# Rail to Digital automated up to autonomous train operation

## D7.4 – Complete Data Annotation

Due date of deliverable: 31/05/2025

Actual submission date: 30/05/2025

Leader/Responsible of this Deliverable: Dr. Philipp Neumaier, DB InfraGO AG

Reviewed: Y/N

| Document status | | |
|---|---|---|
| Revision | Date | Description |
| 01 | 30/05/2025 | First issue |
| 02 | 30/05/2025 | Corrected Protocol Number in Footer |
| 03 | 30/05/2025 | Final version, agreed on with the WP, changed Document Filename due to CT5 Upload error, New Protocol Number in Footer |
| | | |

| Project funded from the European Union's Horizon Europe research and innovation programme | | |
|---|---|---|
| **Dissemination Level** | | |
| **PU** | Public | PU |
| **SEN** | Sensitiv – limited under the conditions of the Grant Agreement | |

Start date: 01/12/2022

Duration: 42 months

## ACKNOWLEDGEMENTS

This project has received funding from the Europe's Rail Joint Undertaking (ERJU) under the Grant Agreement no. 101102001. The JU receives support from the European Union's Horizon Europe research and innovation programme and the Europe's Rail JU members other than the Union.

## REPORT CONTRIBUTORS

| Name | Company | Details of Contribution |
|---|---|---|
| Dr. Philipp Neumaier | DB | Content creator |
| Dr. Martin Köppel | DB | Content creator |
| Sebastian Dubiel | DB | Content creator |
| Philippe David | SNCF | Content creator |
| Philippe Bourdenet | SNCF | Content creator |
| Dr. Cecrick Lelionnais | SNCF | Content creator |
| Bastian Simoni | Alstom | Expert Review |
| Dominik Kevicky | AZD | Expert Review |
| Frederic Antoine | ATSA | Expert Review |
| Hardik Jain | THD/Thales | Expert Review |
| Hájek Jiří | AZD | Expert Review |
| Hjalmar Boulouednine | Siemens | Expert Review |
| Lars-Kristian Vognild | NRD | Expert Review |
| Maik Baehr | Siemens | Expert Review |
| Michal Novak | AZD | Expert Review |
| Michele Bruzzo | Hitachi | Expert Review |
| Oliver Lehmann | SMO | Expert Review |
| Saro Thiyagarajan | FT/Wabtec | Expert Review |
| Sebastiaan Linssen | NS | Expert Review |
| Tom Jansen | NS | Expert Review |

**Disclaimer**

## EXECUTIVE SUMMARY

The rail sector across Europe is facing increasing demands for automation, safety, and operational efficiency. Central to achieving these objectives is the development of perception systems that can reliably interpret complex railway environments under diverse and often challenging conditions. However, the advancement of such systems is constrained by the lack of high-quality, representative, and standardised datasets required to train, test, and validate AI models.

**Deliverable 7.4**, part of the R2DATO project's Work Package 7 "Data Factory", addresses this challenge by establishing a methodology and executing a joint campaign for the collection and annotation of real-world sensor data from operational rail environments. The ultimate objective is to build a robust dataset foundation that supports the development of sensor-based perception models (as foreseen in D7.5) and contributes to the creation of a publicly accessible dataset (under D7.6).

### Context and Objectives

The primary aim of D7.4 is to enable the training and benchmarking of AI perception systems by providing a diverse, well-structured, and high-fidelity dataset of sensor observations collected under real railway conditions. The deliverable responds to critical gaps in the availability of such data, especially those encompassing edge cases like low visibility, occlusion, or weather artefacts—conditions frequently encountered in real-life rail operations but underrepresented in existing datasets.

In line with European digitalisation strategies and automation goals, D7.4 provides a necessary resource to accelerate the development of trustworthy and resilient railway AI systems. It also contributes to the broader effort of harmonising data formats and standards, particularly by adopting the ASAM Open Label format for annotation consistency.

### Scientific and Technical Approach

The work carried out under D7.4 represents a collaborative effort between Deutsche Bahn and SNCF, supported by additional stakeholders from the R2DATO consortium. Between 2022 and 2024, both partners conducted targeted field recordings using multi-modal sensor configurations (RGB, infrared, LiDAR), mounted on operational trains.

Scene selection followed a strategic approach: DB focused on infrastructural and behavioural relevance (e.g. station activity, platform crowding, night scenes), while SNCF contributed sequences reflecting adverse environmental effects, such as raindrops on lenses and visibility impairments. Together, these scenes provide a rich set of perceptual challenges and contextual diversity.

Data annotation was carried out using a combination of manual and semi-automated tools. The process emphasised cross-modal alignment, temporal consistency, and class coherence across 2D and 3D geometries. Quality assurance mechanisms, including key performance indicators and validation routines, were implemented to ensure data integrity. All datasets are integrated within the R2DATO Data Factory infrastructure for downstream usability.

### Main Findings and Added Value

Deliverable 7.4 produced a set of 8 thoroughly annotated scenes comprising more than 200 frames across varying environments. Key achievements include:

- High-quality multi-sensor annotations, covering over 15 object classes;
- Detailed geometric and semantic labeling, consistent across RGB, infrared, and LiDAR;
- Explicit treatment of adverse weather conditions, contributing to model robustness;
- Establishment of a reusable annotation standard (ASAM Open Label);
- Cross-partner interoperability through harmonised practices and formats.

The work represents a significant step forward in data-driven railway automation and provides EU-level added value by creating an interoperable resource that can serve future research and industrial deployment.

**Outlook and Recommendations**

While the deliverable fulfills its core objectives, further work is recommended to increase the geographical and operational diversity of the dataset. Scaling up data acquisition, increasing automation in annotation workflows, and extending coverage to underrepresented scenarios (e.g. rural single-track routes, high-speed segments) will enhance applicability.

Moreover, the annotated dataset developed under D7.4 is poised to directly support the training of perception models in D7.5 and forms a crucial component of the open-access dataset to be delivered under D7.6. These subsequent steps will strengthen the project's contribution to future standardisation efforts and foster greater market acceptance of AI-enabled railway technologies.

In conclusion, D7.4 delivers a scientifically grounded, operationally relevant dataset that addresses critical data gaps in the railway automation domain, laying the foundation for resilient, scalable, and interoperable AI systems in European rail transport.

# ABBREVIATIONS AND ACRONYMS

| Abbreviation | Full Name / Meaning |
|---|---|
| 2D | Two-Dimensional |
| 3D | Three-Dimensional |
| ADIF | Adminstrador de Infraestructuras Ferroviarias |
| AI | Artificial Intelligence |
| API | Application Programming Interface |
| ASAM | Association for Standardization of Automation and Measuring Systems |
| ATO | Automatic Train Operation |
| ATSA | Alstom Transport SA |
| AZD | AŽD Praha s.r.o. |
| BAT | Bounding Box Annotation Tool |
| CVAT | Computer Vision Annotation Tool |
| DB | Deutsche Bahn |
| DBSCAN | Density-Based Spatial Clustering of Applications with Noise |
| Dx.x | Deliverable x.x |
| e.V. | eingetragener Verein (German: registered association) |
| ERJU | Europe's Rail Joint Undertaking |
| FT | Faiveley Transport (likely, as part of Wabtec) |
| GDPR | General Data Protection Regulation |
| GNSS | Global Navigation Satellite System |
| GPS | Global Positioning System |
| GTSD | Not explicitly defined |
| HIL | Hardware in the Loop |
| HPC | High-Performance Computing |
| Hz | Hertz |
| IMU | Inertial Measurement Unit |
| IoU | Intersection over Union |
| IR | Infrared |
| JSON | JavaScript Object Notation |
| JU | Joint Undertaking |
| KPIs | Key Performance Indicators |
| LAN | Local Area Network |
| LiDAR | Light Detection and Ranging |
| ML | Machine Learning |
| NRD | Norwegian Railway Directorate |
| NS | Nederlandse Spoorwegen |
| PCD | Point Cloud Data |
| PEDF | Pan-European Data Factory |
| PNG | Portable Network Graphics |
| PTP | Precision Time Protocol |
| PU | Public |
| px | Pixels |

| | |
|---|---|
| R2DATO | Rail to Digital automated up to autonomous train operation |
| RGB | Red Green Blue |
| SASP | System Access & Security Platform |
| SEN | Sensitive |
| SMO | Siemens Mobility GmbH |
| SNCF | Société Nationale des Chemins de fer Français |
| SSD | Solid-State Drive |
| THD | Thales (division within Thales Group) |
| UI | User Interface |
| UUIDv4 | Universally Unique Identifier version 4 |
| VDL | Vehicle Data Logger |
| WP7 | Work Package 7 |
| YOLOv8 | You Only Look Once version 8 |

# TABLE OF CONTENTS


Acknowledgements ..... 2
Report Contributors ..... 2
Executive Summary ..... 4
Abbreviations and Acronyms ..... 6
Table of Contents ..... 8
List of Figures ..... 11
List of Tables ..... 11
1. Introduction ..... 12
1.2 Purpose of the Document ..... 12
1.2.1 Scope of the Deliverable ..... 12
1.3 Relation to Other Deliverables (D7.1, D7.5) ..... 13
2 Project Overview ..... 14
2.1 Background Information and Relevance to the Data Factory ..... 14
2.2 Objectives of D7.4 ..... 14
2.3 Stakeholders and Responsibilities ..... 15
3 Data Collection and Sensor Setup ..... 16
3.1 Overview of Collected Data ..... 16
3.1.1 Rationale for Scene Selection ..... 17
3.2 Object Classes Overview ..... 17
3.2.1 General Requirements ..... 18
3.2.2 Dynamic Objects ..... 18
3.2.3 Static Objects ..... 19
3.2.4 Special Objects ..... 19
3.3 Overview Condition Classes ..... 20
3.4 Sensor Types and Their Configurations ..... 22
3.5 Data Acquisition Workflow ..... 23
3.5.1 Real-World Data Collection ..... 25
3.5.2 Challenges in Data Collection (Technical, Logistical) ..... 25
4 Data Annotation Framework (DB) ..... 27
4.1 Data Annotation Types (Manual vs. Automated) ..... 27
4.1.1 Manual Annotation ..... 27
4.1.2 Automated Annotation ..... 28
4.1.3 Integration and Workflow Management ..... 28
4.1.4 Manual vs. Automated Annotation – A Comparative Overview ..... 29
4.1.5 A Hybrid Strategy for Quality and Scale ..... 29

5 Annotation Requirements and Standards ..... 30
5.1 Alignment with ASAM Open Label Format (Simulated Data) ..... 30
5.1.1 RailLabel Format for Sensor Data Annotation ..... 30
5.2 Synchronization ..... 32
5.3 Unique Object Identification and Labeling ..... 33
5.4 Attributes and Metadata in Annotations ..... 33
5.5 Legal and Ethical Considerations ..... 34
5.5.1 Legal Aspects ..... 34
5.5.2 Ethical Aspects ..... 35
6 Data Storage and Processing ..... 36
6.1 Data Handling and Storage Infrastructure ..... 36
6.2 Data Touchpoint Functions ..... 37
6.3 Integration with the Data Factory Platform ..... 37
6.4 Data Transfer Protocols and Security Measures ..... 37
6.5 Dataset Management for Open Access ..... 38
6.6 Summary ..... 38
7 Evaluation of Data Annotation Quality ..... 39
7.1 Establishing Metrics for Annotation Quality ..... 39
7.2 Validation Methodologies and Annotation Quality Control Workflow ..... 40
7.2.1 Heuristic and Rule-Based Checks ..... 40
7.2.2 Inter-Modal Consistency Validation ..... 40
7.2.3 Temporal Coherence Checks (Sequential Data) ..... 41
7.2.4 Self-Consistency and Annotation Correlation Analysis ..... 41
7.3 Comparative Analysis: Simulated vs. Real Data ..... 41
7.4 Error Detection and Correction Mechanisms ..... 41
7.5 Idea for a validation process ..... 42
7.5.1 Multi-Stage Validation Pipeline ..... 42
7.6 Implementation Considerations ..... 45
7.6.1 Data Format Standardisation ..... 45
7.6.2 Modular Validation Pipeline ..... 45
7.6.3 Toolchain and Environment Integration ..... 46
7.6.4 Scalability and Performance ..... 46
7.6.5 Quality Thresholds and Decision Rules ..... 46
7.6.6 Governance and Human Oversight ..... 46
8 Results and Discussion ..... 48
8.1 Overview of the Annotated Dataset ..... 48

8.1.1 Scene Profile: Trackside Passengers ........................................ 49
8.1.2 Scene Profile: Night Station Scene ........................................ 51
8.1.3 Scene Profile: Crowd at Platform ........................................ 53
8.1.4 Scene Profile: Incoming Train with Trackside Passengers ........................................ 55
8.1.5 Scene Profile: Platform with customers (SNCF) ........................................ 57
8.1.6 Scene Profile: Shunting yard arrival (SNCF) ........................................ 58
8.1.7 Scene Profile: Crossing train (SNCF) ........................................ 59
8.1.8 Scene Profile: Plain line (SNCF) ........................................ 60
8.2 Summary of Annotated Scenes and Selection Rationale ........................................ 60
8.2.1 Scenes Provided by Deutsche Bahn ........................................ 61
8.2.2 Scenes Provided by SNCF ........................................ 61
8.2.3 Summary ........................................ 61
9 Conclusions ........................................ 63
References ........................................ 66

## LIST OF FIGURES

Figure 1 General Workflow of the Data Factory ........ 23
Figure 2 Basic Components of the Uniform Toolchain ........ 24
Figure 3 Idea for a Validation Process (Proposed Multi-Stage Pipeline)........ 43
Figure 4 Annotated Scene: Morning Platform View ........ 49
Figure 5 Annotated Scene: Night Station Scene ........ 51
Figure 6 Annotated Scene: Crowd at Platform ........ 53
Figure 7 Annotated Scene: Incoming Train with Trackside Passengers ........ 55
Figure 8 Annotated Scene: Platform with Customers ........ 57
Figure 9 Annotated Scene: Shunting yard arrival ........ 58
Figure 10 Annotated Scene: crossing train........ 59
Figure 11 Annotated Scene: Plain line ........ 60

## LIST OF TABLES

Table 1 Selected Scenes of DB & SNCF ........ 16
Table 2 Scope of Dynamic Objects ........ 18
Table 3 Scope of Static Objects ........ 19
Table 4 Scope of Special Objects ........ 19
Table 5 Overview of possible Condition Classes........ 21
Table 6 Overview of used Sensor Types........ 22
Table 7 Manual vs. Automated Annotation – A Comparative Overview........ 29
Table 8 RailLabel Structure........ 31
Table 9 Attribute Schema for Person Entity in Annotation Framework ........ 34
Table 10 Overview annotated Scenes........ 48

# 1. INTRODUCTION

To effectively support the development of perception systems for autonomous rail transport, access to high-quality, annotated sensor data is essential. Chapter 1 introduces Deliverable 7.4 by framing its purpose, scope, and relevance within the R2DATO project. It outlines the specific objectives of the deliverable, defines its relation to other work packages, and positions it as a key enabler for subsequent tasks involving model training (D7.5) and open data dissemination (D7.6). In doing so, it sets the stage for the detailed technical and methodological content presented in the following chapters.

## 1.2 Purpose of the Document

This document provides a detailed account of the activities and results related to **Deliverable 7.4: Joint Collection and Annotation of Sensor Data** within the framework of the **R2DATO** project. The deliverable addresses the acquisition and annotation of **real-world sensor data**, serving as a foundational resource for the development of **advanced perception systems** in **autonomous railway operations**.

