

# Datendiversität, Ontologien und Datengovernance im digitalisierten Bahnsystem

## Data diversity, ontologies and data governance in the digitalised rail system

Waseem ul Aslam Peer | Imran Mehaboob Hussain | Anna Bardakh | Patrick Marsch | Bernardo B Ramalho Immendorff | Daniel Obermeyer

**B**asierend auf dem Roll-out von ETCS und digitalen Stellwerken arbeiten viele Bahnen in Europa wie auch die Deutsche Sektorinitiative Digitale Schiene Deutschland (DSD) an der weiteren Digitalisierung des Bahnsystems. Dazu zählen auch Maßnahmen in Richtung KI-basierter Zugdisposition und vollautomatisiertes Fahren (GoA 4). Im derart weiterentwickelten Bahnsystem spielen Datenhaltung und deren effiziente Handhabung eine entscheidende Rolle. Dieser Beitrag zeigt die Diversität von Daten im digitalisierten Bahnsystem auf und vertieft zwei wesentliche Maßnahmen, die in diesem Zusammenhang wesentlich sind: die Standardisierung von Datenontologien und geeignete Datengovernance.

### 1 Einführung

Das System Bahn stößt aufgrund steigender Nachfrage und langjähriger Infrastrukturherausforderungen an seine Kapazitätsgrenzen. Gleichzeitig ist der Schienenverkehr für nachhaltige Verkehrs- und Klimaschutzziele von entscheidender Bedeutung. Aus diesen und anderen Gründen arbeitet die deutsche Sektorinitiative DSD [1] gemeinsam mit europäischen Partnern daran, die Kapazität, Qualität und Effizienz des Bahnbetriebs weiter zu steigern, u. a. durch weiterentwickelte Leit- und Sicherungstechnik (LST) für dichtere Zugfolgen, auf Künstlicher Intelligenz (KI) basierende Disposition des Bahnbetriebs und durch vollautomatisiertes Fahren (sog. Grade of Automation 4, GoA 4).

Diese Entwicklungen bringen neue Technologien wie KI und Sensorik ins Bahnsystem und erhöhen die Datenvielfalt im System weiter, was sowohl Herausforderungen als auch Chancen birgt. Darüber hinaus ist das Bahnsystem wesentlich auf die Qualität und Integrität der Daten angewiesen.

In diesem Beitrag wird ein Überblick über die Daten im Bahnsystem gegeben und werden die wichtigsten Merkmale erläutert. Anschließend werden drei wesentliche Voraussetzungen für den Umgang mit der Datenvielfalt im Bahnsystem vertieft: die Standardisierung des Datenmanagements und der Datenontologien sowie die Einführung einer geeigneten Daten-governance.

### 2 Übersicht über Daten im digitalisierten Bahnsystem

Das digitalisierte Bahnsystem umfasst eine große Vielfalt an Datenquellen und -formaten. Dies ist in Bild 1 dargestellt, wo Beispiele danach kategorisiert sind, woher die Daten stammen (von der Infrastruktur- oder Zugseite usw.) und wo im Ablaufzyklus des Bahnsystems die Daten verwendet werden.

Die Daten können wie folgt weiter charakterisiert werden:

**M**any European railways and the German Digitale Schiene Deutschland (DSD) sector initiative are striving toward higher levels of automation in the rail system involving AI-based train disposition and fully automated driving (GoA 4) based on the rollout of ETCS and digital interlockings. Data and its efficient management are playing an increasingly vital role in such an evolved rail system. This article depicts the data diversity in the digitalised rail system and highlights two essential means of handling and leveraging this, i.e. the standardisation of data ontologies and appropriate data governance.

### 1 Introduction

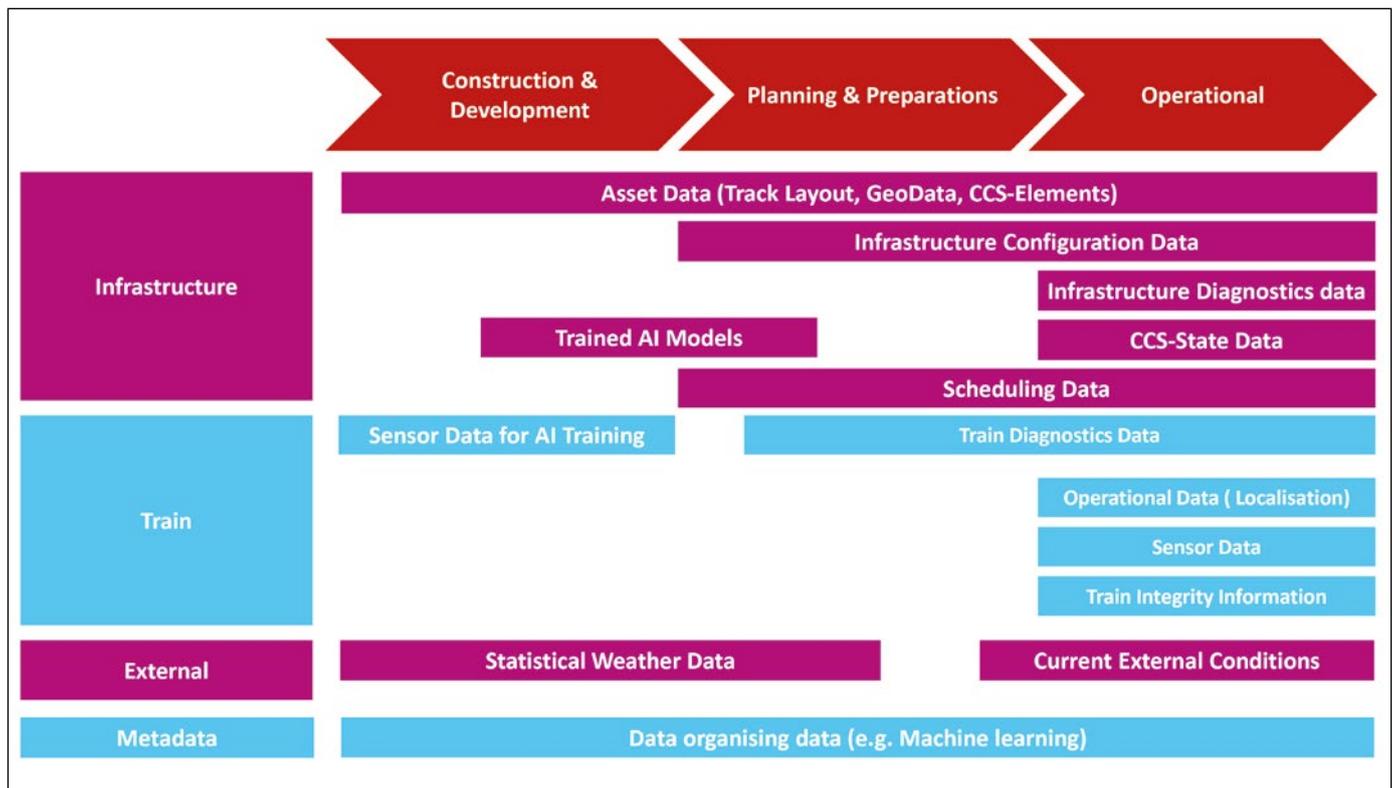
The rail system is reaching its capacity limits due to increasing demand and longstanding infrastructure challenges. Despite this, rail transport remains vital for sustainable transportation and climate protection goals. For these and other reasons, the German DSD sector initiative [1] is working with its European partners on further enhancing the capacity, quality and efficiency of rail operations, for instance using evolved control command and signalling (CCS) for driving at minimal distances, artificial intelligence (AI) based traffic disposition and fully automated driving (so-called Grade of Automation 4, GoA 4).

