laws, regulations and directives analysed in the legal framework:

European Level Policy	National Policy (DE)	DB Policy
AI Act	General Railway Act (AEG)	Deutsche Bahn Confidentiality Classes for Information
Data Act	Federal Data Protection Act (BDSG)	Deutsche Bahn Data Governance Policy
Data Governance Act	Railway Construction and Operation Regulations (EBO)	Deutsche Bahn Data Protection Policy (national / international)
E-Privacy Directive	Regulation on the Use of Vehicle Data (FzTV)	Deutsche Bahn Geo Information Management
E-Privacy Regulation	NIS-2 Implementation and Cybersecurity Strengthening Act (NIS2UmsuCG)	Deutsche Bahn Information Security Policy
General Data Protection Regulation (GDPR)	Regulation on Telecommunications Data Protection (TTDSG)	Deutsche Bahn Legal requirements, video technology, rights of the persons

Directive on Security of Network and Information Systems (NIS Directive)		Deutsche Bahn Use of video technology
Regulation on Data Protection for EU Institutions and Bodies		
Regulation on the Free Flow of Non-Personal Data		

This document provides an initial indication and thus an overview of relevant regulations, directives and laws at EU and German level, including an initial assessment. The guidelines that are valid and binding in the DB Group are also listed. An initial state of affairs regarding legal developments is also presented. **This document does not constitute a legal opinion on the pan-European Data Factory**. Likewise, it is not a concept for a legal model to implement the pan-European Data Factory in a legally compliant manner. A legal assessment was also not carried out. Therefore, **there is no guarantee of completeness with regard to the legal framework**. Likewise, **there is no guarantee of a legally conclusive assessment**.

2.2 Applicability of the Laws, Regulations and Directives on Data in the Data Factory

2.2.1 European Level - In place

2.2.1.1 General Data Protection Regulation (GDPR) - EU Regulation 2016/679

The General Data Protection Regulation is a comprehensive regulation governing data privacy for EU citizens. The GDPR exists to protect individuals whose personal data are used. The GDPR applies directly to all member states of the EU. Due to the fact that personal data will be collected this impacts data usage in the railway sector. On the train side or on the track side, railway undertakings (in regular operation and during test runs) and other stakeholders such as universities or industrial partners use various types of sensors (e.g. cameras, LIDAR, radar) to collect image, location and other data (all these data are hereinafter referred to as "sensor data"). In particular, the camera images can have a personal reference (for example, when filming people on/on the track or on the platform). The GDPR has application priority over the BDSG.

Date of query: 22.09.2023: Datenschutz-Grundverordnung (DSGVO) – Finaler Text der DSGVO inklusive Erwägungsgründe (dsgvo-gesetz.de)

Applicability: High

Area: Video, Images, Trackside and Train

2.2.1.2 Directive on measures for a high common level of cybersecurity across the Union (NIS2 Directive) - EU Directive 2022/2555

EU NIS2 is the European Framework for Critical Infrastructure Operators. EU NIS2 Focuses on ensuring cybersecurity and resilience for critical infrastructure, including railways. Therefore this regulation addresses security and integrity with regard to the movement of data within the EU which is impacting the data processing and usage in the railway sector.

Date of query: 22.09.2023: The Directive on security of network and information systems (NIS Directive) | Shaping Europe's digital future (europa.eu); KRITIS-Dachgesetz für Resilienz ab 2023 – OpenKRITIS

Applicability: High

Area: Networks, Cyber Security

2.2.1.3 Regulation on the Free Flow of Non-Personal Data - EU Regulation 2018/1807 (proposal)

The General Data Protection Regulation (GDPR) already provides for the free movement of personal data within the EU. This Regulation will therefore ensure a comprehensive and coherent approach to the free movement of all data in the EU. Therefore this regulation addresses the free movement of non-personal data across EU borders which is impacting the data processing and usage in the railway sector.

Date of query: 22.09.2023: Free flow of non-personal data | Shaping Europe's digital future (europa.eu)

Applicability: High

Area: Overall non-personal data

2.2.1.4 ePrivacy Directive - EU Directive 2002/58/EC (proposal)

The E-Privacy Directive deals with the privacy of electronic communications which is impacting the data processing and usage in the railway sector.

Date of query: 22.09.2023: Proposal for an ePrivacy Regulation | Shaping Europe's digital future (europa.eu)

Applicability: Low

Area: Certification of data, Security and Trust

2.2.1.5 European Data Governance Act - EU Regulation 2022/868

The Data Governance Act (valid from September 2023) seeks to increase trust in data sharing, strengthen mechanisms to increase data availability and overcome technical obstacles to the reuse of data. Therefore the Data Governance Act will also support the set-up and development of common European data spaces in strategic domains, involving both private and public players, in sectors such as health, environment, energy, agriculture, mobility, finance, manufacturing, public administration and skills. This Regulation will impact public authorities who held data for sharing and also companies and other legal bodies considering voluntary donations of data. It further will have an impact on the railway sector if data sharing via European data spaces will be considered. E.G. creation and publishing of Open Data Sets.