The specific objectives of this document are to:

- Describe the **methods** and **technical setups** employed for **sensor data acquisition**, including the configuration of **camera** and **LiDAR** systems
- Explain the **annotation procedures** and technologies used, with particular attention to the integration of **manual** and **automated approaches**
- Define the **annotation standards** and **requirements**, ensuring compatibility with the **ASAM Open Label format** and the technical specifications outlined in **Deliverable 7.1**
- Evaluate the **quality** and **reliability** of the annotated datasets, with emphasis on their applicability to **machine learning** in **railway scenarios**
- Discuss the **legal**, **logistical**, and **technical challenges** encountered throughout the collaborative data collection and annotation process
- Provide guidance on the application of the resulting dataset in subsequent project phases, especially in relation to the development of **perception models** (**Deliverable 7.5**) and the creation of a **public dataset** (**Deliverable 7.6**).

**Deliverable 7.4** constitutes a key milestone in ensuring the availability of a **high-quality**, **representative dataset** essential for advancing perception capabilities in the context of **automated train operation** across **European rail networks**.

### 1.2.1 Scope of the Deliverable

Deliverable 7.4 focuses on the following areas of work:

**a. Data Collection:**

This task involves the joint acquisition of **sensor data** from **real-world environments**, with a focus on **camera** and **LiDAR systems**. Data are collected from **instrumented trains** in operational settings. The resulting datasets support **perception-related machine learning tasks**.

**b. Data Annotation:**

The real-world sensor data are annotated using a combination of **manual** and **automated methods**. The annotations include **semantic labels**, **object IDs**, and **bounding boxes** in **2D** and **3D**, following the **ASAM Open Label format**. This ensures **consistency** and **reusability** across related deliverables.

**c. Quality Assurance:**

To ensure **data quality**, the annotation output is assessed using defined **key performance indicators (KPIs)**. Special attention is paid to **temporal** and **spatial synchronization** between sensors and to **consistency** between **2D** and **3D annotations**. **Validation routines** are established to detect and correct annotation errors.

**d. Integration with the Data Factory:**

The annotated datasets are integrated into the **infrastructure of the Data Factory** to support further processing and usage across work packages, including **model training** in **D7.5** and **dataset publication** in **D7.6**. Secure **data handling**, **storage**, and **transfer protocols** are part of the implementation scope.

**e. Collaboration and Coordination:**

Data collection and annotation activities are conducted jointly by **Deutsche Bahn** and **SNCF**. While each partner processes its respective datasets, **interoperability** and **comparability** are ensured through **cross-validation** and adherence to shared **annotation standards**.

**f. Deliverable Outcome:**

The final output is a **curated** and **validated dataset** comprising **camera** and **LiDAR data** with **high-quality annotations**. This dataset is designed to serve as a foundation for **machine learning model development** (**D7.5**) and for inclusion in the **publicly available dataset** envisaged in **D7.6**.

## 1.3 RELATION TO OTHER DELIVERABLES (D7.1, D7.5)

Deliverable 7.4 interfaces directly with several other components of **Work Package 7**, forming an integral part of the **R2DATO Data Factory pipeline**.

**D7.1 – Requirements and Specifications for the Data Factory**
D7.4 adheres to the specifications defined in **D7.1** regarding **data formats** and **quality benchmarks**, in particular the adoption of the **ASAM Open Label standard**. It builds on the infrastructure designed in D7.1 by supplying **standardized**, **high-quality annotated data**. [1]

**D7.5 – Development of Data-Factory toolchain and perform ML/AI model training**
Parts of the annotated datasets created in D7.4 might serve as the **training resource** for **object detection models** developed in D7.5. The **accuracy** and **structure** of these annotations are critical to the success of **machine learning applications** in **autonomous railway systems**.

By aligning with these interconnected deliverables, **D7.4** ensures the **continuity** and **coherence** of the data lifecycle within the **R2DATO project**.

[1] (Neumaier & Dubiel, D7.1 Requirements and specifications of the Data, 2024)

# 2 PROJECT OVERVIEW

As the digital transformation of railway operations accelerates, the availability of structured, real-world sensor data becomes a critical enabler for intelligent perception systems. Chapter 2 provides the contextual foundation for Deliverable 7.4 by positioning it within the R2DATO project's overarching strategy. It outlines the deliverable's objectives, explains its role within Work Package 7 – Data Factory, and defines the responsibilities of participating stakeholders. This chapter also clarifies how D7.4 supports the broader goal of fostering data-driven automation in rail transport across Europe.

## 2.1 Background Information and Relevance to the Data Factory

The **R2DATO project** seeks to advance the automation of railway operations through the development of intelligent perception systems, digital twin technologies, and data-centric infrastructures. Within this broader framework, **Work Package 7 (WP7) – Data Factory** plays a central role by providing the foundation for collecting, processing, and delivering high-quality sensor data to support machine learning and system integration across the project.

As part of this effort, **Deliverable D7.4 – Joint Collection and Annotation of Sensor Data** focuses on the acquisition and structured annotation of real-world data from operational railway settings. Its primary function is to translate raw sensor streams into well-organised, labelled datasets that can be directly used for developing perception models (D7.5) and for building a publicly accessible dataset (D7.6).

The relevance of D7.4 lies in its role as a practical interface between on-track data acquisition and AI-readiness. By generating consistent, richly annotated datasets, it provides the empirical groundwork for the Data Factory's machine learning activities.

## 2.2 Objectives of D7.4

The goals of **Deliverable D7.4** can be summarised as follows:

- **Coordinated Collection of Real-World Sensor Data**
  Conducting targeted measurement campaigns using multimodal sensor platforms—primarily **camera** and **LiDAR** systems—mounted on trains in active service, with attention to varied operational contexts.
- **Annotation and Labeling of Collected Data**
  Implementing a combination of **manual** and **automated annotation workflows** to assign object classes, identifiers, and spatial boundaries to observed elements, enabling their effective use in training AI models.
- **Standardisation and Enrichment with Metadata**
  Ensuring all annotated data are aligned with common standards—particularly the **RailLabel** (based on **ASAM Open Label format)** —and supplemented with detailed **metadata** to maximise interpretability and reuse.
- **Preparation for Machine Learning in D7.5**
  Delivering datasets that are directly usable for developing and evaluating perception algorithms under **Deliverable D7.5**, with clear emphasis on data quality, annotation consistency, and traceability.
- **Contribution to the Open Dataset in D7.6**
  Selecting and preparing representative segments of the annotated dataset for inclusion in the **open-access data release** foreseen in **Deliverable D7.6**, in line with R2DATO's commitment to transparency and reusability.

## 2.3 STAKEHOLDERS AND RESPONSIBILITIES

**Deliverable D7.4** is jointly coordinated by **Deutsche Bahn (DB)** and **SNCF**, who share responsibility for the overall planning, execution, and documentation of the data collection and annotation activities.

In addition to DB and SNCF, several other railway operators—**SMO**, **NRD**, **NS**, **GTSD**, and **TRV**—are involved as stakeholders, in accordance with the **R2DATO Grant Agreement**. While these partners do not carry primary implementation duties, they are actively engaged in the project and regularly informed about progress.

Work on D7.4 is structured to ensure transparency and alignment. Regular **biweekly coordination meetings** provide a platform for all consortium members to review ongoing activities, offer feedback, and stay connected to the development process. Once the draft version of the deliverable is prepared, all partners are invited to contribute to its final review and refinement.

This approach fosters a shared understanding and ensures that the resulting dataset is not only technically sound but also broadly applicable and accepted within the consortium—a crucial condition for its long-term use across different applications and partners within R2DATO.

# 3 DATA COLLECTION AND SENSOR SETUP

Capturing representative, high-quality sensor data from operational railway environments requires a carefully coordinated technical setup. Chapter 3 delves into the configuration of the sensor systems and the methods applied during the data acquisition campaigns. It describes the sensor types, their deployment on recording platforms, and the recording parameters applied under real-world conditions. This chapter thereby establishes the technical basis for reproducible data generation and provides transparency into the conditions under which the dataset was created.

## 3.1 Overview of Collected Data

To ensure a representative basis for subsequent annotation and machine learning tasks, nine scenes were selected and recorded under varying environmental and operational conditions. The selection was guided by the goal of capturing diverse and relevant constellations of railway scenarios, ranging from sparse morning settings to dynamic situations with high object density and low-light conditions.

The raw sensor data for each scene include synchronized recordings from multiple modalities: RGB cameras (long-range and mid-range), infrared camera, and a LiDAR system. These raw data form the foundation for the annotation and perception pipeline described in later sections of this deliverable.

| Scene Name | Duration | Time & Lighting | Scene Dynamics | Partner | Environmental Focus |
|---|---|---|---|---|---|
| **Morning Platform View** | 15 s | Morning, daylight | Dynamic | DB | Sparse environment with few people and infrastructure elements clearly visible |
| **Night Station Scene** | 15 s | Late night, artificial light | Dynamic | DB | Low visibility; signal detection and infrastructure under nighttime conditions |
| **Crowd at Platform** | 11 s | Daytime | Static, high density | DB | Many people on the platform; stationary train; realistic crowded platform scenario |
| **Incoming Train with Passengers at Platform** | 12 s | Daytime | Dynamic (train arriving) | DB | Dense crowd, train entering platform; critical for safety-relevant perception |
| **Platform with customers** | 11s | Daytime | Dynamic | SNCF | Rain drops on the lens |
| **Shunting yard arrival** | 70s | Daytime | Dynamic | SNCF | Rain drops on the lens |
| **Crossing train** | 8s | Daytime | Dynamic | SNCF | Rain drops on the lens. Water on a metallic bridge |
| **Plain line** | 35s | Daytime | Dynamic | SNCF | Rain drops on the lens. Water on a metallic bridge. A lot of parking train. |

Table 1 Selected Scenes of DB & SNCF

### 3.1.1 Rationale for Scene Selection

The selection of video sequences was guided by the objective of capturing representative, safety-relevant, and perceptually challenging situations within railway environments. The aim was not merely to document prototypical scenes but to assemble a dataset that reflects the operational diversity and environmental variability characteristic of real-world railway conditions.

1. **Scene Selection by Deutsche Bahn (DB)**

Deutsche Bahn focused on scenarios that combine high structural relevance with clear perceptual demands:

- The **Morning Platform View** provides a baseline for geometric calibration and infrastructure recognition under favourable lighting conditions.
- The **Night Station Scene** introduces low-illumination challenges, enabling evaluation of sensor robustness in environments with limited visibility.
- The **Crowd at Platform** captures complex visual occlusions and movement patterns, offering valuable data for algorithms addressing person detection and group dynamics.
- The **Incoming Train** sequence focuses on the interaction between fast-moving objects and static infrastructure at platform level, highlighting the need for accurate spatial-temporal alignment.

These scenes were selected to ensure a high degree of contextual and visual diversity, thereby supporting the development of models capable of generalising across a broad range of operational settings.

2. **Scene Selection by SNCF**

SNCF contributed a complementary set of recordings with a deliberate focus on environmental artefacts—most notably, **rain droplets on camera lenses**. This phenomenon, though often considered a disturbance, was intentionally included due to its frequency in operational rail contexts and its impact on sensor performance.

The SNCF dataset features:

- **Level Crossing under a Bridge**: A visually constrained scene marked by **lens moisture** and **low-contrast surroundings**, relevant for testing robustness against partial obstruction.
- **Platform with Customers**, **Shunting Yard Arrival**, and **Crossing Train**: All recorded under **rainy conditions**, these sequences include **reflections from wet surfaces**, **droplets on the optics**, and **metallic glare**, posing significant challenges for perception models.
- **Plain Line**: An open-track sequence with **persistent water on the lens** and the presence of **multiple parked trains**, making it valuable for examining long-range object recognition under degraded visual quality.

The inclusion of these scenes extends the dataset beyond idealised conditions and promotes the development of models that can operate reliably even when confronted with **sensor artefacts**, **weather effects**, and **limited visibility**. In doing so, the dataset contributes to more robust and realistic perception system design.

## 3.2 Object Classes Overview

This overview summarizes the key object classes relevant to sensor data annotation in the R2DATO project. It aligns with the needs of perception systems for autonomous train operation and reflects agreement between DB and SNCF on the scope and structure of annotations.

### 3.2.1 General Requirements

- Unique Object IDs for each instance (tracking across frames)
- Attribute tagging (e.g., pose, state, connectivity)
- Alignment with ASAM OpenLABEL / RailLabel schema
- Applicable to both 2D (camera) and 3D (LiDAR) annotations
- Timestamp synchronization across modalities

### 3.2.2 Dynamic Objects

| Object Class | Details | DB Scope | SNCF Scope |
|---|---|---|---|
| **Person** | Adult/child, pose (standing, walking), function (worker/passenger), distraction status, wheelchair/pram link | ✅ | ✅ |
| **Train** | Locomotive, regional/intercity, commuter, construction train; includes ConnectedTo(wagons), IsFront | ✅ | ✅ |
| **Wagons** | Freight, intercity, or regional wagons; annotated only if disconnected or relevant independently | ❌ | ✅ |
| **Crowd** | Group of 6+ people; individual identities not annotated | ✅ | ❌ |
| **Bicycle** | Ridden, pushed, parked; ConnectedTo(person), State(pushed/ridden/stationary) | ✅ | ❌ |
| **Group of Bicycles** | Clustered bicycles, not individually annotated | ❌ | ❌ |
| **Motorbike** | Motorcycles, scooters; parked/moving; ConnectedTo(person) | ❌ | ❌ |
| **Road Vehicle** | Cars, vans, buses, trucks, trailers | ✅ | ✅ |
| **Animal** | Large visible animals (e.g., dogs, deer); species/size optional | ❌ | ❌ |
| **Wheelchair** | Manual/electric; ConnectedTo(person) | ❌ | ❌ |

Table 2 Scope of Dynamic Objects

### 3.2.3 Static Objects

| Object Class | Details | DB Scope | SNCF Scope |
|---|---|---|---|
| **Track** | Rail lines; Attributes: EgoTrack, RailSide(left/right); polyline format | ✅ | ✅ |
| **Switch** | Track switches; Attributes: ConnectedTrack(left/right), State (position) | ✅ | ✅ |
| **Catenary Pole** | Overhead line poles (excluding foundation) | ✅ | ✅ |
| **Signal Pole** | Poles supporting active/inactive railway signals | ✅ | ✅ |
| **Signal** | Annotated with State (e.g., red, green, off) | ✅ | ✅ |
| **Signal Bridge** | Overhead structures supporting signals | ❌ | ✅ |
| **Buffer Stop** | End-of-track stoppers | ❌ | ✅ |
| **Control Box** | Small boxes controlling infrastructure; optional Locked attribute | ❌ | ❌ |
| **Level Crossing** | Road-rail intersection; may include barrier state | ❌ | ✅ |
| **Platform** | Passenger platforms (bounding box or polygon) (Not annotated by DB) | ❌ | ✅ |
| **Fouling point** | Concrete bar indicating the fouling point near the switches | ❌ | ✅ |

Table 3 Scope of Static Objects

### 3.2.4 Special Objects

| Object Class | Details | DB Scope | SNCF Scope |
|---|---|---|---|
| **Flame** | Visible fire from accidents or sources | ❌ | ✅ |
| **Smoke** | From engines or fire; affects visibility | ❌ | ✅ |
| **Fire Bucket** | Small safety devices (e.g., sand or water buckets) | ❌ | ❌ |

Table 4 Scope of Special Objects

### 3.3 Overview Condition Classes

While not directly used as a primary criterion for selecting the four annotated scenes presented in this deliverable, a structured taxonomy of environmental and situational **condition classes** has been established to support future dataset management and extension. These condition classes describe external influences that affect sensor performance, scene visibility, and the interpretability of sensor data. They are of particular relevance for:

- targeted filtering and retrieval of scenes with specific contextual characteristics,
- increasing variability and robustness in training data for machine learning models,
- evaluating system performance under adverse or edge-case scenarios.