These developments are bringing novel technologies into the rail system, such as AI and advanced sensing, and have further increased the data diversity in the system, thereby posing both challenges and opportunities. Furthermore, the rail system is becoming increasingly reliant on the quality and integrity of data. This article provides an overview of the data in the rail system with an introduction to its key characteristics. Two essential prerequisites for handling the data diversity in the rail system are then detailed, i.e. the standardisation of data management and data ontologies and the introduction of a suitable data governance.

### 2 An overview of the data in the digitalised rail system

The digitalised rail system contains a broad diversity of data sources and formats. This is shown in fig. 1, where data examples have been categorised according to where the data originates (on the infrastructure or train side etc.) and when in the rail system's lifecycle the data is used.

The data can be further characterised using the following properties:



**Bild 1: Beispiele von Daten in Entwicklung, Umsetzung und Betrieb des digitalisierten Bahnsystems**  
 Fig. 1: Examples of the data involved in the development, rollout and operation of the digitalised rail system

**2.1 Sicherheitskritikalität**

Unter (funktional) sicherheitskritischen Daten versteht man Informationen, die für den sicheren Bahnbetrieb essenziell sind. Diese müssen strengen Validierungs-, Schutz- und Sicherheitsmaßnahmen unterzogen werden, um sicherzustellen, dass es bei ihrer Generierung, Integration oder Übertragung nicht zu Kompromittierungen kommt. Beispiele hierfür sind:

- Signalisierungsdaten, z. B. bezogen auf Gleissignale, Blocksysteme, Achszähler, Weichenansteuerung
- Zugposition und -integrität, einschließlich des Status sicherheitsrelevanter Systeme wie Bremsen
- Monitoring-Daten, z. B. im Zusammenhang mit kritischen Gleismängeln
- Infrastruktur- oder Konfigurationsdaten, die an sicherheitskritischen Vorgängen beteiligt sind.

Andererseits stehen nicht-sicherheitskritische Daten nicht in direktem Zusammenhang mit dem sicheren Bahnbetrieb, sondern unterstützen Betrieb und die Wartung. Diese Daten erfordern meist weniger Schutz und Validierung und umfassen z. B. Zugfahrpläne und Reiseprofile.

**2.2 Statische, dynamische und echtzeitkritische Daten**

Statische Daten bleiben im Laufe der Zeit weitgehend unverändert und sind für Planung und Analyse essenziell. Beispiele hierfür sind geografische Koordinaten sowie Daten von Weichen, Kreuzungen, Bahnsteigen, Signal- oder Oberleitungsausrüstung. Dynamische Daten ändern sich regelmäßig, z. B. durch Betriebsbedingungen oder Wartungstätigkeiten. Beispiele hierfür sind aktualisierte Zugfahrpläne, Wartungsprotokolle, Gleisbelegungsinformationen oder Wetterdaten.

Kritische Echtzeitdaten umfassen zeitkritische Informationen, die für die sofortige Entscheidungsfindung unerlässlich sind. Beispiele

**2.1 Safety-criticality**

Safety-critical data refers to information that is critical to safe railway operations. It must undergo rigorous validation, protection and security measures in order to ensure that no compromise occurs during its generation, integration or transfer. Examples include:

- signalling system data, e.g. related to trackside signals, block systems, axle counters, point switches,
- train position and integrity, including the status of safety-critical systems like the brakes,
- monitoring data, e.g. related to critical track defects,
- infrastructure or configuration data involved in safety-critical operations.

On the other hand, non-safety critical data is not directly related to safe rail operations, but supports broader rail operations and maintenance. This data might not be as rigorously protected or validated as safe data and it includes train schedules and journey profiles, for instance.

**2.2 Static, dynamic and real-time critical data**

Static data remains largely unchanged over time and is crucial for planning and analysis. Examples include geographical coordinates and data related to switches, junctions, platforms and signalling or catenary equipment.

Dynamic data changes regularly, e.g. influenced by operating conditions or maintenance activities. Examples include updated train schedules, maintenance logs, track occupancy information or weather data.

Real-time critical data encompasses time-sensitive information that is essential for immediate decision-making. Examples include train location and movement data within the context of the European Train Control System (ETCS).

le hierfür sind Zugpositions- und Zugbewegungsdaten im Zusammenhang mit dem European Train Control System (ETCS).

### 2.3 Dateneigentum

In vielen Fällen sind die Daten im Bahnsystem Eigentum von Herstellern, die so die Kontrolle über die Konfigurationen und Diagnosedaten ihrer installierten Komponenten behalten. Andererseits sind Infrastruktur- und Betriebsdaten wie Zugpositionen, Gleisbelegungsinformationen etc. meist im Eigentum von Eisenbahninfrastrukturunternehmen.

Diese unterschiedlichen Eigentumsverhältnisse beim Datenbesitz führen zu Herausforderungen bzgl. Integration und Zugänglichkeit, da Hersteller häufig proprietäre Protokolle für ihre Systeme verwenden. Beispielsweise erfordert die Integration von Konfigurations- und Leistungsdaten verschiedener europäischer Anbieter im ETCS-Kontext einen hohen Aufwand und Koordination.