Date of query: 06.09.2023: European Data Governance Act | Shaping Europe's digital future (europa.eu)

Applicability: Medium

Area: OpenDataset, Data Sharing

2.2.1.6 Data Act - EU Regulation 2022/0047 (COD)

On the 28 June 2023, a political agreement was reached between the European Parliament and the Council of the EU on the Data Act. The Act is now subject to formal approval and once adopted, will enter into force 20 days after Official Journal publication, becoming applicable after roundabout 18 months. As of November 2023, the law is in force. The Data Act clarifies who can create value from data and under which conditions. The Data Act will ensure fairness by setting up rules regarding the use of data generated by Internet of Things (IoT) devices. Only the users of networked devices can decide how data they have helped to create should be handled. Users can be companies as well as consumers. For the Railway sector there will be an impact if IoT-Data will be created or used.

Date of query: 06.09.2023: Data Act | Shaping Europe's digital future (europa.eu); EUR-Lex - 52022PC0068 - EN - EUR-Lex (europa.eu); Data Act — Factsheet | Shaping Europe's digital future (europa.eu); Data Act (dihk.de); Update 16.11.20203: Datengesetz: Neue Regeln für besseren Zugang und bessere Nutzung von Daten | Aktuelles | Europäisches Parlament (europa.eu)

Applicability: Medium

Area: IoT Data

2.2.2 European Level – Upcoming

2.2.2.1 ePrivacy Regulation (Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC)

The European Commission's proposal for a Regulation on ePrivacy aims at reinforcing trust and security in the digital world. The Commission has started a major modernisation process of the data protection framework over the past few years, which culminated in the adoption of the General Data Protection Regulation (GDPR). The ePrivacy legislation needs to be adapted to align with these new rules. The (new) ePrivacy Regulation will repeal the (current) ePrivacy Directive.

Date of query: 22.09.2023: Proposal for an ePrivacy Regulation | Shaping Europe’s digital future (europa.eu)

Applicability: Low

Area: Communications content and metadata

2.2.2.2 Regulation on Data Protection for EU Institutions and Bodies - EU Regulation 2018/1725

Regulation 2018/1725 sets forth the rules applicable to the processing of personal data by European Union institutions, bodies, offices and agencies. It is aligned with the General Data Protection Regulation and the Data Protection Law Enforcement Directive. It entered into application on 11 December 2018. This regulation covers specific data protection rules to EU institutions which are e.g.:

- the European Parliament (Brussels/Strasbourg/Luxembourg);
- the European Council (Brussels);
- the Council of the European Union (Brussels/Luxembourg);
- the European Commission (Brussels/Luxembourg/Representations across the EU);
- Based on the current understanding this regulation has no impact on the data processing and usage in the railway sector.

Date of query: 26.09.2023: Data protection in the EU (europa.eu); Types of institutions, bodies and agencies | European Union (europa.eu)

Applicability: not applicable

Area: --

2.2.2.3 AI Act - EU Regulation 2021/0106 (COD)

The AI Act regulate artificial intelligence (AI) to ensure better conditions for the development and use of this innovative technology. Basically, if an AI system can harm a human being within the European Union, the requirements from the EU AI Act come into force.

Specifically, as soon as an AI system is operated, distributed, or used in the EU market (i.e., even if the operator, provider, or developer is established outside the Union), the AI Act and the possible penalties take effect. According to the current version, non-compliance with the requirements can result in penalties of up to €30,000,000 or 6% of sales.

The impact to the railway sector will be there due to the fact that if AI Systems are used a valuation procedure according to the risk taxonomy has to be conducted. If the AI Systems is rated at "high risk" e.g. assessments for AI development process (end-to-end data management, interpretability, documentation accuracy, quality control) have to be put in place. In addition, it must be taken into account that DB is one of the operators of critical infrastructure and thus, in particular, the ANNEX III HIGHER RISK CI SYSTEMS REGARDING ARTICLE 6(2) has an effect. There is mentioned, that if there are "AI systems intended to be used as security components in the management and operation of road transport and in water, gas, heat and power supply" these are considered a high risk AI system.

Date of query: 06.09.2023: EU AI Act: first regulation on artificial intelligence | News | European Parliament (europa.eu); EUR-Lex - 52021PC0206 - EN - EUR-Lex (europa.eu); Was ist KRITIS? - AG KRITIS

Applicability: High

Area: Simulation, AI Models

2.2.3 Further Monitoring Activies on European Level

As already mentioned in the introduction the purpose of the "Data Factory" for the railway sector is to enable technological and operational conditions in order to fully automated rail operation (so-called GoA4 rail operation). In other industries regulations / directives with respect to data protection in the context of autonomous driving could be identified. E.g. UNECE WP.29 und ISO/SAE 21434 - Cybersecurity Management System (CSMS) and Software Update Management System (SUMS) for automotive in order to enable autonomous driving.

Specific regulations / directives for rail operators are missing.

2.2.4 National Policy (DE) - In place

2.2.4.1 Railway Construction and Operation Regulations (EBO)

This regulation may contain provisions regarding data protection in relation to railway construction, operation, and maintenance activities.

In general, the following recommendation for action should be noted at this point: In the context of specific use cases, the following classification mechanisms / application criteria have to be considered in order to decide on the applicability of this regulation or not.

Railroad facilities and vehicles must be designed to meet safety and security requirements. Data provision in this context would be a possible use case. Whether this regulation applies would have to be evaluated specifically for each use case.