Within the Deutsche Bahn infrastructure, scenes can already be queried based on such condition labels, thanks to the integration of weather metadata via an internal API. This API links the GNSS/GPS coordinates of sensor recordings with external weather data sources, enabling automatic tagging of scenes with relevant environmental descriptors such as *rain*, *fog*, or *snow accumulation*.

The following table provides an overview of defined condition classes that may affect visual clarity, sensor accuracy, and annotation quality. These classes form a foundation for context-aware data selection and scenario diversification in future perception system development.

| Condition Class | Description |
|---|---|
| **Rain** | Light, moderate, or heavy rainfall affecting visibility. |
| **Fog** | Light or dense fog obscuring objects in the scene. |
| **Snow** | Falling snow (light or heavy), impacting visibility and movement. |
| **Snow Accumulation** | Snow buildup on tracks or infrastructure. |
| **Hail** | Hailstorms causing visibility and surface hazards. |
| **Night** | Low-light conditions during nighttime. |
| **Daylight** | Clear daylight with full natural visibility. |
| **Overcast** | Cloudy skies, reduced sunlight but no precipitation. |
| **Cloud Cover** | Partial or full cloud coverage reducing light/shadow contrast. |
| **Dusk/Dawn** | Low-light transitions at sunrise or sunset. |
| **Sun Glare** | Intense sunlight directly affecting sensor exposure. |
| **Shadows** | Large shadows cast by objects, reducing visibility in areas. |
| **Reflections** | Light reflections on wet/glassy surfaces affecting detection. |
| **Artificial Lighting** | Presence of artificial light sources (street/train lights). |
| **Tunnel Lighting** | Reduced or flickering lighting inside tunnels. |
| **Tunnel Exit Light Change** | Sudden light change when exiting a tunnel. |

| | |
|---|---|
| **Wet Surfaces** | Rain/snow causing slippery, reflective surfaces. |
| **Puddles on Track** | Water accumulation on or near tracks. |
| **Flooding** | Significant water presence due to heavy rain. |
| **Ice** | Icy conditions on tracks, platforms, or surroundings. |
| **Heat Haze** | Visual distortion from heatwaves on hot days. |
| **Dust Storm** | High dust levels significantly impairing visibility. |
| **Smoke** | From fire or exhaust, reducing visual clarity. |
| **Obscured Objects** | Items hidden by snow, dirt, vegetation, etc. |
| **Dirt on Camera** | Lens contamination from dust, mud, etc. |
| **Mosquitoes/Insects** | Insects on lenses affecting image clarity. |
| **Strong Winds** | Environmental movement (e.g., vegetation, debris). |
| **High Humidity** | Fogging/condensation on lenses. |
| **Reduced Visibility** | General term covering fog, smoke, rain, dust. |
| **Electrical Interference** | EMI affecting sensor accuracy. |
| **Track Vibration** | Train-induced movement blurring camera view. |
| **Train Vibration Impact** | Continuous shaking from train motion. |
| **Construction Zones** | Temporary obstacles or changes near tracks. |
| **Track Maintenance Equipment** | Presence of service tools or machinery. |
| **Trackside Vegetation** | Plants partially obscuring track or objects. |
| **Graffiti on Signs** | Sign visibility reduced by vandalism. |
| **Graffiti on Objects** | Object surfaces obscured or distorted. |
| **Debris on Track** | Branches, trash, or materials obstructing view. |
| **Emergency Lighting** | Flashing lights from emergency vehicles. |
| **Heavy Traffic at Crossings** | Congestion near road-rail intersections. |
| **Bridge Shadow** | Shade from overpasses reducing visibility. |
| **Water Spray from Vehicles** | Splashing onto sensors from nearby traffic. |
| **Blocked Crossings** | Vehicles or barriers obstructing train paths. |
| **Trackside Workers** | Presence of personnel during maintenance. |
| **Disruptions Due to Derailment** | Obstructions from derailments affecting scene. |
| **Environmental Pollution** | Dust, chemicals, or fumes reducing visibility. |

Table 5 Overview of possible Condition Classes

## 3.4 Sensor Types and Their Configurations

The sensors are strategically placed to ensure that the data collected is comprehensive and accurate:

1. **LiDAR Placement**:
   - **Positioning**: LiDAR sensors are mounted on the top front of the train, offering a wide field of view with the ability to detect objects both around the train and directly ahead. This positioning allows for comprehensive situational awareness with focused detection in the train's forward path.
   - **Configuration**: The sensors' fields of view are configured to overlap slightly, ensuring continuous environmental monitoring, particularly in areas with complex track geometry.
2. **Camera Placement**:
   - **Positioning**: Cameras are mounted at the front of the train, providing a clear view of the track ahead.
   - **Configuration**: The cameras are configured to capture wide-angle views, ensuring that the tracks and adjacent areas are thoroughly monitored within the simulated environment. Adjustments were made to the cameras based on received technical specifications, ensuring proper alignment and functionality within the simulation.

| Sensor | Sensor Name | Sensor Data | Number of Sensors | Resolution | Sampling Rate | Partner | Notes |
|---|---|---|---|---|---|---|---|
| **LiDAR** | lidar_merged | 3D point clouds (PCD) | 6 | 6 x 10,240 points/frame | 4.67 Hz | DB | Horizontal with overlapping field of view |
| **Mid-Range RGB Camera** | camera_mid_range | Color images (8-bit, PNG) | 1 | 5472 x 3648 px | 4.67 Hz | DB | Center-mounted |
| **Long-Range RGB Camera** | camera_long_range | Color images (8-bit, PNG) | 1 | 3088 x 2064 px | 4.67 Hz | DB | Center-mounted |
| **Infrared Camera** | camera_infrared | Grayscale images (8-bit, PNG) | 1 | 640 x 512 px | 4.67 Hz | DB | Center-mounted |
| **GPS/GNSS** | – | Latitude & Longitude (WGS84) | 1 | – | 100/10 Hz | DB | – |
| **SNCF sensors** | Lowlight camera SEE3Cam_CV30 | Color images | 1 | 1280 x 720 px | 60 Hz | SNCF | Center-mounted |
| **Lidar** | AEVA AERIS 2 lidar | 4D point cloud | 1 | 1000 lines/frame | 10 Hz | SNCF | right-mounted |

Table 6 Overview of used Sensor Types

## 3.5 Data Acquisition Workflow

The **data acquisition process** forms the entry point of the **Data Factory** pipeline. It is designed to ensure that **sensor data** is systematically **captured**, **securely transferred**, and **prepared for further processing** in a manner that guarantees **traceability** and **technical consistency**. This stage is crucial for building a robust **data backbone** that feeds subsequent tasks such as **annotation**, **simulation**, and **model development**, as outlined in **Deliverable 7.1**.

**Sensor data** is recorded continuously during train operation using a suite of **onboard sensing systems**. These include **high-resolution cameras**, **LiDAR units**, and **GNSS/GPS modules**, all integrated into the **Vehicle Data Logger (VDL)**. As the train moves through its environment, the VDL captures **multimodal raw data streams**—covering **visual**, **positional**, and **environmental** aspects relevant to both infrastructure and operational context.

Once the journey concludes and the vehicle reaches a designated parking location, the data is **offloaded** via a dedicated interface: the **Data Touchpoint**. This **secure transfer hub** establishes the link between the **vehicle-side acquisition** and the **centralized infrastructure**. The data is then forwarded to a **Data Center**, where it undergoes an initial **pre-processing phase**. This phase includes **filtering**, **data compression**, **integrity checks**, and **format harmonization**, ensuring readiness for use within the **Data Factory's standardized processing environment**.

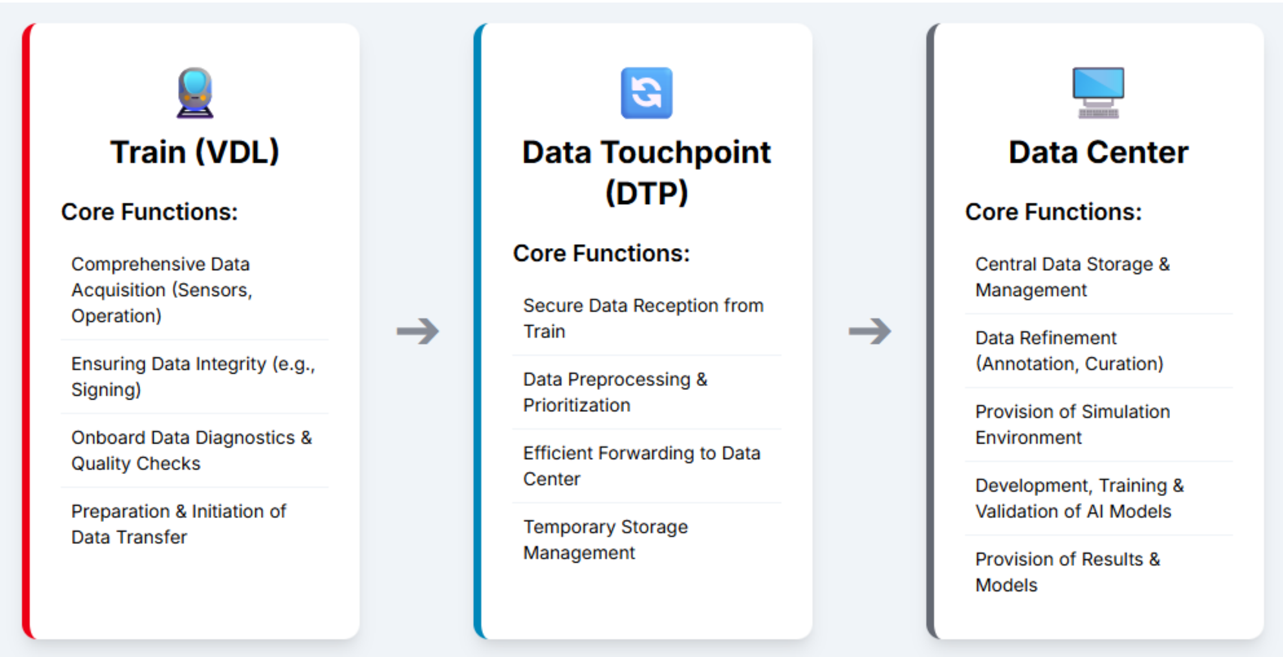

Figure 1 General Workflow of the Data Factory

The entire workflow is **modular** and **scalable**, built around a **uniform toolchain** that facilitates seamless transitions between **data acquisition**, **processing**, **annotation**, **simulation**, and **training**. It not only supports efficient data flow but also ensures **high data quality** and **operational reproducibility**. Ultimately, this acquisition framework enables **consistent data provisioning** for **machine learning workflows** and **simulation-based testing**—key components in developing **trustworthy perception systems** for autonomous rail transport.

The **Uniform Toolchain** is a standardized, modular suite of integrated platforms designed to support the entire lifecycle of data and AI development. This toolchain ensures **interoperability**, **scalability**, and **traceability** across all stages of the data processing and model training pipeline.

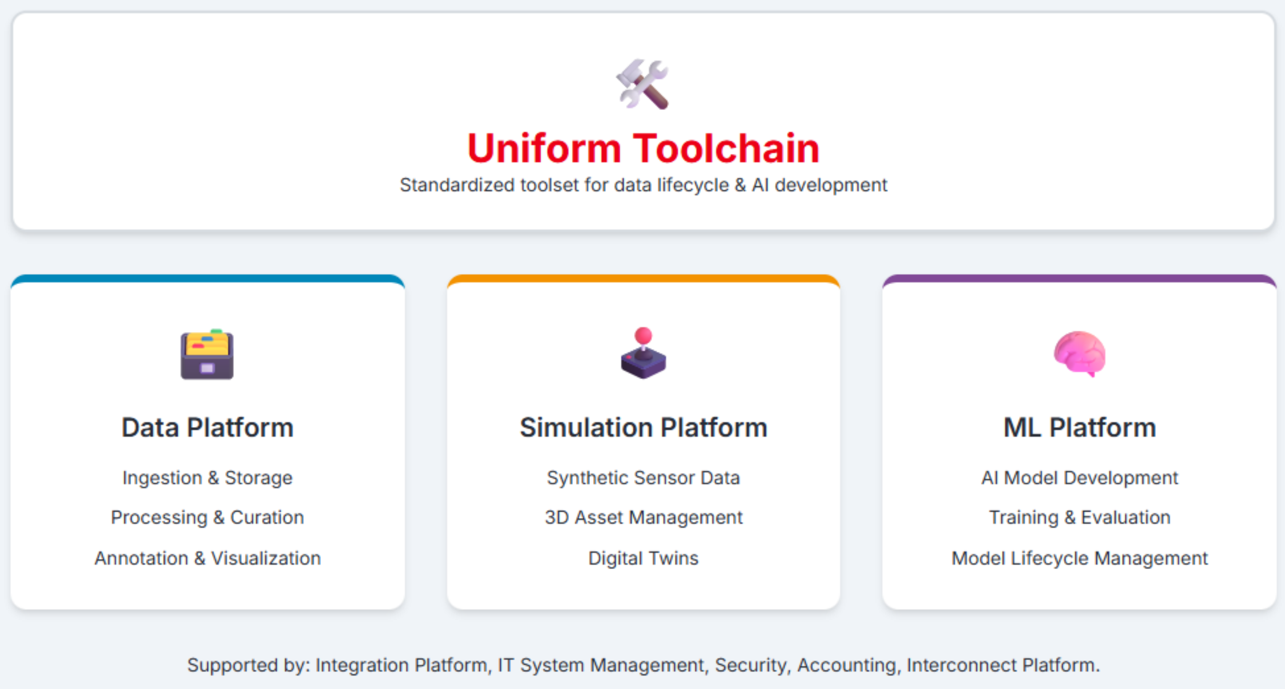


**Figure 2 Basic Components of the Uniform Toolchain**

The **Data Platform**, as detailed in D7.1[2] and expanded in D7.4, serves as the central node for **data ingestion, curation, storage**, and **annotation management**. It underpins all activities related to human and machine-driven labeling, versioning, and quality control of multimodal sensor datasets.

Complementing it, the **Simulation Platform**—introduced in D7.1 and elaborated in D7.3[3]—enables the creation of **synthetic sensor data**, **scenario-based simulations**, and **digital twins** of railway environments. This is crucial for training AI models under rare or safety-critical conditions that are difficult to capture in the real world.

The **ML Platform**, covered in D7.1 and central to D7.5, orchestrates the **training, evaluation**, and **lifecycle management** of AI models. It integrates model versioning, performance tracking, and experiment management, ensuring consistent reproducibility and alignment with certification requirements.

Together, these platforms form the pillars of the Uniform Toolchain, supported by shared services such as **IT System Management**, **Integration & Interconnect Platforms**, and **System Access & Security**. The architecture—designed to be **modular** and **extendable**—supports a wide range of use cases and aligns with the vision of a **Pan-European Data Factory (PEDF)**, facilitating cross-organizational collaboration without compromising data sovereignty.