### 2.4 Strukturierte und unstrukturierte Daten

Strukturierte oder modellierte Daten sind solche, die einer Verfeinerung und Organisation unterzogen wurden, um eine präzise Analyse zu ermöglichen. Umgekehrt werden Daten, die sich noch in einem unverarbeiteten Zustand befinden, wie z. B. rohe Sensordaten, als unstrukturiert bezeichnet.

## 3 Neuartige Datentypen und -systeme im weiterentwickelten Bahnsystem

Die Weiterentwicklung des Bahnsystems bringt neuartige Datentypen mit sich:

### 3.1 Umfangreiche Fahrzeugsensordaten

Beim vollautomatisierten Bahnbetrieb (GoA 4) [2, 3] werden LiDAR-, Radar- und Kamerasensoren an Zügen eingesetzt, um Objekte und Gefahren auf den Gleisen zu erkennen. Jeder Sensortyp erzeugt unterschiedliche Dateneigenschaften, von hochauflösenden Kamerabildern bis hin zu dreidimensionalen Darstellungen der Umgebung. Die Daten können gemeinsam mit den Daten des Global Navigation Satellite System (GNSS) und Odometriesensoren verarbeitet werden, um eine hochpräzise Lokalisierung mittels Landmarkenerkennung zu ermöglichen [4]. Insgesamt erfordert GoA 4 umfassende Fahrzeugsensordaten, wie in Bild 2 dargestellt, mit anspruchsvoller Handhabung.

### 2.3 Ownership

In many cases, the data in the rail system is owned by the equipment suppliers who maintain proprietary control over the configurations and diagnostics for their installed components. On the other hand, infrastructure and operating data such as train positions, track occupancy information etc. is usually owned by the rail infrastructure managers.

Such heterogeneous data ownership results in challenges related to integration and accessibility, as suppliers often use proprietary protocols for their systems. For instance, the integration of configuration and performance data from various European suppliers within an ETCS context requires extensive effort and coordination.

### 2.4 Structured and unstructured data

Structured or modelled data refers to information that has undergone refinement and organisation in order to enable a precise analysis. Conversely, data that is still in an unprocessed state, such as raw sensor data, is characterised as unstructured.

## 3 Novel data types and systems in the evolved rail system

The further evolution of the rail system has led to some novel data types:

### 3.1 Comprehensive vehicle sensor data

Fully-automated train operations (GoA 4) [2, 3] will involve LiDAR, radar and camera sensors on trains to detect objects and hazards on the tracks. Each sensor type leads to different data properties, ranging from high-resolution camera images to three-dimensional representations of the surroundings. The data may be processed together with Global Navigation Satellite System (GNSS) and odometry sensor data for high-precision localisation based on landmark detection [4]. Overall, GoA 4 will require comprehensive vehicle sensor data, as shown in fig. 2, with sophisticated handling requirements.

### 3.2 Annotated and simulated data within the context of AI training

The training of the AI involved in the automated hazard detection further requires the raw sensor data to be complemented with annotated and simulated data. Raw data serves as the foun-

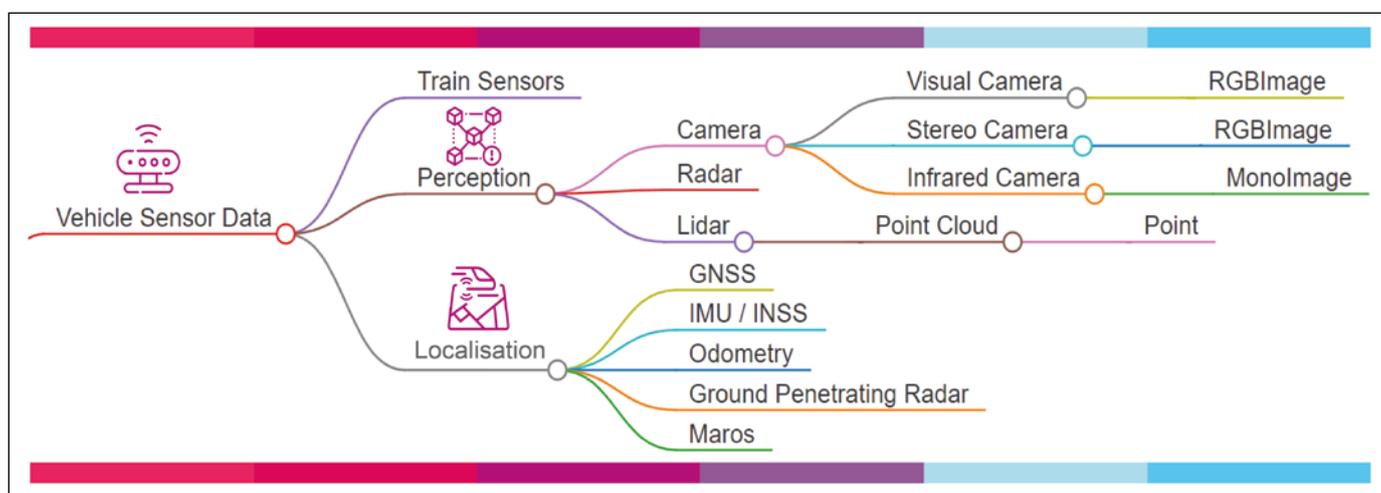


Bild 2: Erwartete zugseitige Sensordaten im Kontext des vollautomatisierten Fahrens (GoA 4)

Fig. 2: The anticipated vehicle sensor data set within the context of fully-automated driving (GoA 4)

### 3.2 Kommentierte und simulierte Daten im Kontext des KI-Trainings

Für das Training der KI zur automatisierten Gefahrenerkennung ist es zudem erforderlich, rohe Sensordaten durch annotierte und simulierte Daten zu ergänzen. Rohdaten dienen als grundlegende Basis, annotierte Daten liefern eine kontextbezogene Kennzeichnung, die für überwachtetes Lernen unerlässlich ist, und simulierte Daten ermöglichen Modelltests unter kontrollierten virtuellen Bedingungen.