Date of query: 13.09.2023: EBO - nichtamtliches Inhaltsverzeichnis (gesetze-im-internet.de)

Applicability: Very Low

Area: --

2.2.4.2 Federal Data Protection Act (BDSG)

Governs the processing of personal data by federal institutions and private entities operating in Germany's railway sector. It ensures data protection rights and obligations are upheld. This regulation is highly signification if data with personnel reference or personal referenceability is used or shared or processed. As mentioned in the GDPR context the camera images and videos can have a personal reference (for example, when filming people on/on the track or on the platform) which implies the applicability of the new version of BDGS.

In general, the following recommendation for action should be noted at this point: In the context to specific use cases following classification mechanisms / application criteria has to be considered in order to decide on the applicability of this regulation or not:

- Personal data;
- Personal referenceability data.

Date of query: 13.09.2023: BDSG - nichtamtliches Inhaltsverzeichnis (gesetze-im-internet.de)

Applicability: Medium

Area: Personal data, personal referenceability data

2.2.4.3 General Railway Act (AEG)

The AEG establishes regulations concerning the operation and management of railways in Germany. It may include provisions related to data protection and privacy in the railway context. The General Railway Act (AEG) obliges railroads to operate safely and to keep infrastructure, vehicles and equipment in a safe condition. It also contains the authorizations for railroad supervision, the requirements for plan approval or the regulations on the need for and receipt of operating licenses or safety certificates.

In general, the following recommendation for action should be noted at this point: In the context of specific use cases following classification mechanisms / application criteria have to be considered in order to decide on the applicability of this regulation or not:

- Data Provisions regarding Railroad Accident Investigation including Data Protection restrictions;
- Data Provisioning regarding Passenger information;
- Regarding the applicability of the law for Digitale Schiene Deutschland (DSD) Projects context / data use case Specific research must be done.

Date of query: 13.09.2023: AEG - nichtamtliches Inhaltsverzeichnis (gesetze-im-internet.de)

Applicability: Medium to High

Area: Data Provisions regarding Railroad Accident Investigation including Data Protection restrictions

2.2.4.4 Regulation on Telecommunications Data Protection / Telekommunikations-Telemedien Datenschutzgesetz (TTDSG)

This regulation focuses on the protection of telecommunications data, which may be relevant to railway communication systems. The Telecommunications Telemedia Data Protection Act (TTDSG) renews the national implementation framework of the e-privacy directive.

In general, the following recommendation for action should be noted at this point: In the context of specific use cases following classification mechanisms / application criteria have to be considered in order to decide on the applicability of this regulation or not:

The storage of and access to information in the terminal equipment of an End User by the Telemedia Services. Whether this regulation applies would have to be evaluated specifically for each use case if Telemedia services will be provided.

Date of query: 13.09.2023: TTDSG.pdf (gesetze-im-internet.de)

Applicability: Low

Area: Protection of telecommunications data

2.2.4.5 Railway Specific Employment Data Protection

May include specific policies or guidelines addressing the protection of employee data within the railway sector, ensuring confidentiality and compliance with relevant laws.

Based on the current understanding this regulation has no impact on the data processing and usage due to the fact, that the regulation handle the responsibility of railroad companies for employees in cooperation with other stakeholders.

Date of query: 13.09.2023: EBA - Fachmitteilungen - Verantwortung von Eisenbahnunternehmen für Mitarbeiter in der Zusammenarbeit mit anderen Akteuren (bund.de)

Applicability: Very Low

Area: Protection of Employee Data

2.2.4.6 Proof of equal safety / Nachweis gleicher Sicherheit

The proof of equal safety results from a comparison with a reference system. This approach is gaining importance in the future as it is one of three ways to define or prove safety requirements given by the European Railway Agency (ERA). Based on the current understanding this regulation has no impact since we are not doing operations of the trains and the data is purely for AI related domain, this data does not decide actually operation of the train so it is safe to say "further research can be done to check validity"

Date of query: 06.09.2023: EBA - Homepage - Anforderungen an den Nachweis gleicher Sicherheit (bund.de)

Applicability: TBD

Area: N/A

2.2.4.7 Second act on increasing the security of IT systems (German IT Security Act 2.0)

The IT Security Act 2.0 provides for a number of obligations for operators of critical infrastructures. Among other things, operators are to Provide for minimum security standards for critical infrastructures (e.g., the use of intrusion detection systems according to § 8a IT Security Act 2.0), comply with security requirements for critical components (see below), Comply with information obligations and reporting requirements vis-à-vis the Federal Office for Information Security (abbreviated to "BSI") (e.g., list all IT products that are important for the functionality of critical infrastructures, report malfunctions).

This regulation seems to be highly significant, due to the fact that the DB infrastructure is considered critical infrastructure.

Date of query: 26.09.2023: BSI - The German IT Security Act 2.0 (bund.de)

Applicability: TBD

Area: N/A

2.2.5 National Policy (DE) – Upcoming

2.2.5.1 NIS-2 Implementation and Cybersecurity Strengthening Act./ NIS-2-Umsetzungs- und Cybersicherheitsstärkungsgesetz – NIS2UmsuCG/

On 27.12.2022, the NIS-2 Directive was published in the EU Official Journal and came into force on 16.01.2023. The NIS-2 Directive must be transposed into national law by September 2024. For the implementation of the NIS-2 Directive in Germany, there is already a draft of the Federal Ministry of the Interior (NIS2UmsuCG). (Status July 2023) Affected companies and organizations must take appropriate measures in areas such as cyber risk management, supply chain security, business continuity management, penetration testing and incident response, and reporting to the authority and remediation.