[2] (Neumaier & Dubiel, D7.1 Requirements and specifications of the Data, 2024)

[3] (Neumaier & Dubiel, D7.3 Perform simulation with initial, 2024)

### 3.5.1 Real-World Data Collection

In the context of the **Data Factory**, a dedicated **sensor-equipped train** serves as the primary platform for **real-world data collection**. This train was equipped with a **sensor suite**—including **RGB & IR cameras**, **LiDAR**, **GNSS**, and **inertial measurement units**—integrated seamlessly with an onboard **Vehicle Data Logger (VDL).** The aim is to capture **authentic**, **multimodal sensor data** during **regular service operations**.

As the train traverses its designated route, the **sensor array** continuously records **environmental** and **situational information** in **real time**. This includes **dynamic scenes** featuring **infrastructure**, **terrain**, **vegetation**, **animals**, and other **railway-specific elements**. Such data is invaluable in reflecting the **heterogeneity** and **complexity** of **operational railway environments** across different **seasons** and **lighting conditions**.

During the journey, all **raw sensor data** is **securely stored** on the **VDL**. Upon completion of the route—typically when the train reaches its **parking position**—the data is **offloaded** through a secure **Data Touchpoint.** Before full ingestion into the **Data Center**, the data undergoes **pre-processing procedures** such as **synchronization**, **quality filtering**, and **formatting**, ensuring readiness for subsequent **annotation** and **usage**.

This **single-train data collection strategy** provides a **controlled** yet **comprehensive dataset** that mirrors the **real-world operational setting.** It establishes a solid **empirical foundation** for developing **AI perception models** and contributes directly to the creation of **certified datasets** within the **R2DATO framework**.

### 3.5.2 Challenges in Data Collection (Technical, Logistical)

The collection of real-world sensor data on-board of a moving railway vehicle presents a range of challenges, both technical and logistical, that must be systematically addressed to ensure data integrity, operational feasibility, and compliance with the overall R2DATO architecture.

**Technical Challenges**

1. **Sensor Synchronization and Calibration**

   Ensuring precise temporal and spatial synchronization across heterogeneous sensors (e.g., cameras, LiDAR, GNSS, IMU) is critical. Misalignments may compromise data quality and affect downstream processes such as annotation and model training.

2. **Data Volume and Storage Constraints**

   The high-resolution and high-frequency nature of sensor recordings generates substantial data volumes within a short timeframe. Managing this data locally on the Vehicle Data Logger (VDL) requires robust storage solutions.

3. **Environmental Interference**

   Variable weather conditions, lighting dynamics (e.g., glare, shadows), and occlusions can introduce noise and reduce the consistency of captured data. Adaptive sensor configurations and redundancy are often necessary to mitigate these effects.

4. **Hardware Reliability Under Motion**

   Sensors and logging hardware must withstand the mechanical vibrations, temperature fluctuations, and electromagnetic interference characteristic of the railway environment. Ensuring long-term operational stability is non-trivial.

**Logistical Challenges**

1. **Access and Deployment Timing**

   Coordinating access to the train for sensor installation and maintenance requires alignment with railway operations, often constrained by tight schedules and limited-service windows.

2. **Data Offloading Infrastructure**

   The availability of a fixed, reliable Data Touchpoint at designated parking locations is crucial for seamless data transmission. Any disruption in connectivity or power at these points can delay data ingestion into the Data Center.

3. **Operational Route Variability**

   To capture a representative dataset, the train must operate across diverse geographies and conditions. Planning such coverage while remaining within the bounds of scheduled service routes poses logistical complexity.

4. **Security and Data Protection**

   Transferring and storing sensitive operational data demands rigorous cybersecurity measures and adherence to data protection protocols, especially under multi-stakeholder and cross-border scenarios.

# 4 DATA ANNOTATION FRAMEWORK (DB)

Transforming raw sensor recordings into actionable training data requires a robust annotation pipeline. Chapter 4 introduces the procedures and technologies used to annotate the collected sensor data. It describes the interplay between manual and semi-automated annotation methods, outlines the quality assurance mechanisms applied, and presents the adoption of the ASAM Open Label standard as a unifying framework. The chapter serves to document how semantic consistency, technical validity, and cross-modality alignment were achieved across all annotated scenes.

## 4.1 DATA ANNOTATION TYPES (MANUAL VS. AUTOMATED)

Within the R2DATO Data Factory, the precise annotation of multimodal sensor data is fundamental to producing high-quality datasets for AI-based perception systems in autonomous rail operations. While Deliverable D7.1[4] defines the overarching standards and technical infrastructure, Deliverable D7.4 focuses on the concrete annotation processes—both manual and automated—that address the complexity and variability inherent in railway environments.

### 4.1.1 Manual Annotation

Manual annotation remains indispensable for scenarios requiring nuanced semantic interpretation, particularly in complex or ambiguous scenes. Conducted by trained annotators using dedicated interfaces within the Data Platform, manual annotation ensures:

- **Detailed labeling** across sensor modalities (e.g. RGB, infrared, LiDAR, RADAR),
- **Application of context-sensitive rules**, such as occlusion handling, truncation indicators, and semantic continuity across frames and modalities,
- **Use of geometry types** tailored to object class and sensor: 2D/3D bounding boxes, polygons, and semantic segmentations,
- **Assignment of persistent object IDs** (UUIDv4) with attention to **amodal representation**, i.e. inclusion of occluded object parts.

Manual annotation is particularly critical for:

- **Dynamic scenes** with partial visibility or visual clutter,
- **Rare or ambiguous classes**, such as smoke, track obstructions, or human trespassers,
- **Ground truth generation**, serving as a reference for validating automated pipelines.

The rail domain poses unique annotation challenges: fixed camera angles, elongated infrastructure (e.g. masts, cable trays), variable weather and lighting (e.g. fog, tunnels, snow), and rare but safety-

[4] (Neumaier & Dubiel, D7.1 Requirements and specifications of the Data, 2024)

critical edge cases. Annotators are required to apply domain expertise, especially in detecting subtle anomalies or non-standard elements.

While time- and resource-intensive, manual annotation provides unmatched semantic fidelity. In R2DATO, annotations adhere to the **RailLABEL schema**, a railway-specific extension of **ASAM OpenLABEL**, ensuring interoperability, traceability, and alignment with digital twin standards.

### 4.1.2 Automated Annotation

To enable large-scale annotation at speed, R2DATO employs automated techniques that complement manual efforts. These methods are powered by:

- **Pre-trained AI models** integrated into the Data Factory's ML platform,
- **Projection-based techniques**, such as transferring LiDAR labels onto RGB or radar images,
- **Rule-based augmentation** using digital maps and sensor fusion to automate context-aware labeling.

Automated annotation is particularly effective for:

- **Repetitive, well-defined objects** such as tracks, poles, and signals,
- **Low-complexity scenes** with minimal occlusion,
- **Intermediate labeling steps** in semi-supervised workflows.

This automation dramatically accelerates throughput: for example, pole detection from LiDAR data can process thousands of frames overnight. However, it also has limitations—misclassifications may occur in scenes with occlusion, environmental noise (e.g. snow, smoke), or rare object classes.

To address this, R2DATO implements a **hybrid human-in-the-loop approach**:

1. **Pre-annotation** using automated tools,
2. **Manual review and refinement**,
3. **Version-controlled validation and release** via the Annotation Platform.

Every annotation is tracked, versioned, and verified for **spatial, temporal, and semantic consistency** before being released for training or publication.

### 4.1.3 Integration and Workflow Management

Manual and automated annotations are orchestrated through a centralized **Annotation Platform**, which coordinates:

- **Task distribution** between human annotators and automated pipelines,
- **Version management** and change tracking,
- **Cross-modality consistency checks** to ensure data quality across sensors, timeframes, and object instances.

All annotations conform to the **RailLABEL JSON schema**, built on **ASAM OpenLABEL**, enabling seamless integration with other components of the Data Factory and ensuring long-term data reusability.

### 4.1.4 Manual vs. Automated Annotation – A Comparative Overview

| Aspect | Manual Annotation | Automated Annotation |
|---|---|---|
| **Performed by** | Human annotators | AI models, scripts, rule-based engines |
| **Best for** | Complex scenes, rare objects, occlusion-rich situations | Repetitive classes, low-complexity environments |
| **Annotation Types** | 2D/3D bounding boxes, polygons, semantic masks | Bounding boxes, projected masks |
| **Accuracy** | High semantic precision, context-aware | Variable; depends on training data and scene complexity |
| **Review Process** | Reviewed for correctness, consistency, and schema compliance | Requires manual validation and correction |
| **Tooling** | GUI-based tools with semantic tagging | Scripting tools, projection frameworks, ML pipelines |
| **Data Suitability** | Ground truth creation, complex edge cases | High-volume, structured sensor streams |
| **Standard Compliance** | Fully compliant with **RailLABEL** and **ASAM OpenLABEL** | Aligned with **RailLABEL**, requires validation |
| **Integration** | Annotation Platform (UI, task manager) | ML and Simulation Platforms for projections and automation |

Table 7 Manual vs. Automated Annotation – A Comparative Overview

### 4.1.5 A Hybrid Strategy for Quality and Scale

The R2DATO Data Factory adopts a **hybrid annotation strategy**, combining the interpretive depth of manual labeling with the scalability of automation. This approach, informed by best practices from both the automotive and railway domains, enables high-fidelity datasets to be built at scale.

Manual annotations provide reliable ground truth for model training and quality control. Automated methods, when properly validated and integrated, significantly boost efficiency and coverage. Through the consistent use of **RailLABEL**, centralized annotation workflows, and rigorous quality control, R2DATO ensures that annotated datasets meet the precision and robustness required for the deployment of **Level 4 autonomous train systems**.

# 5 ANNOTATION REQUIREMENTS AND STANDARDS

The utility of annotated datasets depends not only on annotation accuracy, but also on the clarity and consistency of the semantic structure behind them. Chapter 5 introduces the classification scheme and label attributes that guide the annotation of objects across multiple sensor types. It outlines the taxonomy adopted in the R2DATO context and describes how labeling rules were defined to ensure reproducibility and comparability across teams and tools. This chapter thus lays the groundwork for coherent interpretation and scalable annotation practices.

## 5.1 ALIGNMENT WITH ASAM OPEN LABEL FORMAT (SIMULATED DATA)

### 5.1.1 RailLabel Format for Sensor Data Annotation

As part of Deliverable D7.4, which focuses on the joint collection and annotation of real-world sensor data, the **RailLabel** format has been adopted or evaluated as a structured and extensible solution for railway-focused annotations.

RailLabel is an open-source annotation data format, originally developed by Deutsche Bahn within the *Digitale Schiene Deutschland* initiative. The format is hosted and maintained publicly at https://github.com/DSD-DBS/raillabel. (DB InfraGO AG, 2025)

#### 5.1.1.1 Purpose and Motivation

RailLabel was designed to provide a consistent, machine-readable, and semantically rich format for annotating multimodal sensor data in railway environments. It aims to:

- Support a wide range of annotation types (e.g., 2D bounding boxes, 3D cuboids, semantic segmentation, classification, etc.).
- Handle data from multiple synchronized sensors such as RGB cameras, LiDAR, radar, GNSS, and IMUs.
- Enable time-synchronized annotations across modalities.
- Serve as an intermediate format for training, testing, and benchmarking perception systems for autonomous train operations.

Its extensibility and domain specificity make it particularly suitable for use in the **R2DATO project**, where diverse sensor configurations, platforms (e.g., different trains), and annotations must be harmonized into a unified structure for further processing and ML training (as needed in D7.5 and D7.6).

#### 5.1.1.2 Structure of the RailLabel Format

RailLabel uses **JSON** as the container format. The top-level structure consists of the following core fields:

| Field | Description |
|---|---|
| **context** | Metadata about the annotation dataset: date, location, sensors used, coordinate systems. |
| **frames** | A dictionary of time-indexed frames (keyed by timestamps) containing sensor-specific annotations. |
| **streams** | Definitions of each sensor stream (e.g., front_camera, lidar_top) including type, modality, resolution, and calibration data. |
| **objects** | A list of all objects annotated throughout the dataset, each with a unique ID and class label. |
| **metadata** | Optional user-defined metadata including annotator information, tools used, version, etc. |

Table 8 RailLabel Structure

Each frame includes:

- Timestamp
- Stream-specific annotations (e.g., 2D bounding boxes for RGB camera, 3D cuboids for LiDAR)
- References to object IDs (for tracking across time)

This enables **temporally consistent annotations** across modalities, crucial for use cases involving object tracking, scene understanding, and predictive modelling.

#### 5.1.1.3 Supported Annotation Types

RailLabel supports (but is not limited to) the following annotation types:

- bbox: 2D bounding boxes for camera imagery
- cuboid: 3D bounding boxes for LiDAR or stereo
- poly2d and poly3d: polygonal annotations
- classification: Object-level semantic labels
- attribute: Key-value pairs describing scene conditions or object states (e.g., weather, occlusion, speed)

The annotation structure allows multiple annotation types to coexist per frame and per sensor, thus supporting multi-task training (e.g., detection + tracking + classification).

#### 5.1.1.4 Interoperability and Comparison to ASAM OpenLABEL

While RailLabel is tailored for railway scenarios, it shares conceptual similarities with the **ASAM OpenLABEL** standard:

- Both emphasize **sensor-agnostic structures**
- Both support **multi-sensor annotations**
- Both allow for **temporal consistency** and **object permanence**

In fact, **conversion tools** between RailLabel and ASAM OpenLABEL are under development, facilitating interoperability and easing integration with other EU-funded projects or external toolchains.

#### 5.1.1.5 Usage in R2DATO (Context for D7.4)

Within the scope of WP7, RailLabel is being used or benchmarked for:

- **DB's annotation pipeline** for real-world LiDAR and RGB data
- **SNCF's ingestion workflows** for synchronizing pre-annotated data
- Aligning real and synthetic datasets via a **common object model** and metadata standard
- Ensuring annotation quality and traceability across sensors and data sources

As discussed during multiple D7.4 meetings, RailLabel's extensible schema makes it suitable for:

- Capturing weather attributes
- Handling annotation confidence levels and manual/automatic origins
- Annotating multi-agent scenes (e.g., trains, humans, vehicles)

## 5.2 SYNCHRONIZATION

**Accurate synchronization** of sensor data is essential for enabling reliable annotation across multimodal recordings. It ensures that all **data streams**—including **video**, **infrared**, **LiDAR**, and other **sensor modalities—**are precisely **aligned in time**. This **temporal alignment** allows annotators to consistently relate events across different sources, such as correlating **motion patterns**, **object trajectories**, or **environmental changes** between modalities.

Without proper synchronization, annotations risk being **misplaced**, **inconsistent**, or **semantically incorrect**, which can severely compromise **data quality** and negatively affect **training datasets**, **simulation fidelity**, and **machine learning model performance**. Ensuring accurate time alignment is therefore critical to maintaining the **semantic integrity** and **contextual accuracy** of annotations.

To achieve this, all recorded data was synchronized using the **Precision Time Protocol (PTP)**. **PTP synchronization** is a high-accuracy method used to coordinate clocks across a network, often achieving synchronization within **microseconds** or even **nanoseconds**. The protocol operates by exchanging timing messages between a **master clock**—typically synchronized to a GPS or atomic source—and one or more **slave clocks**, which adjust their time accordingly.

PTP also accounts for **network delay**, measuring the time it takes for messages to travel between devices and compensating for it to maintain a **highly precise, shared time base** across all sensors. This enables **frame-level alignment** and ensures that annotations across modalities represent the **true temporal structure** of recorded scenes.