### 3.3 Trainierte KI-Modelle

KI-Modelle werden darauf trainiert, Gleise, Oberleitungen, Brücken und andere Objekte eindeutig zu identifizieren. Da diese Objekte jedoch von Land zu Land unterschiedlich sein können, müssen Züge, die in mehreren Ländern verkehren, mit KI-Modellen ausgestattet sein, die speziell auf die genaue Erkennung von Objekten in jeder Region zugeschnitten sind.

Darüber hinaus werden neuartige Datensysteme etabliert:

### 3.4 Digitales Register

Das in Europe's Rail standardisierte Digitale Register (DR) [5] ist ein Repository für Infrastrukturdaten, das als „Single Source of Truth“ (SSOT) für verschiedene Bahn-Subsysteme dient. Es ist eine wesentliche Komponente für den hoch- bis vollautomatisierten Bahnbetrieb (GoA 2 – GoA 4) und für ETCS Level 3 Moving Block.

### 3.5 Data Factory

Das Konzept einer (potenziell paneuropäischen) Rail Data Factory [6, 7] als gemeinsame Infrastruktur von Bahnen, Herstellern, Behörden und anderen wird als Wegbereiter für die Erfassung und Verarbeitung großer Mengen von Zugsensordaten gesehen. Dieses soll durch simulierte Daten für die Schulung und Zertifizierung von KI-Modellen für den GoA 4-Betrieb ergänzt werden.

### 3.6 Herausforderungen im Zusammenhang mit der Datenvielfalt

Die Handhabung derart vielfältiger Daten stellt eine erhebliche Herausforderung dar. Insbesondere erfordert die Integration der Daten erweiterte Verarbeitungsmöglichkeiten, um plattformübergreifende Kompatibilität, Genauigkeit und Konsistenz sicherzustellen. Zudem erfordern Unterschiede in den Datenqualitätsstandards, Formaten und Aktualisierungszyklen zwischen den Systemen ausgefeilte Datenverwaltungsprotokolle und Transformationsprozesse. Mit zunehmendem Datenfluss aus mehreren Quellen nehmen auch die Risiken für Datensicherheit und Datenschutz zu, was robuste Schutzmaßnahmen erfordert. Die effektive Bewältigung dieser Komplexität ist entscheidend für die Schaffung einer einheitlichen und zuverlässigen Datenumgebung, die eine genaue Entscheidungsfindung unterstützt und Echtzeitabläufe ermöglicht.

## 4 Datenvielfalt durch Standardisierung managen

Standardisierung im Datenmanagement bezieht sich auf die Erstellung einheitlicher Datendefinitionen, -formate und -protokolle über alle Systeme hinweg, um Interoperabilität und Transparenz zu gewährleisten. Im Kontext der Datenvielfalt dient Standardisierung insbesondere als Grundlage für die Harmonisierung von Eingaben aus unterschiedlichen Quellen. Sie vereinfacht die Integration, unterstützt Kollaboration und ermöglicht fundierte, datengesteuerte Entscheidungen, die für die digitale Transformation im Bahnsystem unerlässlich sind. Die Vorteile der Standardisierung, wie auch in Bild 3 dargestellt, sind:

dational input, annotated data provides the contextual labeling essential for supervised learning and simulated data enables model testing under controlled virtual conditions.

### 3.3 Trained AI models

AI models are trained to reliably identify tracks, catenaries, bridges and other objects. However, trains operating across multiple countries must be equipped with AI models specifically tailored to accurately recognise objects within different regions as the characteristics of these objects may vary between countries. Further novel data systems are being established:

### 3.4 The Digital Register

The Digital Register (DR) [5], as standardised in Europe's Rail, is a repository for the infrastructure data that serves as a “single source of truth” (SSOT) for various rail subsystems. It is an essential component for highly to fully automated rail operations (GoA 2 – GoA 4) and for ETCS level 3 moving block.

### 3.5 The Data Factory

The concept of a (potentially pan-European) Rail Data Factory [6, 7] as a joint infrastructure among railways, suppliers, authorisation bodies and others is seen as a precursor for the collection and processing of large amounts of train sensor data complemented with simulated data for the training and certification of AI models for GoA 4 operations.

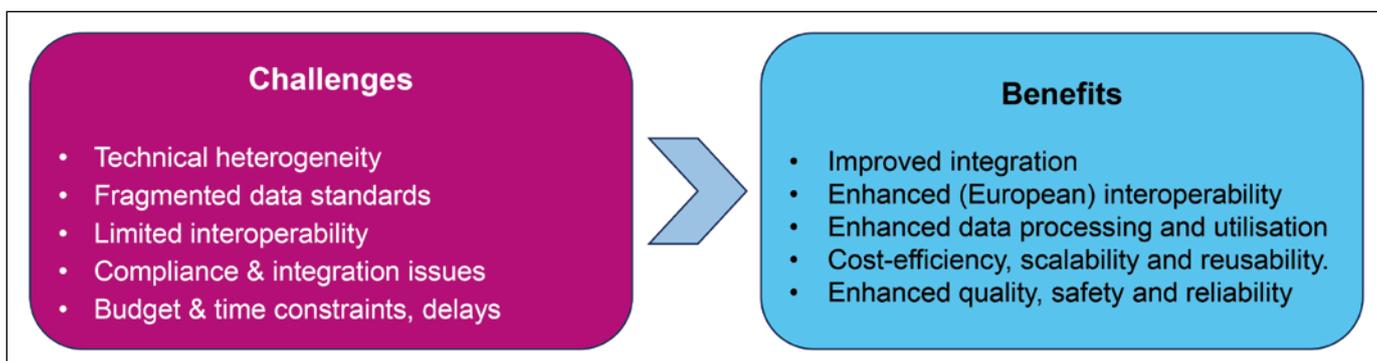
### 3.6 Challenges related to data diversity

Managing such diverse data sources introduces significant challenges. In particular, the integration of this data variety requires advanced processing capabilities so as to ensure compatibility, accuracy and consistency across platforms. Additionally, differences in data quality standards, formats and update cycles across systems call for sophisticated data management protocols and transformation processes. Data security and privacy risks will also escalate with the increasing data flow from multiple sources, thereby necessitating robust protective measures. Effectively addressing these complexities is crucial to creating a unified and reliable data environment that supports accurate decision-making and enables seamless real-time operations.