This regulation seems to be highly significant, due to the fact that the DB belongs to the critical infrastructure.

Date of query: 22.09.2023: Referentenentwurf des BMI: NIS-2-Umsetzungs- und Cybersicherheitsstärkungsgesetz – NIS2UmsuCG - AG KRITIS; Microsoft Word - CI1_17002_41_22 3-16 (Referentenentwurf NIS2UmsuCG - 03-04-2023 09-00) (kritis.info)

Applicability: High

Area: Networks, Cyber Security

2.2.5.2 Regulation on Telecommunications Data Protection (TTDSG)/ Telekommunikations-Telemedien Datenschutzgesetz (TTDSG)

This regulation focuses on the protection of telecommunications data, which may be relevant to railway communication systems. The Telecommunications Telemedia Data Protection Act (TTDSG) renews the national implementation framework of the e-privacy directive.

In general the following recommendation for action should be noted at this point: In the context of specific use cases following classification mechanisms / application criteria have to be considered in order to decide on the applicability of this regulation or not:

The storage of and access to information in the terminal equipment of an End User by the Telemedia Services. Whether this regulation applies would have to be evaluated specifically for each use case if Telemedia services will be provided.

Date of query: 13.09.2023: TTDSG.pdf (gesetze-im-internet.de)

Applicability: Low

Area: Protection of telecommunications data

2.2.6 Further Monitoring Activies on National level

2.2.6.1 Regulation on the Use of Vehicle Data (FzTV)

This regulation addresses the handling and use of vehicle data including data protection measures. There couldnt be a specific regulatory for the railway sector indentifiied. The mentioned regualtions refers to vehicles which are part of the Road Traffic Act. Therefore it seems not relevant so fare

Date of query: 13.09.2023: FzTV - nichtamtliches Inhaltsverzeichnis (gesetze-im-internet.de)

Applicability: NA

Area: NA

2.2.7 Deutsche Bahn (DB) Policies - In place

2.2.7.1 Deutsche Bahn Geo Information Management

The policy provides a regulatory framework for the geo information management in the DB group to regulate the handling of geo-resources.

Date of query: 06.09.2023

Applicability: High

Area: Processing Geodata

2.2.7.2 Deutsche Bahn Data Protection Policy

The policy is an overarching policy that outlines Deutsche Bahn's commitment to protecting personal data of employees, customers, and stakeholders. It also covers data collection, processing, retention, and security measures. All the Policies within the DB context do also follow the ISO standard that cover ISO 27000 Series as well.

Date of query: 29.08.2023

Applicability: High

Area: ISMS for Video, Images, Trackside and Train.

2.2.7.3 Data Governance Policy

The aim of the policy is to define the target state of the data organisation and also to define the data and metadata management in the DB Group.

Date of query: 29.08.2023

Applicability: High

Area: Data ownership

2.2.7.4 Information Security Policy

The aim of the policy is to establish and operate the Information Security Management System (ISMS) in the BU/SU and on corporate level:

- Establishment of Chief Information Security Officer (CISO) role and tasks;
- Clear structure, process and documentation of the ISMS.

Date of query: 29.08.2023

Applicability: High

Area: ISMS within Data Factory and on level of the sector programme “Digitale Schiene Deutschland” (DSD)

2.2.7.5 Confidentiality classes for information

The aim of the policy is to assignment of confidentiality classes with respect to the protection requirements.

Date of query: 29.08.2023

Applicability: High

Area: Open data vs. non open data and different projects or user roles

2.2.7.6 Use of video technology

The aim of the policy is to ensure the use of video technology in the DB Group complies with data protection and security requirements.

Date of query: 29.08.2023

Applicability: High

Area: Data protection with respect to video on trackside and on trains.

2.2.7.7 Legal requirements, video technology, rights of the persons

The aim of the policy is to ensure that the use of video technology in the DB Group complies with data protection.

Date of query: 29.08.2023

Applicability: High

Area: Legal restrictions with respect to Video technology and personal data

2.2.7.8 DB-internal Principles of Privacy

This is a summary of the DB internal regulations regarding data processing, purpose, minimisation, transparency and quality, as shown in Figure 4.