In summary, **PTP-based synchronization** underpins the integrity of the annotation process in the R2DATO Data Factory, forming a technical prerequisite for the creation of temporally consistent and semantically accurate datasets.

## 5.3 Unique Object Identification and Labeling

In the annotation files, a real-world object can be captured simultaneously from several sensors and at several time instances. A real-world object is a real object in the scene, for example a person or a catenary pole. If these real-world objects are captured from different sensors, one object can lead to several annotations. To track a real-world object over time and space, it is therefore important to give them unique identifier. Hence, the annotations contain an object ID which references to the specific object in the scene. To test tracking and fusion algorithms, this object ID is crucial, to determine the performance.

## 5.4 Attributes and Metadata in Annotations

Attributes in annotations are structured pieces of information that describe the properties of labeled objects in a dataset in more detail. They are often essential for many machine learning and computer vision tasks.

There are attributes, that are essential and present in all object classes, such as “occlusion”. “Occlusion” describes how visible an object is within a scene. Sometimes objects are covered by other structures leading to occlusion and e decreased visibility.

Other attributes are object specific. An example may the attribute “pose” for person. This attribute describes if a person is “standing”, “sitting”or “lying”. Such attributes are may important when a decision needs to be made. Decisions can be for example that the train brakes.

An example for an object class and its attributes is given in the Table 9 Attribute Schema for Person Entity in Annotation Frameworkbelow. The class is in this case “person”. The attributes are: function, isDummy, age, aid, isDistracted, carrieng, connectedTo, pose, occlusion. The attributes can then have different characteristics. The values are then predefined and can be selected by an annotator. For example the attribute “age” has four predefined values, i.e. “baby”, “child”, “adult” and “unknown”. The attributes also can have various types. We provided three types:

1) single-select: One option of the predefined values can be chosen.
2) multi-select: Several options of the predefined values can be chosen.
3) boolean: The option to be chosen is either true or false.

```yaml
person:
  function:
    attribute_type:
      type: single-select
      options:
        [passenger, worker, security, staff, uniformed, other, unknown]
  isDummy:
    attribute_type: boolean
  age:
    attribute_type:
      type: single-select
      options: [adult, baby, child, unknown]
  aid:
    attribute_type:
      type: multi-select
      options: [none, whiteCane, walkingAid, other, unknown]
  isDistracted:
    attribute_type:
      type: single-select
      options: ["true", "false", unknown]
  carrying:
    attribute_type:
      type: multi-select
      options: [none, suitcase, backpack, umbrella, baby, other]
  connectedTo:
    attribute_type: multi-reference
  pose:
    attribute_type:
      type: single-select
      options: [upright, sitting, lying, other, unknown]
  occlusion:
    attribute_type:
      type: single-select
      options: [0-25%, 25-50%, 50-75%, 75-100%, undefined]
```

Table 9 Attribute Schema for Person Entity in Annotation Framework

## 5.5 Legal and Ethical Considerations

In the following, the legal and ethical considerations for datasets are described. First the legal and following the ethical aspects are outlined:

### 5.5.1 Legal Aspects

**Copyright and Licensing**: Datasets are typically released under specific licenses that govern how they may be used, modified, and distributed. It is important to carefully review these licenses to ensure that any annotation work is legally compliant and does not  on the rights of the data creators.

**Privacy and Data Protection**: The proposed datasets include images and videos recordings that may capture identifiable individuals, vehicles, or locations. Depending on the specific law of a country, legal frameworks such as data protection laws may require that personal information is anonymized or that specific safeguards are in place to protect the privacy of individuals.

**Informed Consent and Data Collection**: The data used for annotation should be collected with appropriate consent when necessary. Using data that was gathered without proper authorization or transparency can result in legal liabilities, especially if individuals' rights are violated.

**Jurisdictional and Export Considerations**: Certain datasets or their annotated forms may be subject to legal restrictions based on national security, export controls, or jurisdiction-specific regulations. Compliance with such laws is necessary to avoid legal consequences.

### 5.5.2 Ethical Aspects

**Privacy and Anonymization**: Even when not legally required, it is ethically responsible to anonymize identifiable elements within a dataset, such as faces, license plates. In such a way individual privacy is secured.

**Bias and Representation**: Annotators and person that select the data should be trained to recognize and minimize bias in the annotation process. Unbalanced datasets can lead to biased machine learning models.

**Transparency and Documentation**: Ethical annotation involves that it is documented which, tools, and criterias were used during the process. This ensures transparency for the individuals.

**Working Conditions and Fair Treatment**: Ethical data annotation also consideres the correct treatment of annotators. This includes fair wages, reasonable working hours and proper training.

# 6 DATA STORAGE AND PROCESSING

The successful implementation of Deliverable D7.4—specifically, the joint collection and annotation of multimodal sensor data—relies heavily on a robust and scalable data infrastructure. This infrastructure is defined in **Deliverable D7.1**[5], which lays out the full technical architecture of the **Data Factory**. It covers all stages of the data lifecycle: from acquisition and secure transmission to long-term storage, processing, annotation, and provision for downstream use.

The platform is designed to handle the **volume, variety, and velocity** of data generated by sensor-equipped trains operating in live railway environments. It enables efficient **data-driven development of AI models** for perception systems while ensuring **interoperability, security, and regulatory compliance**.

## 6.1 DATA HANDLING AND STORAGE INFRASTRUCTURE

Data from sensor-equipped vehicles is collected by the **Vehicle Data Logger (VDL)**, a specialised edge computing system designed for railway environments. Upon acquisition, the data is cryptographically signed and validated before it is transferred via the **Data Touchpoint** to the central infrastructure. This ensures **authenticity**, **traceability**, and **tamper resistance** from the earliest point in the data pipeline.

At the heart of the Data Factory is the **Data Center**, hosted and maintained by **DB InfraGO**, which includes:

- **5 Petabyte tiered storage architecture** combining SSDs for high-speed ingestion and magnetic media for long-term archival;
- **Two high-performance computing (HPC) clusters**, with dedicated purposes:
  - One for **machine learning model training and evaluation**;
  - One for **data simulation and synthetic data generation** (relevant to D7.3[6] but supporting annotation alignment);
- **Cloud-native extensions**, allowing dynamic scalability and redundancy for compute and storage-intensive tasks;
- A **spine-leaf LAN topology**, enabling low-latency, high-bandwidth communication between compute, storage, and annotation systems.

The infrastructure has been stress-tested under operational conditions and supports parallel ingestion and processing of large-scale, high-frequency data streams, such as those produced during full-day recording campaigns. Importantly, it is structured to ensure **future extensibility**, allowing the integration of additional data sources or partner-specific pipelines.

---

[5] (Neumaier & Dubiel, D7.1 Requirements and specifications of the Data, 2024)

[6] (Neumaier & Dubiel, D7.3 Perform simulation with initial, 2024)

## 6.2 Data Touchpoint Functions

The **Data Touchpoint** is more than a passive conduit. It comprises subcomponents like the **Automated Offload** and **DC Transmitter**, which:

- Establish reliable network connections at predefined railway parking locations.
- Manage the secure duplication and removal of data post-transfer.
- Relay diagnostic metadata, such as data sanity reports and transfer logs, to IT management subsystems.

These mechanisms ensure that only validated, non-redundant data enters the Data Factory's processing pipeline.

## 6.3 Integration with the Data Factory Platform

The **Integration Platform** serves as the orchestration layer of the Data Factory. It provisions:

- **Interactive computation environments** for data analysts and ML engineers.
- **Process automation**, enabling scheduled tasks, workflow execution, and containerized deployments.
- **Centralized interconnectivity**, linking internal toolchains (e.g., Data Platform, ML Platform, Simulation Platform) with external IT assets and other Data Factories.

This interoperability ensures that each data item flows seamlessly from ingestion to annotation, training, and ultimately to dissemination. The **Integration Platform** is the orchestration layer that connects raw data to downstream applications. It bridges ingestion, annotation, training, simulation, and ultimately, dataset dissemination.

It enables:

- **Interactive computing environments** (e.g., JupyterHub, VS Code Web) for data scientists, annotators, and ML engineers;
- **Process orchestration** using containerised workflows (e.g., with Kubernetes and Airflow), supporting scalable annotation and training pipelines;
- **Data lineage and tracking tools**, enabling reproducibility and impact assessment;
- **Secure interconnection** between the Data Factory's internal toolchains—Annotation Platform, ML Platform, and Simulation Platform—and external services such as remote storage, partner infrastructures, or other EU Data Factories.

This design ensures that **each data item is traceable, version-controlled, and reproducible**, making the platform compliant with FAIR principles (Findable, Accessible, Interoperable, Reusable).

## 6.4 Data Transfer Protocols and Security Measures

Security and access governance are managed by the **System Access and Security Platform (SASP)**. It provides an integrated suite of capabilities designed to uphold **confidentiality, integrity, and availability (CIA)** across all layers of the infrastructure.

Key elements include:

- **Role-based access control (RBAC)**, with differentiated roles such as Data Owner, Steward, and Consumer;
- **Multi-factor authentication (MFA)** and federated login support for authorised users;
- **End-to-end encryption** for data in transit and at rest;
- **Comprehensive audit trails** and **anomaly detection**, enabling forensic analysis in case of policy violations or cyber threats;
- **Policy enforcement engines** for managing data retention, anonymisation, and export controls (e.g., in compliance with GDPR and cybersecurity directives).

The security architecture is aligned with **ENISA recommendations** and supports future certification under **ISO/IEC 27001**.

## 6.5 Dataset Management for Open Access

The **Data Platform** serves as the unified interface for managing, annotating, and curating datasets for public release in D7.6. It includes:

- **Automated metadata enrichment**, including spatial-temporal tags, sensor calibration parameters, and annotation quality metrics;
- A **semantic data catalogue**, allowing users to query the dataset by scene type, sensor modality, object class, or annotation confidence;
- **Integrated annotation tools**, supporting manual review, polygon/cuboid editing, and class refinement;
- **Data export and policy compliance modules**, enforcing anonymisation rules, classification by data sensitivity, and controlled access via token-based authentication.

Together, these components ensure that only **validated, high-quality, and regulation-compliant data** are published to external stakeholders—maximising the scientific and societal value of the dataset while mitigating legal and reputational risks.

## 6.6 Summary

Chapter 6 outlines the end-to-end data infrastructure underpinning the Data Factory, as defined in Deliverable D7.1. It describes how real-world sensor data—collected from equipped trains—is securely offloaded, processed, annotated, and provisioned for machine learning and open data publication.

Key components include:

- A **Vehicle Data Logger** for authenticated onboard data acquisition;
- A **Data Touchpoint** for secure, automated offloading and integrity checks;
- A central **Data Center**, equipped with petabyte-scale storage and HPC clusters for AI training and simulation tasks;
- An **Integration Platform** to orchestrate data workflows, toolchain interoperability, and interactive ML development;

- A **Security Layer** (SASP) enforcing access control, encryption, and audit logging to ensure data confidentiality and compliance;
- A **Data Platform** for metadata enrichment, annotation, and policy-based open data management, supporting downstream tasks in D7.5 and D7.6.

# 7 EVALUATION OF DATA ANNOTATION QUALITY

**Disclaimer:**
***The following section outlines a conceptual framework for the evaluation and assurance of annotation quality as part of Deliverable D7.4. At this stage, the methodology remains theoretical and is intended as a set of recommendations for future implementation. The proposed processes, metrics, and tools are designed to support a scalable and systematic approach to quality control but have not yet been operationalised or empirically validated within the project.***

## 7.1 ESTABLISHING METRICS FOR ANNOTATION QUALITY

In the context of autonomous railway perception systems, the quality of annotated datasets has a direct and profound impact on the reliability, generalisability, and safety of machine learning models trained thereupon. Unlike domains with well-defined benchmarks, Deliverable D7.4 addresses the challenge of ensuring annotation quality in the absence of definitive ground truth references. This necessitates a robust framework of validation metrics and processes tailored to multi-sensor, real-world data.

Key metrics established for annotation quality include:

- **Formal Schema Integrity**: Verifying that annotation files (in JSON or similar structured formats) conform to a well-defined schema. This involves checking for required fields (e.g., class, coordinates, frame reference), proper data types, and absence of missing or null values.
- **Geometric Validity**: Ensuring that 2D bounding boxes are contained within the image boundaries and that 3D boxes possess positive, non-zero dimensions. Checks are implemented for values outside plausible spatial ranges (e.g., boxes extending beyond sensor field of view or containing negative depth).
- **Semantic Correctness**: All labels must correspond to a pre-approved class ontology. Typographical errors, inconsistent naming, or use of deprecated classes are automatically flagged. KPIs here include the frequency of invalid or unrecognized labels.
- **Distributional Regularity**: Statistical metrics such as class frequency histograms, box size distributions, and spatial heatmaps are used to identify outliers or skewed data. KPIs include z-scores and clustering deviations that exceed defined thresholds.

- **Redundancy and Duplication**: Automated comparisons of box overlaps (via IoU) detect near-identical annotations for the same object. A KPI is the percentage of boxes per image with an IoU > 0.8 and identical class label.
- **Temporal Continuity (for sequential data)**: Measures include object persistence, stability of class identity across frames, and smoothness of motion trajectory. Inconsistencies are logged, and KPIs reflect average IoU consistency across time.

## 7.2 VALIDATION METHODOLOGIES AND ANNOTATION QUALITY CONTROL WORKFLOW

### 7.2.1 Heuristic and Rule-Based Checks

Fundamental checks are the first line of defence against annotation errors:

- **Size Constraints**: Bounding boxes must lie within realistic limits for each class. For instance, a pedestrian in LiDAR should have a height between 1.4–2.0 meters. Extremely small or oversized boxes are flagged.
- **Bounding Box Coherence**: Rules check that boxes are well-formed, with x_min < x_max and y_min < y_max for 2D, and similar logic for 3D.
- **Null or Empty Annotations**: Instances where expected objects are missing are flagged based on contextual expectations (e.g., station platforms should contain at least one pedestrian in specific frames).
- **Incorrect Label-Box Correlation**: Annotations where the class label does not match the object's physical size or position (e.g., "train" label assigned to a box the size of a "signal post") are flagged.

These checks are implemented via custom Python scripts leveraging libraries such as json, NumPy, and bbox. The results of these analyses generate per-image quality flags.[7]

### 7.2.2 Inter-Modal Consistency Validation

Given the multi-sensor nature of the dataset (RGB, IR, LiDAR), cross-modal consistency is a critical dimension of validation. Three principal approaches are used:

- **3D→2D Projection Consistency**: Using intrinsic and extrinsic calibration parameters, each 3D bounding box is projected onto the 2D planes of the RGB and IR cameras. The resulting 2D projections are compared to existing 2D annotations using IoU. Discrepancies (e.g., IoU < 0.5) may indicate calibration errors, misannotations, or missing labels.
- **Frustum Alignment Checks**: From 2D boxes in camera images, a 3D frustum is generated. LiDAR annotations are validated to ensure they fall within the expected frustum volume.
- **Semantic Consistency Across Modalities**: Object classes assigned across different sensors must be consistent. For example, a "bicycle" labeled in RGB imagery should not be marked as an "unknown object" in IR, unless occlusion or ambiguity is justified.