## 4 Managing data diversity through standardisation

Standardisation in data management refers to the creation of uniform data definitions, formats and protocols across systems so as to ensure interoperability and clarity. Within the context of data diversity, standardisation specifically acts as the foundation for harmonising inputs from various sources. It simplifies integration, supports cross-department collaboration and enables informed, data-driven decisions that are essential for digital transformation in railway networks. The benefits of standardisation, as also depicted in fig. 3, are:

- Interoperability across borders. For instance, the infrastructure data provided for automated railway operation in the levels GoA 2 to GoA 4 by the DR must be made equally available to all railway undertakings. It is hence necessary to standardise the underlying data model and interfaces.
- Enhanced data processing and utilisation. Standards enable platforms like the Data Factory that build upon the broad exchange and joint use of railway data across Europe, for instance for training AI models for automated rail operations.



**Bild 3: Vorteile der Standardisierung von Datenmanagement**

Fig. 3: The benefits of standardisation in data management

- Grenzüberschreitende Interoperabilität: So müssen die vom DR [5] für den automatisierten Bahnbetrieb (GoA 2 bis GoA 4) bereitgestellten Infrastrukturdaten allen Bahnunternehmen gleichermaßen zur Verfügung gestellt werden. Entsprechend ist es notwendig, das zugrunde liegende Datenmodell und die Schnittstellen zu standardisieren.
- Verbesserte Datenverarbeitung und -nutzung: Standards ermöglichen Plattformen wie die Data Factory [6], die auf einem breiten Austausch und einer gemeinsamen Nutzung von Bahn- daten in ganz Europa aufbaut, z. B. zum Trainieren von KI-Modellen für den automatisierten Bahnbetrieb.
- Kosteneffizienz und Ressourcenoptimierung: Standards reduzieren den Bedarf an projektspezifischen Lösungen und Nacharbeiten und gewährleisten eine kosten- und ressourceneffiziente Produktentwicklung im Bahnsektor.
- Datenqualität und -sicherheit: Systeme wie das DR [5] halten hohe Standards der Datenintegrität ein und gewährleisten Sicherheit und Zuverlässigkeit im Bahnsystem. Diese Systeme liefern maßgeschneiderte Daten, einschließlich wichtiger 3D-Topologieinformationen, um betriebliche Erkenntnisse zu verbessern und den Wartungsaufwand zu reduzieren.

#### 4.1 Standardisierte Datenontologien im Eisenbahnbereich

Ein wesentlicher Aspekt der Standardisierung des Datenmanagements ist die Einführung von Datenontologien: Dies sind strukturierte, maschinenlesbare Frameworks, die Datenkategorien und -beziehungen definieren und ein gemeinsames Vokabular und Regeln für die Organisation bereitstellen. Im Gegensatz zu herkömmlichen Methoden wie Datenmodellen, Anwendungsprogrammierschnittstellen (API) und kontrollierten Vokabularen, denen es oft an semantischem Verständnis und Anpassungsfähigkeit mangelt, erhöhen Ontologien die Flexibilität, gewährleisten eine einheitliche Terminologie, ermöglichen logische Schlussfolgerungen und vereinfachen die Handhabung unterschiedlicher Datensätze. Sie spielen eine Schlüsselrolle bei

- der Gewährleistung semantischer Interoperabilität, insbesondere bei der Verwaltung komplexer und unterschiedlicher Datensätze,
- der Ermöglichung eines nahtlosen Datenaustauschs und einer systemübergreifenden Integration,
- der Verbesserung der Datenqualität und -konsistenz, z. B. durch die Definition von Validierungsregeln,
- der Ermöglichung verbesserter Entscheidungsfindung, z. B. durch erweiterte Analysen basierend auf einer strukturierten und semantisch reichen Datendarstellung, und
- der Ermöglichung von Innovation, z. B. durch Methoden zum Suchen und Nutzen von Wissen.

- Cost efficiency and resource optimisation. Standards reduce the need for custom solutions and reworking, thereby ensuring cost and resource efficient product development in the rail sector.
- Data quality and safety. Systems like the DR [5] maintain ambitious data quality standards, thereby ensuring safety and reliability in rail operations. These systems provide tailored data, including critical 3D topology information, to enhance operating insights and reduce maintenance overheads.

#### 4.1 Standardised data ontologies in the railway domain

Standardising data management involves introducing data ontologies, i.e. structured, machine-readable frameworks that define data categories and relationships and provide a shared vocabulary and rules for organisation. Unlike conventional methods like data models, Application Programming Interfaces (APIs), and controlled vocabularies, which often lack semantic understanding and adaptability, ontologies enhance flexibility, ensure consistent terminology, allow for semantic reasoning and simplify the handling of diverse datasets. They play a key role in

- ensuring semantic interoperability, especially in managing complex and diverse datasets,
- enabling seamless data exchange and cross-system integration,
- improving data quality and consistency, for instance through the definition of validation rules,
- enabling improved decision making, for instance through advanced analytics based on a structured and semantically rich representation of the data, and
- enabling innovation, for instance through means of discovering and reusing knowledge.

Data ontologies are further essential for formalised knowledge representation, serving as a foundation for knowledge graphs. They play a critical role in AI and machine learning (ML) systems by enhancing reasoning, learning and data organisation and enabling more precise data annotation. Initiatives such as Shift2Rail, LinX4Rail and UIC's OntoRail highlight how ontologies enable seamless data exchange, improved decision-making and innovation in the rail system.

The following are key ontologies that have been defined:

- The ERA Ontology [8] provides a structured framework for terms, concepts and relationships in the railway domain. It ensures interoperability and consistency across railway systems and facilitates digital transformation in the rail system. It is aligned with EU railway directives and supports data integration and knowledge management.

Datenontologien sind außerdem für eine formalisierte Wissensdarstellung unerlässlich und dienen als Grundlage für Wissensgraphen. Sie spielen eine entscheidende Rolle in der KI und bei maschinellem Lernen (ML), indem sie das logische Schlussfolgern, Lernen und die Datenorganisation verbessern und eine präzise Datenannotation ermöglichen. Initiativen wie Shift2Rail, LinX-4Rail und OntoRail der UIC zeigen, wie Ontologien einen nahtlosen Datenaustausch, eine verbesserte Entscheidungsfindung und Innovation im Bahnsystem ermöglichen.