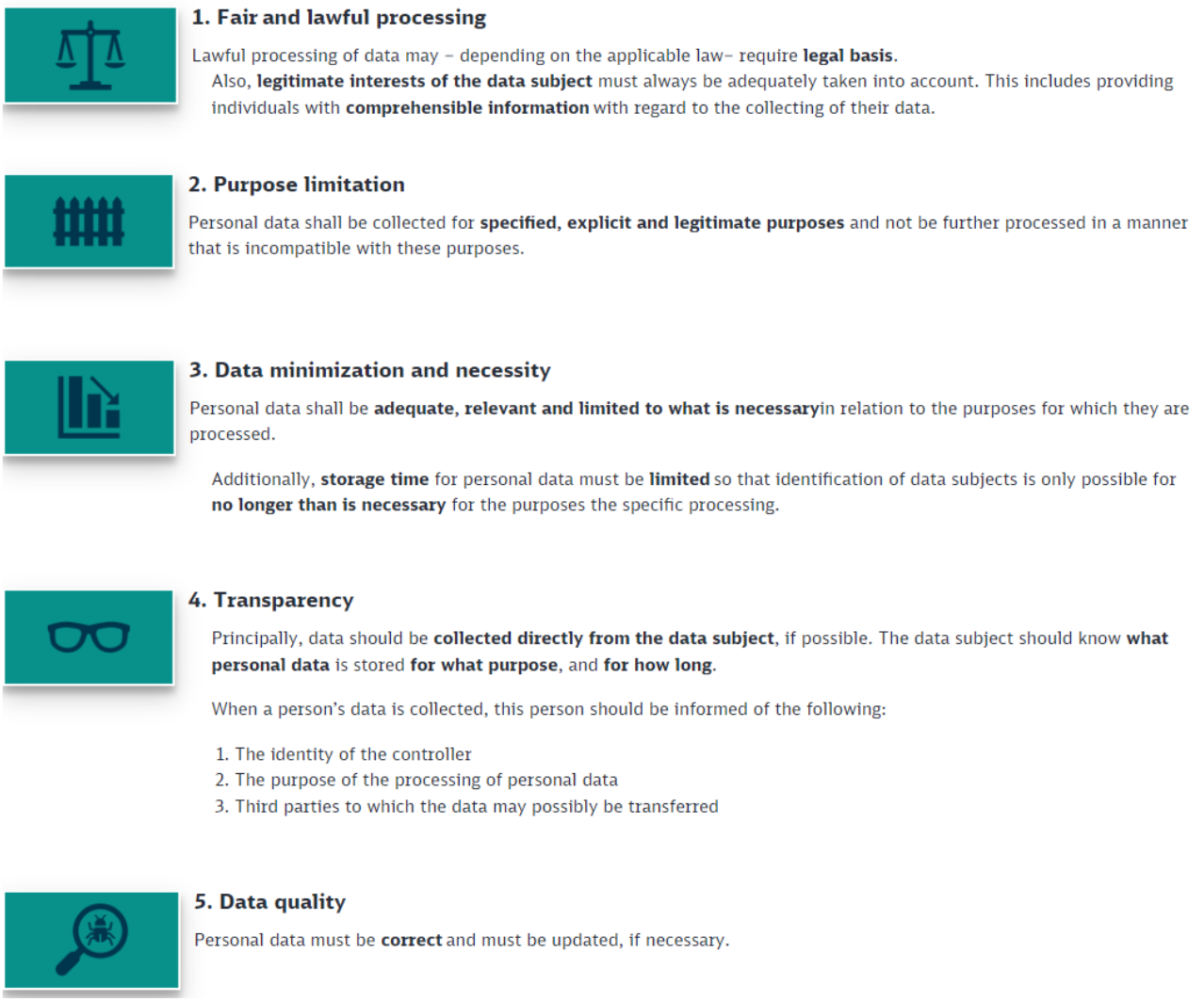

Figure 4. DB-internal principles of privacy.

Date of query: 16.10.2023

Applicability: High

Area: Data processing

2.3 Comparison of the Regulations and Policies from different IMs within Europe

Comparison of EU and national laws

All of these regulations must comply with the national laws of the member countries of the infrastructure service providers and the GDPR.

In order to gain a comprehensive understanding of the different additions to the GDPR in different countries, we have carefully analysed the national legislation of these countries and compared it with the provisions of the GDPR.

In Table 1, a comparison of the most significant topics in this context is presented.

Table 1. Comparison of EU and national laws.

	EU GDPR	**GER** Federal Data Protection Act	**FR** Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	**NL** Uitvoeringswet Algemene verordening gegevensbescherming
Source	https://gdpr-info.eu/	https://www.gesetze-im-internet.de/englisch_bdsg/index.html (Deviation from GDPR)	https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/2018-05-25/ (Deviation from GDPR)	autoriteitpersoonsgegevens.nl/uploads/imported/uavg.pdf (Deviation from GDPR)
Legal principles	When processing personal data, all principles listed in Art. 5 GDPR must be followed.			
Legal basis	The processing of personal data is lawful if at least one of the six conditions of Art. 6 I GDPR is met and all conditions of Art. 5 GDPR are met.	Pursuant to Section 24 BDSG, private entities may, under certain circumstances, process personal data for a purpose other than that for which the data was collected.	The legal bases for lawful data processing are defined in Art. 7. The structure differs from the GDPR in that consent must be given for processing unless another legal basis (the same as in the GDPR) is relevant.	
Sensitive data	Art. 9 and Art. 10 GDPR list certain categories of data whose processing is subject to stricter regulations:	Section 22 I No. 1 BDSG allows the processing of special categories of personal data despite Art. 9 I DSGVO, if	Art. 8 excludes the applicability of the general ban on processing personal data under certain conditions.	Articles 18, 23, 24, 25, 26, 27, 29, 30, 32 and 33 of the UAVG define exceptions and specifications in which the