[7] (encord-team/encord-active: The toolkit to test, validate, and ..., 2025)

### 7.2.3 Temporal Coherence Checks (Sequential Data)

When annotations span sequences (e.g., video or LiDAR sweeps), consistency over time is verified:

- **Trajectory Smoothness**: Bounding box coordinates for the same object should change smoothly over time. Sharp displacements or fluctuations are indicative of tracking or annotation errors.
- **Identity Stability**: Object IDs must persist across frames. Re-use or abrupt re-appearance of IDs is detected through tracking metrics.
- **Temporal IoU Analysis**: The average IoU of boxes across successive frames is computed. A KPI of >0.7 across frames is indicative of high temporal consistency.

### 7.2.4 Self-Consistency and Annotation Correlation Analysis

- **Inter-Annotator Agreement**: For samples annotated by multiple users, agreement is computed using metrics like Cohen's Kappa or pairwise IoU. Low agreement flags ambiguity or guideline issues.
- **Statistical Outlier Detection**: Distributional anomalies in box size, aspect ratio, or class frequency are flagged using histogram analysis and clustering (e.g., DBSCAN).
- **Contextual Anomaly Checks**: Annotations violating physical logic (e.g., a pedestrian inside a train's bounding box) are detected using relational rules.[8]

## 7.3 Comparative Analysis: Simulated vs. Real Data

The R2DATO Data Factory uses simulated data as a benchmark for annotation quality due to its inherently consistent generation. Comparative evaluation reveals typical divergences:

- **Box Geometry Divergence**: Real data shows higher variability in box shapes, sizes, and alignment due to manual input and sensor noise.
- **Class Label Drift**: Mislabeling in real data (e.g., confusion between "signal" and "lamp post") is rare in simulation.
- **Cross-Modal Misalignments**: Real-world sensor desynchronization introduces projection errors that are absent in simulation environments.

Metrics from real and synthetic datasets are cross-compared to develop empirical baselines for annotation quality expectations.

## 7.4 Error Detection and Correction Mechanisms

The correction mechanism is tiered to prioritize severity and impact:

- **Automated Flagging**: All aforementioned validation steps populate a report with flagged annotations categorized by error type and confidence level.

[8] (Automated QA, Review & Honeypots, 2025)

- **Machine-Learning Validation**: Pre-trained detectors (e.g., YOLOv8, Cleanlab ObjectLab) are run against the dataset. Discrepancies between model output and human annotation are quantified and visualized.
- **Visual Debugging and Manual Review**: High-confidence errors and a statistical sample of unflagged data are reviewed in tools like CVAT, FiftyOne, and Open3D.
- **Annotation Update Loop**: Confirmed errors are either corrected directly or returned to the labeling teams with feedback. Metrics on correction rates and re-review cycles feed back into the KPI dashboard.[9]

## 7.5 IDEA FOR A VALIDATION PROCESS

The overarching goal of the validation process is to systematically evaluate the **reliability, completeness, and internal consistency** of the annotated data collected under D7.4. Given the absence of a definitive ground truth, particularly for real-world sensor data (RGB, IR, LiDAR), the process relies on a suite of indirect verification techniques.

The validation process is designed to:

- Detect formal, semantic, spatial, and temporal inconsistencies in annotations.
- Provide feedback loops for manual correction and dataset refinement.
- Support downstream tasks such as model training and benchmarking with high-confidence data.[10]

### 7.5.1 Multi-Stage Validation Pipeline

The proposed validation process is structured as a **multi-stage pipeline**, allowing for layered error detection, increasing the depth and reliability of quality control.

[9] (Automated Quality Assurance for Object Detection Datasets, 2025)

[10] (Data Collection and Annotation - Ultralytics YOLO Docs, 2025)

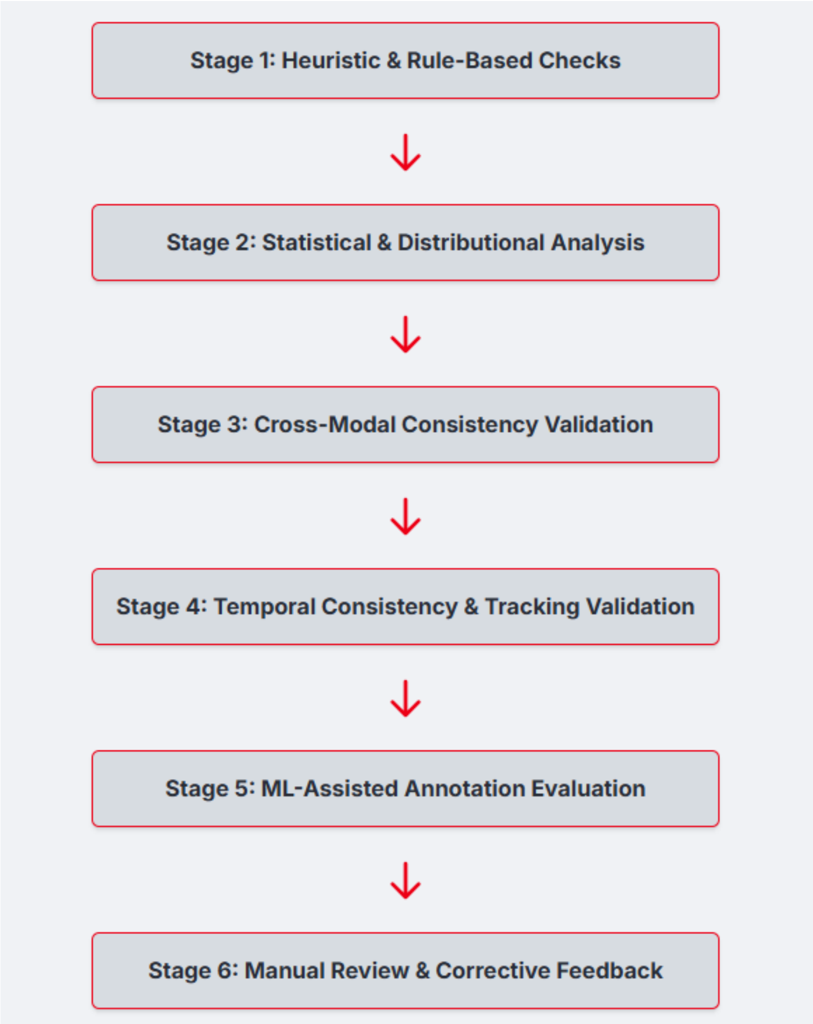


Figure 3 Idea for a Validation Process (Proposed Multi-Stage Pipeline)

**Stage 1: Heuristic and Rule-Based Checks**

**Purpose:** Detect blatant errors using simple, fast, and deterministic methods.

**Example checks:**

- Invalid bounding box dimensions (zero or negative values).
- Coordinates outside image bounds or LiDAR frame limits.
- Class labels not found in the project ontology.
- Redundant or duplicate bounding boxes with IoU > 0.8.

**Tools:** Python (NumPy, json, pybboxes), Deepchecks, custom validation scripts.

**Stage 2: Statistical and Distributional Analysis**

**Purpose:** Identify subtle anomalies based on statistical deviation from expected patterns.

**Key techniques:**

- Frequency analysis of class distributions.

- Detection of bounding box outliers using z-scores or clustering (e.g., DBSCAN).
- Visualization of size and shape distributions per class.

**Output:** Statistical report and flagged samples.

**Stage 3: Cross-Modal Consistency Validation**

**Purpose:** Assess consistency between annotations across different sensor modalities (RGB, IR, LiDAR).

**Methods:**

- **3D→2D Projection:** LiDAR boxes are projected into image space. The Intersection over Union (IoU) with existing 2D annotations is computed.
- **Frustum Validation:** Checks whether 3D objects fall within the volume implied by corresponding 2D annotations.
- **Semantic Consistency:** Ensures object classes are uniformly labeled across modalities.

**KPI Examples:**

- Percentage of cross-modal objects with IoU > 0.5.
- Class label agreement ratio across sensors.

**Stage 4: Temporal Consistency and Tracking Validation**

**Purpose:** Ensure annotations remain logically coherent across consecutive time steps (video or scan sequences).

**Checks include:**

- Continuity of object IDs across frames.
- Smooth trajectories and plausible motion vectors.
- Detection of object "teleportation" or disappearance.

**KPI Examples:**

- Frame-to-frame IoU consistency.
- Dropout rate of persistent objects across time.

**Stage 5: ML-Assisted Annotation Evaluation**

**Purpose:** Use weak supervision via machine learning to highlight potential label issues.

**Approaches:**

- Run pre-trained detectors (e.g., YOLO, CenterPoint) on the dataset.
- Compare model predictions with human annotations.
- Use tools like Cleanlab's ObjectLab to quantify probable errors (e.g., missed object, misclassification, spatial misalignment).

**Outcome:** Prioritized review list of suspect annotations.

**Stage 6: Manual Review and Corrective Feedback**

**Purpose:** Human-in-the-loop validation of critical cases.

**Methods:**

- Visual inspection using tools like CVAT, FiftyOne, or 3D BAT.
- Structured review sessions focusing on flagged data and a representative sample.

**Best practices:**

- Annotator-reviewer separation (4-eyes principle).
- Logging of error types and review outcomes.
- Systematic feedback loop into annotation guidelines and training.

## 7.6 Implementation Considerations

While the multi-stage validation process provides a comprehensive conceptual framework, its successful application hinges on a set of technical, organizational, and infrastructural considerations. The implementation of this framework must be both **scalable and adaptable** to the specific requirements of the R2DATO Data Factory. This section outlines the key factors and practical requirements that should be addressed in the operationalisation of annotation validation.

### 7.6.1 Data Format Standardisation

The foundation of any automated validation system is the availability of well-structured and standardised annotation data. All sensor annotations (2D and 3D) must conform to a project-wide **data schema**, which should define:

- Required fields (e.g., class, bbox, sensor_id, timestamp, object_id)
- Accepted data types (e.g., integers for coordinates, strings for labels)
- Format constraints (e.g., 2D boxes in pixel space, 3D boxes in metric world coordinates)
- Semantic ontologies (e.g., a controlled vocabulary of object classes)

A centralised schema definition should be maintained in a version-controlled repository and used to validate incoming JSON or equivalent files using schema validation tools (e.g., jsonschema in Python).

### 7.6.2 Modular Validation Pipeline

The proposed validation steps should be implemented as **independent, modular components**, each responsible for a specific type of check (heuristic, statistical, cross-modal, temporal, ML-based). This modular architecture offers several advantages:

- **Reusability**: Validation modules can be adapted for different datasets or sensor setups.
- **Parallelisation**: Components can be executed concurrently on subsets of data to improve throughput.
- **Extensibility**: New validation routines can be added without disrupting the existing pipeline.

Each module should be implemented as a **self-contained script or service** with a clear input/output interface, such as annotated JSON input and a quality report or error list output in CSV or structured JSON format.

### 7.6.3 Toolchain and Environment Integration

The validation pipeline should be tightly integrated into the R2DATO Data Factory's toolchain. This includes:

- **Annotation Tools**: Validation should be triggered automatically after annotation submission (e.g., in CVAT) or as part of the review workflow.
- **Data Storage and Access**: Scripts must have secure and efficient access to both raw sensor data (images, point clouds) and annotation files via network-mounted volumes or APIs.
- **Feedback Channels**: Errors detected by validation scripts should be linked back to the annotator UI to facilitate correction (e.g., clickable issue markers in CVAT or FiftyOne).

Support for **Python-based environments** with key libraries (NumPy, pandas, OpenCV, Open3D, pybboxes, bbox, jsonschema, etc.) is recommended.

### 7.6.4 Scalability and Performance

Given the large volume of data expected in D7.4 (multi-modal, time-sequenced sensor streams), the validation infrastructure must support:

- **Batch Processing**: Annotated datasets should be validated in segments (e.g., by day, location, or sensor configuration).
- **Compute Resource Allocation**: ML-based and projection-intensive validation steps (e.g., 3D→2D transformations) should be performed on high-performance nodes with GPU or multi-core CPU support.
- **Logging and Monitoring**: All validation runs should produce comprehensive logs, including timing, error rates, and failure summaries.

A dashboard interface could be developed to monitor validation progress and KPIs in real-time.

### 7.6.5 Quality Thresholds and Decision Rules

Each validation module should define **thresholds** and **decision criteria** for flagging errors, e.g.:

- IoU below 0.5 for cross-modal projections.
- Z-score > 3 for bounding box area anomalies.
- Motion discontinuity between frames > 5m within 100ms for the same object ID.

These thresholds should be configurable and iteratively adjusted based on empirical results and expert feedback.

### 7.6.6 Governance and Human Oversight

To ensure the robustness of the system:

- **Review Guidelines** should be created to interpret flagged errors and define when a manual correction is required.
- **Quality Gatekeepers** (e.g., senior annotators or domain experts) should oversee the review process and update annotation guidelines as needed.

- **Feedback Loops** should ensure that frequent error types lead to corrective action upstream—e.g., clarifying label definitions or improving annotation tool ergonomics.[11]

[11] (Huan, 2025)

# 8 RESULTS AND DISCUSSION

Building on the technical and methodological groundwork established in the preceding chapters, Chapter 8 presents the annotated dataset created through the joint efforts of Deutsche Bahn and SNCF. It introduces each annotated scene, discusses the diversity of captured environments, and provides insights into the annotation scope, sensor-specific contributions, and object visibility across modalities. In addition to presenting the results, the chapter critically reflects on key findings and challenges encountered during the process, linking the dataset to its intended use in model development and public dissemination.