Wesentliche, für das Bahnumfeld relevante Ontologien sind:

- Die ERA-Ontologie [8] bietet einen strukturierten Rahmen für Begriffe, Konzepte und Beziehungen im Eisenbahnbereich. Sie gewährleistet Interoperabilität und Konsistenz zwischen Bahnsystemen und erleichtert die digitale Transformation im Bahnsystem. Sie ist auf EU-Richtlinien abgestimmt und unterstützt die Datenintegration und das Wissensmanagement.
- Die Ontologie des W3C Semantic Sensor Network (SSN) [9] bietet einen Rahmen für die Beschreibung von Sensoren, deren Beobachtungen und der Systeme, zu denen sie gehören, und unterstützt das semantische Verständnis in Sensornetzwerken. SOSA [10], als Teilmenge von SSN, konzentriert sich auf Kernkonzepte im Zusammenhang mit Sensoren, Beobachtungen und Akteuren und erleichtert damit die Integration in breitere Anwendungen.
- Die Sensor Model Language (SensorML) [11] ist ein XML-basierter Standard, der vom Open Geospatial Consortium (OGC) entwickelt wurde. Er wird verwendet, um Sensoren und Messprozesse einschließlich ihrer Konfigurationen, Operationen und Ausgaben zu modellieren und zu beschreiben. Er wird häufig in der Umweltüberwachung, in autonomen Systemen und im Internet of Things (IoT) verwendet.
- Die Railway Markup Language (RailML) [12] ist ein standardisiertes XML-Schema, das für den Datenaustausch im Eisenbahnsektor entwickelt wurde. Es enthält Ontologieprinzipien, um Beziehungen zwischen Eisenbahninfrastruktur, Fahrzeugen und Fahrplan zu definieren und so einen nahtlosen Datenaustausch zwischen verschiedenen Systemen zu ermöglichen.

Diese Standards bilden nicht nur eine solide Grundlage für den Bahnsektor, es sind auch verschiedene Erweiterungen denkbar. So erscheint es z. B. sinnvoll, die ERA-Ontologie um Sensoraspekte zu erweitern, wie etwa eine Zugsensor-Ontologie, die im Rahmen des Projekts AutomatedTrain [3] entwickelt wurde. Auch domänenspezifische Unterstützung wäre hilfreich, etwa in Bezug auf das Robot Operating System (ROS), das häufig im Automatisierungskontext angesprochen wird. Schließlich wird ein großer Wert in der Entwicklung eines harmonisierten Konzepts für Bahnplanungsdaten gesehen, wie es im Europe's Rail FP1-MOTIONAL Projekt [13] für Digital Asset Engineering verfolgt wird. Der Fokus liegt dabei auf der Einführung eines standardisierten digitalen Formats für effiziente Planungs- und Managementprozesse in der LST.

## 5 Die Rolle der Datengovernance im digitalisierten Bahnsystem

Wie erwähnt, ist die Datenvielfalt im digitalisierten Bahnsystem enorm und bietet viele Möglichkeiten für datenbasierte Entscheidungen. Gleichzeitig kann es, wenn diese Daten nicht auf standardisierte Weise verwaltet und geregelt werden, zu Verstößen gegen Datenkonformität und Datenschutzbestimmungen kommen, die zu Geldbußen, Vertrauensverlust und Reputationsschäden der beteiligten Entitäten führen können.

- The W3C Semantic Sensor Network (SSN) [9] ontology provides a framework for describing sensors, their observations and the systems they are part of, thereby supporting semantic understanding in sensor networks. As a lightweight subset of SSN, SOSA [10] focuses on core concepts related to sensors, observations, and actuators, thereby making it easier to integrate with broader applications.
- The Sensor Model Language (SensorML) [11], is an XML-based standard developed by the Open Geospatial Consortium (OGC). It is used to model and describe sensors and measurement processes, including their configurations, operations and outputs. It is widely used in environmental monitoring, autonomous systems and Internet of Things (IoT).
- The Railway Markup Language (RailML) [12] is a standardised XML scheme developed for data exchange within the railway sector. It incorporates ontological principles to define the relationships between railway infrastructure, rolling stock, and timetabling, thereby enabling seamless data exchange between different systems.

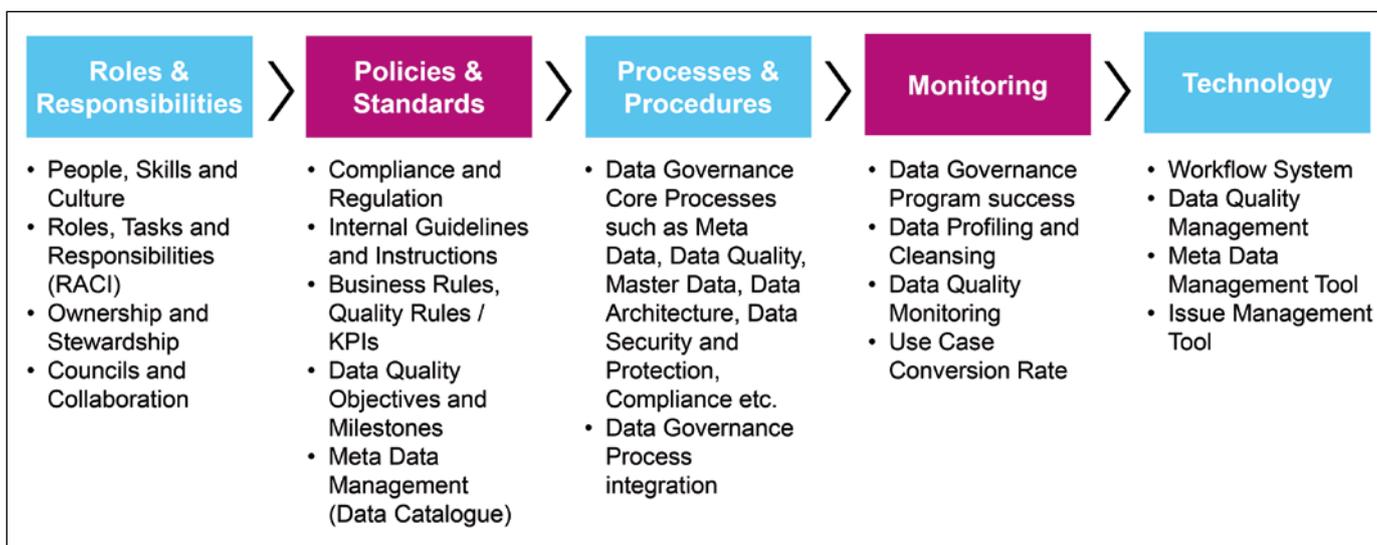
While these standards provide a robust foundation for the rail sector, various extensions are also conceivable. For instance, it appears to make sense to extend the ERA ontology to cover sensing aspects, such as a train sensor ontology that has been developed within the context of the AutomatedTrain project. Domain specific support would also be helpful, for instance in relation to the Robot Operating System (ROS) often used in automation contexts. Finally, significant value is seen in establishing a harmonised concept for railway planning data, as pursued in Europe's Rail FP1-MOTIONAL project [13] for Digital Asset Engineering. The focus is on introducing a standardised digital format for efficient planning and management processes in CCS deployments.