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
	The processing of these data is generally prohibited. However, exceptions are defined.	necessary to exercise the rights derived from the right to social security and social protection. For example, also for the purpose of assessing the employee's ability to work.	Art. 9 specifies by whom the processing of personal data on criminal convictions and criminal offenses pursuant to Art. 10 GDPR may take place.	processing of sensitive data is allowed.
Information requirements	The controller must provide certain information to the data subject at the time the data is collected. The information that must be communicated depends on whether the data controller received the data directly from the data subject or from another entity.	The BDSG defines exceptions to the information requirements. For example, they do not apply under certain conditions if the further processing of data takes place in analog form The information obligations also do not apply if their provision would have a detrimental effect on the well-being of Germany and the interest of the data controller in not providing the information outweighs the	Article 32 specifies the information requirements in more detail. For example, the obligation to name the data protection officer does not apply if the appointment is mandatory. Also, for example, no information obligations need to be fulfilled if the data are subsequently processed for purposes pursuant to Art. 89 GDPR.	According to Art. 41(1) UAVG, data controllers may disregard the information obligations (Art. 13, 14 DSGVO) to the extent necessary and proportionate to ensure certain objectives. These objectives are, for example: - the prevention, investigation, detection and prosecution of breaches of professional rules for regulated professions

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
		interest of the data subjects. Furthermore, they also do not apply if the provision of information to data subjects would prejudice the establishment, exercise or defense of legal claims, Further, other cases were also mentioned in which a duty to provide information would not apply, such as in the case of video surveillance of public areas.	Furthermore, it is provided, among other things, that if collected personal data are anonymized within a short period of time by a procedure previously approved by the CNIL, only the identity of the controller and the purpose of the processing must be informed.	- the protection of the data subject or the rights and freedoms of others
Automated decision-making	Individuals have the right not to be subject to decisions based solely on automated processing (without human intervention in the decision-making process),	Exceptions: Insurance contract and credit rating	According to Art. 10, no judicial decisions evaluating the behavior of individuals may be issued on the basis of automated data processing with the aim of	Those responsible may make automated individual decisions if this is necessary to fulfill a legal obligation or to perform a task in the public interest.

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
	including profiling. Data subjects may exercise this right where such decisions produce legal effects or other substantial effects concerning them.		personality assessments.	However, this exception only applies to automated individual decisions that are not based on profiling (Art. 40 UAVG).
Rights of data subjects	The direct obligations to ensure the rights of data subjects are the responsibility of the controller, with the assistance of the processors. Rights: Right of access to relevant data and information Right to rectification Right to erasure Right to restriction of processing Right to object	According to Section 27 II BDSG, the rights of the data subject in Art. 15 GDPR (access), Art. 16 GDPR (rectification) Art. 18 GDPR (restriction of processing) Art. 21 GDPR (objection) limited to the extent that such rights may render impossible or seriously prejudice the achievement of the research or statistical purposes and such restrictions	Art. 32 specifies further rights of the data subjects. In particular, the information rights of the data subjects.	Art. 41, UAVG and the following articles define exceptions where the rights of the data subjects may be limited.

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
	Right to data portability	are necessary for the fulfillment of the research or statistical purposes. There are also further regulations for cases of confidentiality obligations or a disproportionate effort for deletions.		
Processing on behalf of a controller	Controllers may outsource their processing activities to other companies under certain conditions. (Art. 28 GDPR)			
Records of processing activities	Controllers and processors are required to keep records. The content requirements are very similar for both. The records must be kept in writing, including in electronic form. They must be made available to the supervisory	Initial guidance from the Data Protection Conference on how the General Data Protection Regulation should be applied in practical enforcement: https://www.datenschutzkonferenz-		

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
	authority upon request.	online.de/media/kp/dsk_kpnr_1.pdf		
Data breaches	The controller is required to document all data breaches, which include the relevant facts about the breach, its effects and the remedial actions taken. In addition, when the Processor learns of a data breach, it must notify the Controller "without undue delay."	The obligation to inform the data subject of a personal data protection breach pursuant to Art. 34 GDPR does not apply if this would disclose confidential information that must be kept secret by law or by its nature, in particular because legitimate interests of third parties prevail. The data subjects must be informed if their interests outweigh the need for secrecy (Section 29 BDSG).	Further conditions are formulated. For example, the competent supervisory authority must be informed within 72 hours of becoming aware of the data breach.	
Data protection impact assessment	A controller is obliged to carry out a data protection impact assessment if its processing activities are likely to present a high risk to the rights and	The Data Protection Conference has published a list of processing activities for which a data protection impact assessment is mandatory.	CINIL has published a list of processing operations where the performance of a data protection impact assessment is required:	

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
	freedoms of individuals. Such a high risk may occur when new technologies are used.	The detailed list includes, for example: Collection of personal data via interfaces of personal electronic devices that are not protected against unauthorized access and cannot be detected by data subjects (NFC payment) Use of artificial intelligence to process personal data to control interaction with the data subject or to evaluate personal aspects of the data subject Geolocation of employees https://www.lda.bayern.de/media/dsfa_muss_liste_dsk_de.pdf	Processing of health data used by health care institutions or medico-social institutions to care for people Processing of genetic or biometric data of particularly vulnerable individuals (e.g., employees) Profiling of individuals for human resources management purposes Monitoring of employees Processing operations that may result in data subjects being excluded from a contract or in the termination of a contract Profiling with data from external sources A list has also been published containing processing	