## 8.1 Overview of the Annotated Dataset

| Scene Name | Sequence ID/frame ID scheme | Date & Time | Scene Description | Frames (Approx.) | Partner | Notes |
|---|---|---|---|---|---|---|
| **Morning Platform View** | 2022-11-23_07-46-50__2022-11-23_08-27-46+16 | 23.11.2022, 07:46–08:27 | Morning, few people, platform at ego track | ~32 | DB | Bridge visible in background |
| **Night Station Scene** | 2022-11-30_23-29-58__2022-11-30_23-52-39+13 | 30.11.2022, 23:29–23:52 | Night, few individuals, platform at ego track, train seen | ~26 | DB | Low-light scene on platform |
| **Crowd at Platform** | 2022-12-09_09-48-40__2022-12-09_09-51-15+12 | 09.12.2022, 09:48–09:51 | Daytime, platform at ego track, many people, train present | ~24 | DB | – |
| **Incoming Train with Passengers at Platform** | 2022-12-09_09-48-40__2022-12-09_09-50-39+12 | 09.12.2022, 09:48–09:50 | Daytime, many people visible, train present | ~24 | DB | People visible in the background |
| **Platform with customers** | frame_1 to frame_15 | 21.06.2024, 13:04–13:05 | Daytime, rain drops, platform and customers | ~15 | SNCF | |
| **Shunting yard arrival** | frame_16 to frame_21 | 21.06.2024, 13:22–13:24 | Daytime, rain drops, stationary trains | ~6 | SNCF | A switch on front of the train |
| **Crossing train** | frame_22 to frame_27 | 21.06.2024, 13:03–13:04 | Daytime, rain drop Main line and | ~5 | SNCF | a crossing train on the right |
| **Plain line** | frame_28 to frame_101 | 21.06.2024, 12:48–13:25 | Daytime, rain drop, main line, | ~75 | SNCF | Main line at different places |

Table 10 Overview annotated Scenes

### 8.1.1 Scene Profile: Trackside Passengers

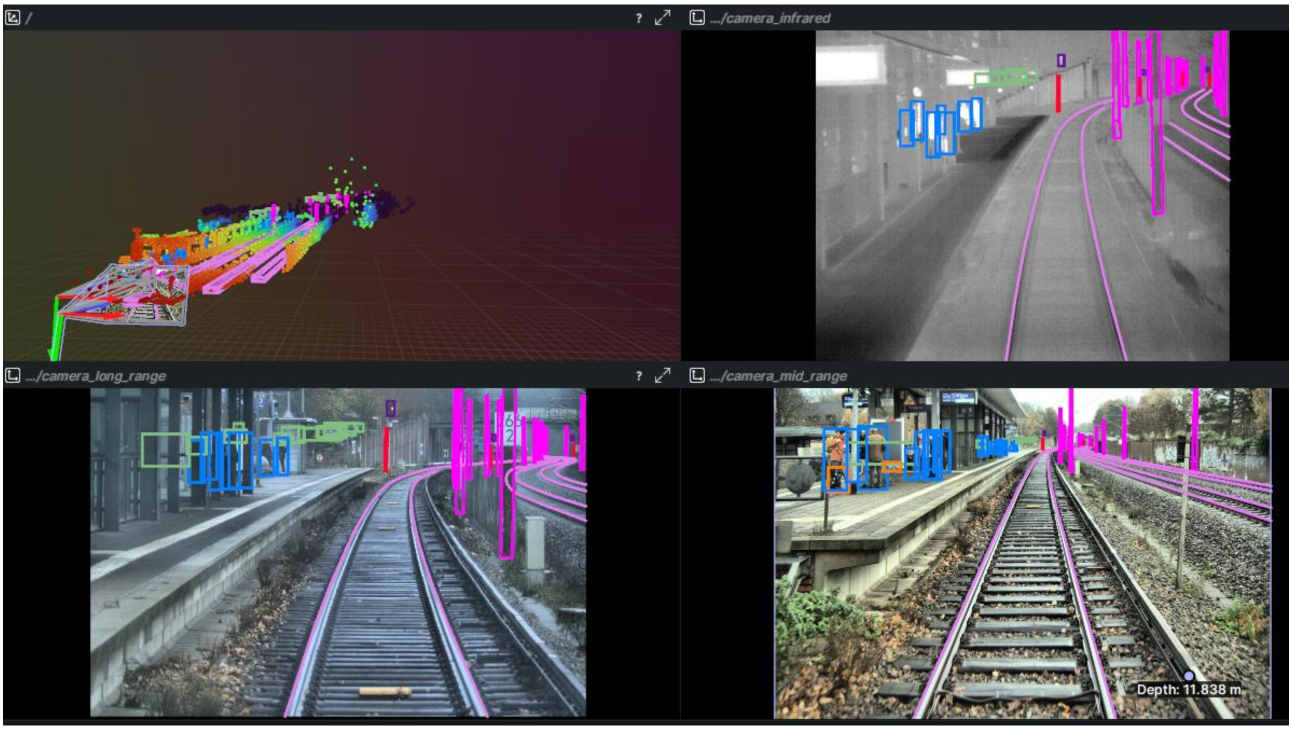


Figure 4 Annotated Scene: Morning Platform View

**Sequence ID:** 2022-12-09_09-48-40__2022-12-09_09-50-39+12
**Duration:** 15 seconds
**Sensor Modalities:** RGB (long-range, mid-range), Infrared, LiDAR
**Environment:** Daytime, station platform, multiple persons and infrastructure elements visible

#### 8.1.1.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Infrared Camera** | Bbox, Poly2D | person, road_vehicle, signal, catenary_pole, signal_pole, track |
| **Long-Range RGB Camera** | Bbox, Poly2D | person, road_vehicle, signal, catenary_pole, signal_pole, track |
| **Mid-Range RGB Camera** | Bbox, Poly2D | bicycle, person, personal_item, road_vehicle, signal, catenary_pole, signal_pole, track |
| **LiDAR (lidar_merged)** | Cuboid, Seg3D | person, road_vehicle, signal, signal_pole, catenary_pole, track, personal_item, bicycle |

#### 8.1.1.2 Technical Highlights

- **Annotation Scope:**
  - All dynamic and static objects in the railway environment are annotated.

  - LiDAR annotations include both **cuboid volumes** and **3D segmentations** for enriched spatial context.
  - The railway track is marked across all modalities using 2D polylines or 3D segmentation.
  - The mid-range RGB camera additionally captures fine-grained details such as bicycle and personal_item.
- **Object Tracking:**
  - Uniform object IDs across sensors enable temporal and spatial consistency in multimodal annotation.
- **Environmental Characteristics:**
  - The scene shows a well-lit platform environment with several persons near the tracks and visible infrastructure (e.g., signals, poles).
  - Vehicles and smaller personal items (e.g., luggage) are annotated where visible in LiDAR data.
- **Modality-Specific Notes:**
  - *Infrared:* Thermal contrast aids in the identification of persons under partial occlusion.
  - *Mid-Range RGB:* Offers rich detail, supporting bounding box and polygon annotations for a broad range of object types.
  - *LiDAR:* High-resolution point cloud enables depth estimation and geometric annotation of non-human objects like bicycles.

### 8.1.2 Scene Profile: Night Station Scene

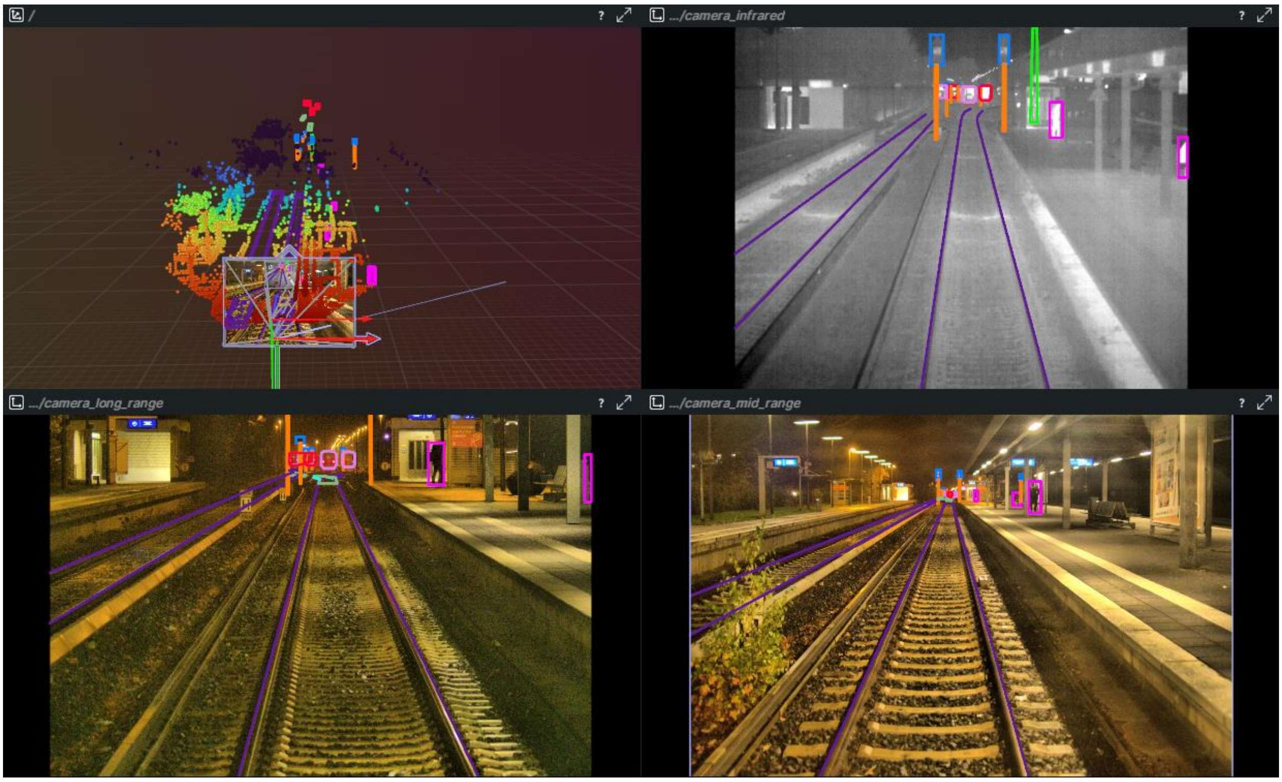


Figure 5 Annotated Scene: Night Station Scene

**Sequence ID:** 2022-11-30_23-29-58__2022-11-30_23-52-39+13
**Duration:** 13 seconds
**Sensor Modalities:** RGB (long-range, mid-range), Infrared, LiDAR
**Environment:** Nighttime, illuminated station, trains and infrastructure prominently visible

#### 8.1.2.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Infrared Camera** | Bbox, Poly2D | person, signal, catenary_pole, signal_pole, track, train, train_front |
| **Long-Range RGB Camera** | Bbox, Poly2D | person, signal, signal_pole, switch, track, train, train_front, transition |
| **Mid-Range RGB Camera** | Poly2D only | catenary_pole, signal_pole, switch, track, train, train_front |
| **LiDAR (lidar_merged)** | Cuboid, Seg3D | person, signal, signal_pole, catenary_pole, switch, track, train |

#### 8.1.2.2 Technical Highlights

- **Annotation Scope:**
    - This scene includes extensive railway infrastructure: multiple trains, tracks, signals, switches, and poles.
    - Infrared data supplements RGB by revealing persons and objects under reduced visibility.
    - Train fronts are distinctly annotated across all modalities (where applicable), supporting orientation analysis.
    - The inclusion of transition in RGB annotations suggests temporal transitions (e.g., door status, light states).
- **Cross-Modality Object Continuity:**
    - train_0000 through train_0003 and matching train_front annotations appear in all relevant modalities, allowing detailed geometric comparison and sensor fusion.
    - Object IDs are consistent across RGB, LiDAR, and infrared, facilitating time-sequence annotation and tracking.
- **Environmental Characteristics:**
    - Despite being captured at night, the scene benefits from strong artificial lighting along the platform.
    - Persons are sparse but identifiable; signals and train contours are sharply defined due to reflective elements.
- **Sensor-Specific Notes:**
    - *Infrared:* Shows clear outlines of the trains and platform features despite limited ambient light.
    - *LiDAR:* Delivers accurate 3D structure, useful for object localization and classification under challenging lighting conditions.
    - *RGB:* Both long- and mid-range images show excellent resolution and clarity, allowing dense annotation of railway infrastructure.

### 8.1.3 Scene Profile: Crowd at Platform

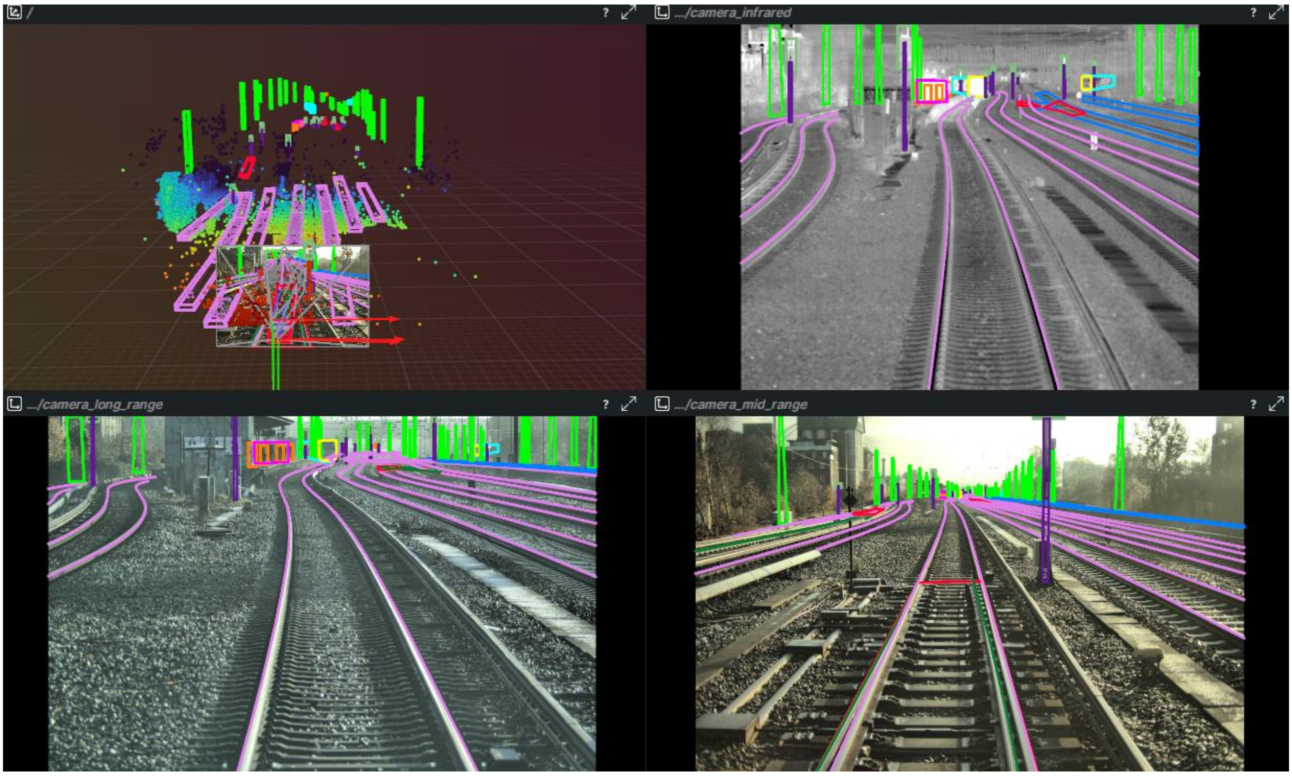


Figure 6 Annotated Scene: Crowd at Platform

**Sequence ID:** 2022-12-09_09-48-40__2022-12-09_09-51-15+12
**Duration:** 11 seconds
**Sensor Modalities:** RGB (long-range, mid-range), Infrared, LiDAR
**Environment:** Daytime, platform view, stationary train, large group of people

#### 8.1.3.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Infrared Camera** | Bbox, Poly2D | person, signal, catenary_pole, signal_pole, switch, track, train, train_front, crowd, ignoreTracks |
| **Long-Range RGB Camera** | Bbox, Poly2D | person, signal, catenary_pole, signal_pole, switch, track, train, train_front, crowd, ignoreTracks, transition |
| **Mid-Range RGB Camera** | Bbox, Poly2D | person, signal, catenary_pole, signal_pole, switch, track, train, train_front, crowd, ignoreTracks, transition |
| **LiDAR (lidar_merged)** | Cuboid, Seg3D | person, signal, catenary_pole, signal_pole, switch, track, train, transition, crowd |

#### 8.1.3.2 Technical Highlights

- **Annotation Focus:**

  - This scene centers on a high-density passenger presence (crowd) on the platform while a **train remains stationary**.
  - Crowd elements are annotated both in 2D (via Poly2D) and 3D (via LiDAR Cuboids), reflecting the complexity of the social scene.
  - Detailed infrastructure elements (e.g., switch, transition, ignoreTracks) are consistently marked for geometry-based model training.
- **Multimodal Consistency:**
  - All major elements—particularly the train and its front—are coherently labeled across modalities with identical object IDs.
  - Persons within the crowd are labeled individually and as grouped entities, enabling flexible use in detection and segmentation models.
- **Environmental Context:**
  - Daylight conditions support high annotation quality in RGB and IR imagery.
  - The scene includes long view corridors with high track complexity—multiple rails, switches, and signal poles are annotated.
  - Crowd distribution is captured particularly well on the long-range view, showing groups waiting or moving on the platform.
- **Modality-Specific Notes:**
  - *Infrared:* Useful for differentiating individual heat signatures within dense groups.
  - *RGB:* Excellent spatial resolution enables high-precision polygonal annotations for both infrastructure and people.
  - *LiDAR:* Enables 3D geometry extraction even in dense environments with overlapping individuals.