### 5 The role of data governance in the digitalised rail system

As pointed out, the data diversity in the digitalised rail system is vast and as such provides many opportunities for data-driven decision-making. At the same time, if this data is not managed and governed in a standardised way, it can lead to violations of data compliance and privacy regulations that can result in fines, a lack of trust and a bad reputation for the entities involved.

Data governance is an organisational framework and set of practices that ensure high data quality, data management consistency and compliance with data related policies. It involves defining responsibilities, establishing processes and implementing policies to guarantee that the data is accurate, available and secure and used effectively across the entire organisation. Fig. 4 outlines the key components that data governance entails.

Data governance helps solve the issues that entities collecting and generating a lot of diverse data typically face. One of these challenges concerns the fact that the data is often very siloed and fragmented, which makes it difficult to have a unified and holistic view. A governance framework helps facilitate standardisation and cross-collaboration between teams as it assigns ownership and accountability to data producers and data consumers. Having clear data ownership is also the starting point for ensuring that metadata is easily discoverable, accessible and manageable, which helps foster a data-driven culture. Successful data governance can also help data producers and consumers arrive at similar definitions of how data is interpreted and at an SSOT. Further benefits of effective data governance include:



**Bild 4: Wesentliche Elemente einer Datengovernance**

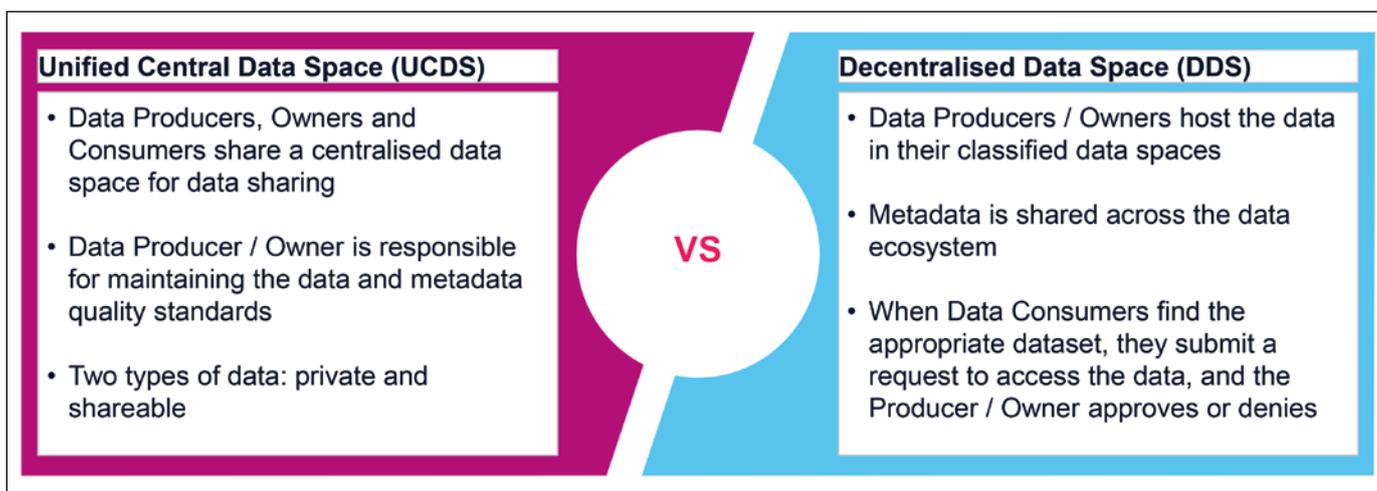
Fig. 4: The key data governance components

Datengovernance ist ein organisatorischer Rahmen und eine Reihe von Praktiken, die eine hohe Datenqualität, Konsistenz der Datenverwaltung und Einhaltung datenbezogener Richtlinien gewährleisten. Dazu gehört die Definition von Verantwortlichkeiten, die Einrichtung von Prozessen und die Umsetzung von Richtlinien, um zu gewährleisten, dass die Daten in der gesamten Organisation korrekt, verfügbar, sicher und effektiv genutzt werden. Bild 4 zeigt die wesentlichen Säulen der Datengovernance.

Datengovernance hilft, die Probleme zu lösen, mit denen Entitäten, die viele unterschiedliche Daten sammeln und generieren, normalerweise konfrontiert sind. Eine dieser Herausforderungen besteht darin, dass die Daten oft sehr isoliert und fragmentiert sind, was eine einheitliche und ganzheitliche Sicht erschwert. Ein Governance-Framework erleichtert die Standardisierung und teamübergreifende Zusammenarbeit, da es Datenproduzenten und Datenkonsumenten Eigentum und Verantwortung zuweist. Ein klar definiertes Dateneigentum ist auch die Grundlage dafür, dass Metadaten leicht auffindbar, zugänglich und verwaltbar sind, was zur Förderung einer datengetriebenen Kultur beiträgt. Eine erfolgreiche Datengovernance kann Da-

- Data Quality: it ensures that the data is accurate, complete and reliable, thereby fostering trust in the decision-making processes.
- Compliance: it helps organisations adhere to any regulatory requirements, thereby avoiding the legal and financial risks associated with data mishandling.
- Decision-Making: it facilitates informed decision-making by providing reliable and consistent data across departments.
- Efficiency: it streamlines data-related processes, thereby reducing any redundancies and improving overall operating efficiency.

A detailed analysis of the possible unified (or centralised) and decentralised governance approaches for the development of fully-automated rail operations has been conducted, as depicted in fig. 5. The unified approach uses a common dataspace to govern the data. The decentralised approach has also been adopted in Europe's Rail FP1-MOTIONAL Project [13], thereby aligning it with the principles of data sovereignty. This approach facilitates seamless data exchange among European railway part-



**Bild 5: Zentralisierte und dezentralisierte Ansätze zu Datenräumen**

Fig. 5: Centralised vs decentralised dataspace approaches

tenproduzenten und -nutzern auch dabei helfen, zu ähnlichen Definitionen der Dateninterpretation und zu einer SSOT zu gelangen.