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
			operations without obligation of a data protection impact assessment.	
Data protection officer	It is necessary because the data protection officer is the link between the controller and the data subjects and the supervisory authority.	According to Section 38 BDSG, a data protection officer is required if a data protection officer generally employs at least ten persons who are involved in the automated processing of personal data. In addition, a data protection officer is required for private entities with fewer than ten persons involved in the automated processing of personal data if the data protection officer or processor is subject to a data protection impact assessment or if they commercially process personal data for the purpose of		

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
		transmission, anonymized transmission or market or opinion research.		
Data transfer	Data transfer within the EU is guaranteed to flow freely is guaranteed, but data transfers outside its borders are subject to stricter rules.			
Data protection for employees	Member States may provide for more specific rules to ensure the protection of rights and freedoms with regard to the processing of personal employee data.	Section 26 BDSG permits the processing of employees' personal data if this is necessary for employment-related purposes. Can also be based on consent, which must be voluntary.		Art. 30 I UAVG allows employers to process health data of their employees in the course of employment. In accordance with Art. 33 III UAVG, personal data on criminal convictions and criminal offenses against employees may only be processed if permitted under the Works Council Act (Wet op de ondernemingsraden).

	EU GDPR	GER Federal Data Protection Act	FR Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés	NL Uitvoeringswet Algemene verordening gegevensbescherming
Archiving, scientific and historical research	Art. 89 regulates the processing of personal data for archiving purposes in the public interest or for scientific, historical or statistical research. Although "scientific research" is not defined, Recital 159 clarifies that processing for scientific research purposes should include, for example, technological development and demonstration, basic research, applied research and privately funded research.	Section 27 of the BDSG permits data processing without the consent of the data subject for the purposes of scientific or historical research and for statistical purposes.	Pursuant to Art. 36, the Code du patrimoine provides for regulations regarding the security of data processing for the purposes specified in Art. 89 of the GDPR, as well as regarding retention and deletion periods.	Articles 24 and 32 UAVG allow the processing of special categories of personal data as well as data related to criminal matters if this is necessary for scientific or historical research purposes serving a public interest or statistical purposes. It is also permitted to process such data when it is not possible or feasible to request the explicit consent of the data subject, provided that appropriate safeguards are in place to protect the privacy of the data subject. According to Article 28 UAVG, the processing of genetic data is permitted under the same conditions.

Each EU member state has the opportunity to specify certain aspects of the GDPR in national law or establish additional regulations. This can lead to certain differences in the individual member states. These differences relate to specific national provisions and regulations that go beyond the general provisions of the GDPR.

This overview provides an overview of the different approaches in Germany, France and the Netherlands. The comparison shows that the individual additions made by the member states are not so far apart that a common framework cannot be created.

Comparison of the policies from different IMs within Europe

We have carefully compiled the international and internal guidelines of Deutsche Bahn (DB) and identified data protection provisions and legal requirements contained therein. Likewise, we analysed the publicly available information of SNCF regarding data protection as well as that of NS ProRail.

This comprehensive comparison aims to find out the importance of data protection in these renowned companies and to identify the relevant regulations. We want to find out if there is any overlap or if there is a common understanding of these laws within these organisations.

With respect to Deutsche Bahn (DB), we rely on the ISO 27001 standards, which are implemented in our own regulations. We have carefully compared these with the publicly available and shared regulations of other infrastructure service providers. For example, SNCF has published a comprehensive overview of its data protection policies on its website (www.sncf.com/en/personal-data). NS ProRail, on the other hand, internally relies on the "Ethical Approach of the use of AI", which describes how artificial intelligence can be used ethically and responsibly. They also have internal "Data Ownership" and "Data Usage Boards" that define specific rules for data sharing.

The data protection provisions listed here, in particular the Artificial Intelligence Act, must be implemented across all connected data factories in order for the data to be useable across the entire system. The resulting requirements should therefore be covered by a harmonized framework in the future.

3 CERTIFICATION AND HOMOLOGATION ASPECTS

The standards EN 50126, EN 50128, and EN 50129 are European standards that define requirements for the safety of railway applications. They are part of the regulatory framework necessary for the development, implementation, and certification of safety-relevant railway systems in Europe.

EN 50126 – Railway Applications - Specification of requirements for Reliability, Availability, Maintainability, and Safety (RAMS): This standard addresses the Reliability, Availability, Maintainability, and Safety (RAMS) of railway systems. It defines the processes, activities, and tasks required to achieve and maintain a specific level of safety.

For the pan-European Data Factory, standardisation is central to addressing the challenges associated with ensuring the reliability and safety of development, especially for AI given their inherent complexity and potential unpredictability. It is equally important to ensure the maintainability and availability of these new algorithms under different operating conditions. By adhering to

established standards, it is possible to mitigate risks and achieve a higher level of safety in terms of the performance and security of the algorithms used in different applications.

EN 50128 – Railway Applications - Software for railway control and protection systems: This standard specifies requirements for the software lifecycle for railway applications. It establishes procedures, techniques, and documentation that must be considered during the development and validation of safety-relevant systems.

Demonstrating that railway applications meet safety and quality standards requires extensive testing and validation, which is complex due to the inherent non-determinism of some AI models. Ensuring traceability and integrity of development data is critical.

EN 50129 – Railway Applications - Safety-related electronic systems for signalling: This standard deals with safety-related electronic systems, especially focusing on hardware and interfaces to other systems. It sets requirements for the safety and reliability of these systems.

In particular, the dynamic nature and complexity of AI algorithms can be a challenge in ensuring that all potential failures are detected and properly addressed. It is important to ensure that the system can fail safely and that all safety requirements are met throughout. It must always be ensured that the system can be brought to a safe state.

For certification and homologation, manufacturers and developers must demonstrate that their products and systems meet the requirements of these main standards with many requirements, like development for operation in railway tunnels. This typically includes:

- **Risk Analysis and Risk Assessment**: Identification and evaluation of potential risks and definition of measures for risk reduction.
- **Development and Documentation**: Compliance with the development processes and techniques specified in the standards and the creation of the necessary documentation.
- **Validation and Verification**: Proof that the system meets the specified requirements and that the safety measures are effective.
- **Independent Assessment**: An independent safety assessment by a notified body (Notified Body) may be necessary to verify compliance with the standards. For the different safety integrity levels fixed role models in the standards are defined.

For AI applications in railways, a notable challenge across all three standards is the intrinsic variability and non-determinism of AI algorithms. Demonstrating compliance requires addressing these issues, ensuring transparency, reliability, and safety throughout the system's lifecycle. Additionally, ethical considerations, such as data privacy and bias in decision-making, should be addressed during the application of AI in such critical systems.

Data-Driven AI view:

- **Data Quality and Integrity**: Ensuring the quality and integrity of the development data is essential, as the performance of AI algorithms heavily relies on the input data;

- **Verification and Validation**: Rigorous verification and validation (V&V) processes are required to ensure that the AI algorithms perform as expected under all conditions;
- **Traceability**: Traceability of the development process, including data sources and processing steps, is crucial to ensure compliance with these standards;
- **Safety Assurance**: Providing comprehensive safety assurance for data-driven AI algorithms can be challenging due to their adaptive and learning nature.

In summary, compliance to EN 50126, EN 50128, and EN 50129 standards is critical in development and implementation of data-driven engineering the railway industry, but can also be challenging due to the complex, dynamic, and sometimes non-deterministic nature of AI algorithms. Rigorous testing, validation, and documentation, along with ensuring data quality and integrity, are key to overcomeing these challenges and achieving certification and homologation. The standard DIN EN 50128 is to be withdrawn with a transitional period and the subsequent standard DIN EN 50657 "Railway applications - Applications for railway vehicles - Software for railway vehicles, except railway signalling applications" is to be applied [11].

At the current level of development, it remains still a challenging question how a future fully automated railway system could be proven to be safe or how a homologation could be argued because in an open world scenario (which has to be assumed in the general railway context), according to current regulation, the system must be proven beforehand to react correctly to all possible scenarios under all possible circumstances (while considering accepted limitations).

It is a common and accepted assumption (also following other domains such as automotive) that the approaches listed in the previous paragraphs will play a big role in a safety argumentation, e.g. in the context of scenario-based development. However, the rail environment poses some specific challenges here because the typical frequencies of objects around rail tracks can be very low – meaning that in a scenario-based development approach, many different variations of very rarely-occurring situations or obstacles need to be used for training and testing. For a system that relies purely on recorded data, this amount of sample data can be infeasibly high to record which is why simulation of input data is proposed as a possible solution.

Simulation of sensor data in a scenario-based development makes it possible to generate multiple versions of slightly different scenarios (e.g., by generating various colors, textures or velocities of objects) and to use these data for training or testing during development.

In the context of data protection, it has to be considered that on the one hand for virtual scenarios derived from recorded scenarios, similar restrictions as for recorded data may apply because there might be data cues transferred from the real world into simulated data. On the other hand, the data used for generation should likely be of high, undistorted quality – meaning that e.g. anonymization techniques should be avoided or used carefully so as not to tamper the (potentially multiple times) generated data. The same applies to augmentation techniques which modify recorded data in order to create further variation in the scenarios.

Another part of a future homologation will likely be explainability approaches for AI (“XAI”) that aim to explain the reasoning behind a result obtained by using AI methods. Such methods could help to find undesired biases in the training data or could also identify systematic biases in decisions drawn by AI. In this sense they can constitute good means to identify the issues described in the previous paragraph.

4 CONCLUSION

In conclusion, this document has provided an overview of the legal and regulatory factors essential for achieving technological autonomy in the processing of EU railway data within the digital computing infrastructure. It has also offered initial guidance on data protection policies to ensure compliance with data protection and security requirements at both EU and national levels, as well as within the organisation.

However, it is important to note that this document does not address the operational implementation and integration of these rules, laws, and regulations into the data processing processes of the data factory. Further work is needed to specify and evaluate their impact on data sharing along the data flows and to validate any necessary enhancements to internal guidelines.

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] DIN EN 50128, see https://www.eba.bund.de/SharedDocs/Fachmitteilungen/DE/2017/09_2017_Uebergangsregelgung_zur_Abkuendigung_der_DIN_EN_50128.html