Weitere Vorteile einer effektiven Datengovernance sind:

- **Datenqualität:** Sie stellt sicher, dass die Daten genau, vollständig und zuverlässig sind, und fördert das Vertrauen in Entscheidungsprozesse.
- **Compliance:** Sie hilft Organisationen, gesetzliche Anforderungen einzuhalten und rechtliche und finanzielle Risiken im Zusammenhang mit dem falschen Umgang mit Daten zu vermeiden.
- **Entscheidungsfindung:** Sie erleichtert fundierte Entscheidungen, indem sie abteilungsübergreifend zuverlässige und konsistente Daten bereitstellt.
- **Effizienz:** Sie rationalisiert datenbezogene Prozesse, reduziert Redundanzen und verbessert die allgemeine Betriebseffizienz.

Für die Entwicklung eines vollautomatisierten Bahnbetriebs wurde eine detaillierte Analyse möglicher einheitlicher (bzw. zentralisierter) und dezentraler Governance-Ansätze durchgeführt, wie in Bild 5 dargestellt. Beim einheitlichen Ansatz wird ein gemeinsamer Datenraum zur Verwaltung der Daten verwendet. Der dezentrale Ansatz wurde auch im Europe's Rail FP1-MOTIONAL Projekt [13] übernommen und entspricht den Grundsätzen der Datensouveränität. Dieser Ansatz erleichtert den nahtlosen Datenaustausch und fördert Interoperabilität und Zusammenarbeit unter den europäischen Bahnen.

## 6 Schlussfolgerung

Für die Digitalisierung des Bahnsystems ist ein umfassendes Verständnis der Daten, die die Funktionsweise und Entscheidungsfindung im Bahnbetrieb beeinflussen, von entscheidender Bedeutung. Wie in diesem Beitrag erläutert, ist insbesondere die Standardisierung von Datenontologien und die Einrichtung einer robusten Datengovernance für die konsistente und integrierte Handhabung unterschiedlicher Datensätze von entscheidender Bedeutung. Durch die Identifizierung und Kategorisierung dieser kritischen Daten können Unternehmen ihre Verwaltungspraktiken rationalisieren, die Komplexität verringern und die Gesamteffizienz steigern. Diese Maßnahmen sind essenziell für die Erreichung der ehrgeizigen Digitalisierungsziele im europäischen Bahnsektor. ■

## LITERATUR | LITERATURE

- [1] Digitale Schiene Deutschland, see <https://digitale-schiene-deutschland.de>
- [2] Europe's Rail FP2-R2DATO project, see <https://projects.rail-research.europa.eu/eurail-fp2/>
- [3] AutomatedTrain project, see <https://digitale-schiene-deutschland.de/en/projects/AutomatedTrain>
- [4] Sensors4Rail project, see <https://digitale-schiene-deutschland.de/en/Sensors4Rail>
- [5] Kreyenberg, D. et al.: "Digital Register – the "Single Source of Truth" for the infrastructure data used in fully automated driving", SIGNAL+DRAHT 6/2024
- [6] Data Factory, see <https://digitale-schiene-deutschland.de/en/news/2024/Pan-European-Railway-Data-Factory>
- [7] Pan-European Rail Data Factory, see <https://digitale-schiene-deutschland.de/en/news/2024/Pan-European-Railway-Data-Factory>
- [8] ERA Ontology, see <https://data-interop.era.europa.eu/era-vocabulary/>
- [9] W3C, see <https://www.w3.org/>
- [10] OpenGroup SOSA, see <https://www.opengroup.org/sosa>
- [11] OGC SensorML, see <https://www.ogc.org/publications/standard/sensorml/>
- [12] RailML, see <https://www.railml.org/en/>
- [13] Europe's Rail FP1-MOTIONAL project, see <https://projects.rail-research.europa.eu/eurail-fp1/>

ners, promoting interoperability and collaboration across the network.

## 6 Conclusions

A comprehensive understanding of the data that influences the functioning and decision-making within the rail system is crucial for advancing digitalisation initiatives in the sector. As detailed in this article, the standardisation of data ontologies and the establishment of robust data governance frameworks in particular are essential for managing diverse datasets in a consistent and integrated manner. By identifying and categorising such critical data, organisations can streamline management practices, reduce complexity and enhance overall efficiency. These measures are fundamental to achieving the ambitious digitalisation goals set in the European rail sector. ■

## AUTOREN | AUTHORS

### Waseem ul Aslam Peer

Lead Data Strategy, Governance and Architecture Circle  
Digitale Schiene Deutschland  
DB InfraGO AG  
Adresse / Address: EUREF-Campus 3, D-10829 Berlin  
E-Mail: waseem-ul-aslam.peer@deutschebahn.com

### Imran Mehaboob Hussain

MBSE & Data Intern  
Digitale Schiene Deutschland  
DB InfraGO AG  
Adresse / Address: EUREF-Campus 3, D-10829 Berlin  
E-Mail: imran-mehaboob.hussain@deutschebahn.com

### Anna Bardakh

Data Governance Architect  
Digitale Schiene Deutschland  
DB InfraGO AG  
Adresse / Address: EUREF-Campus 3, D-10829 Berlin  
E-Mail: anna.bardakh@deutschebahn.com

### Dr. Patrick Marsch

Head of Connectivity, IT and Data Platforms and IT/OT Security, DSD  
Digitale Schiene Deutschland  
DB InfraGO AG  
Adresse / Address: EUREF-Campus 3, D-10829 Berlin  
E-Mail: patrick.marsch@deutschebahn.com

### Bernardo B Ramalho Immendorff

Data Management Expert  
DB Systel GmbH  
Adresse / Address: Jürgen-Ponto-Platz 1, D-60329 Frankfurt a.M.  
E-Mail: bernardo.b.ramalho-immendorff@deutschebahn.com

### Daniel Obermeyer

Senior Business Engineer  
DB Systel GmbH  
Adresse / Address: Jürgen-Ponto-Platz 1, D-60329 Frankfurt a.M.  
E-Mail: daniel.obermeyer@deutschebahn.com