### 8.1.4 Scene Profile: Incoming Train with Trackside Passengers

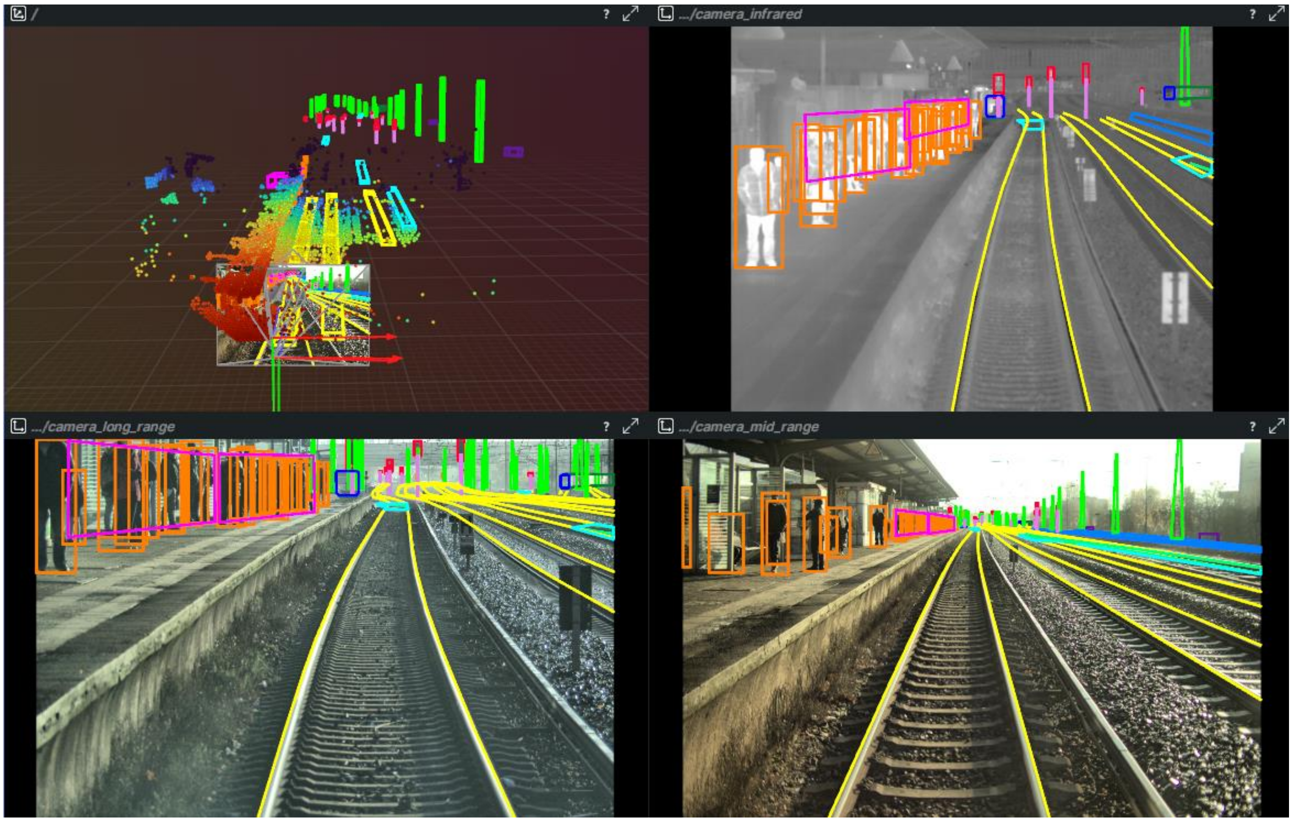

Figure 7 Annotated Scene: Incoming Train with Trackside Passengers

**Sequence ID:** 2022-12-09_09-48-40__2022-12-09_09-50-39+12
**Duration:** 12 seconds
**Sensor Modalities:** RGB (long-range, mid-range), Infrared, LiDAR
**Environment:** Daytime, passengers waiting on platform, train arriving

#### 8.1.4.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Infrared Camera** | Bbox, Poly2D | person, signal, catenary_pole, signal_pole, switch, track, train, train_front, transition, crowd, ignoreTracks |
| **Long-Range RGB Camera** | Bbox, Poly2D | person, signal, catenary_pole, signal_pole, switch, track, train, train_front, transition, crowd |
| **Mid-Range RGB Camera** | Bbox, Poly2D | person, road_vehicle, pram, signal, catenary_pole, signal_pole, switch, track, train, train_front, transition, crowd, ignoreTracks |
| **LiDAR (lidar_merged)** | Cuboid, Seg3D | person, road_vehicle, pram, signal, catenary_pole, signal_pole, switch, track, train, transition, crowd |

### 8.1.4.2 Technical Highlights

- **Scene Dynamics:**
  - This scene captures the dynamic event of a **train arriving** at a platform where passengers are waiting.
  - The presence of a **pram** is annotated in both mid-range RGB and LiDAR, marking a critical use case for safety-related perception.
  - Crowd density is moderate to high along the platform edge.
- **Annotation Specifics:**
  - All train-related elements (train, train_front, transition) are annotated with multiple IDs for distinct cars or segments.
  - The crowd class is used in all sensor types to represent grouped passenger formations.
  - ignoreTracks are annotated to exclude visually misleading or inactive rails from learning and evaluation.
- **Sensor Insights:**
  - *Infrared:* Effective in highlighting human presence near the platform under shadows or partial occlusions.
  - *RGB (Long/Mid-Range):* High-resolution imagery provides detailed annotations of both dynamic actors and railway infrastructure.
  - *LiDAR:* Accurate cuboid and segmentation data for all structural elements and passengers, enhancing 3D modeling of the scene.
- **Infrastructure Complexity:**
  - The scene includes numerous parallel tracks, switches, and signaling elements, creating a complex but well-annotated sensor environment.
  - Several transition states are captured (possibly related to train doors or light changes), especially in RGB and LiDAR.

### 8.1.5 Scene Profile: Platform with customers (SNCF)

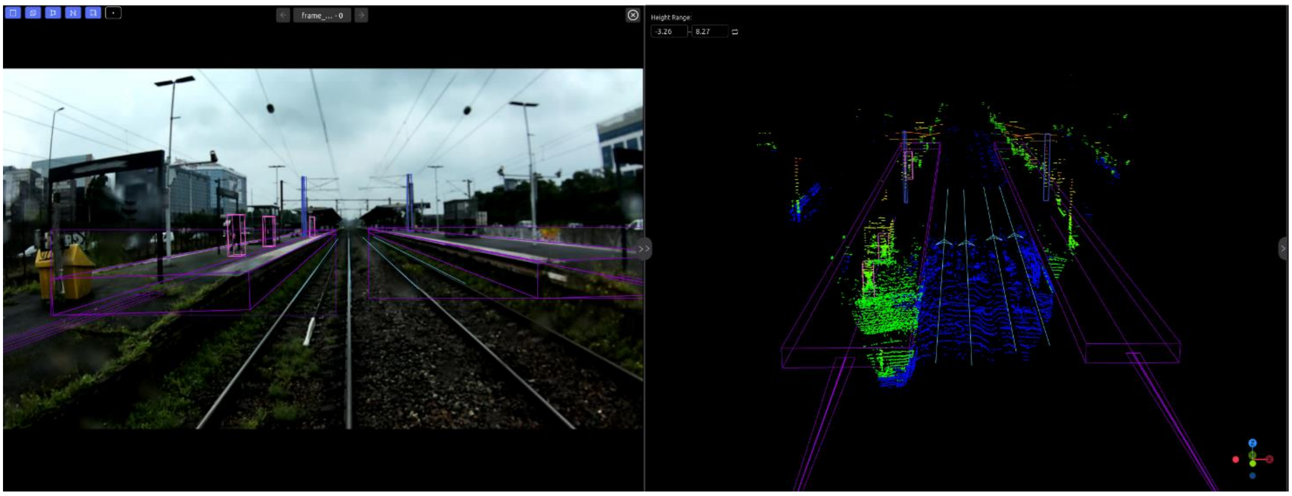

Figure 8 Annotated Scene: Platform with Customers

**Sequence ID:**
**Duration:** 11 seconds
**Sensor Modalities:** RGB (low_light), LiDAR
**Environment:** Daytime, small drops on the lens, passengers waiting on platform

#### 8.1.5.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Low-Light RGB Camera** | Bbox, Poly2D,Bbox3D | person, catenary_pole, track, |
| **LiDAR** | Cuboid, Poly3D | person, catenary_pole, track, |

#### 8.1.5.2 Technical Highlights

- **Scene Dynamics:**
  - Crowd density is low along the platform edge.
- **Sensor Insights:**
  - *RGB (Long/Mid-Range):* High-resolution imagery provides detailed annotations of both dynamic actors and railway infrastructure.
  - *LiDAR:* Accurate cuboid and segmentation data for all structural elements and passengers, enhancing 3D modelling of the scene.
- **Infrastructure Complexity:**
  - The scene includes 2 parallel tracks.

### 8.1.6 Scene Profile: Shunting yard arrival (SNCF)

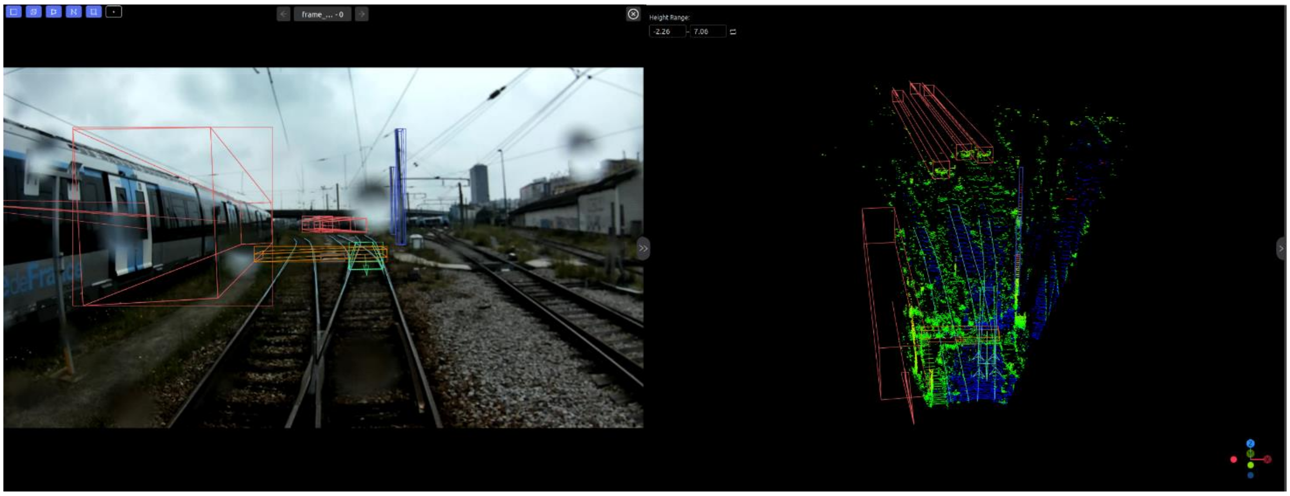

Figure 9 Annotated Scene: Shunting yard arrival

**Sequence ID:**
**Duration:** 70 seconds
**Sensor Modalities:** RGB (low-light), LiDAR
**Environment:** Daytime, multiple tracks and a switch on front of the train, train stopped on the left, rain drops on the lens

#### 8.1.6.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Low-Light RGB Camera** | Bbox, Bbox3D, Poly2D | catenary pole, switch, track, flouding point, train |
| **LiDAR** | Cuboid, Poly3D | catenary pole, switch, track, flouding point, train |

#### 8.1.6.2 Technical Highlights

- **Annotation Specifics:**.
  - Switch is annotated with a cuboid in the point cloud from for tip of the spinc tot he heart of the switch. Flouding point is also annotated with a Cuboid
- **Sensor Insights:**
  - *RGB (low-light):* High-resolution imagery provides detailed annotations of both dynamic actors and railway infrastructure.
  - *LiDAR:* Accurate cuboid and segmentation data for all structural elements enhancing 3D modelling of the scene.
- **Infrastructure Complexity:**
  - The scene includes numerous parallel tracks, switches,

### 8.1.7 Scene Profile: Crossing train (SNCF)

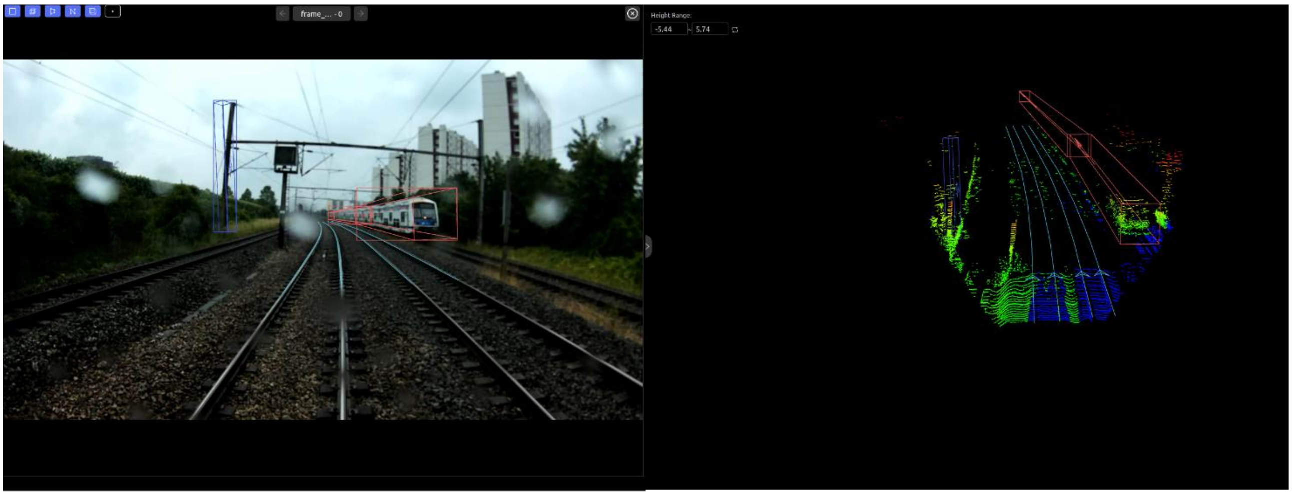

Figure 10 Annotated Scene: crossing train

**Sequence ID: Duration:** 8 seconds
**Sensor Modalities:** RGB (low_light), LiDAR
**Environment:** Daytime, passengers waiting on platform, train arriving, rain drops on the lens

#### 8.1.7.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Low-Light RGB Camera** | Bbox, Bbox3D, Poly2D | catenary pole, train, |
| **LiDAR** | Cuboid, Seg3D | catenary pole, train, |

#### 8.1.7.2 Technical Highlights

- **Scene Dynamics:**
  - This scene captures the dynamic event of a **train arriving**
- **Sensor Insights:**
  - *RGB (low-light):* High-resolution imagery provides detailed annotations of both dynamic actors and railway infrastructure.
  - *LiDAR:* Accurate cuboid and segmentation data for all structural elements and passengers, enhancing 3D modelling of the scene.
- **Infrastructure Complexity:**
  - The scene includes 2 parallel tracks, a crossing train coming on the right.

### 8.1.8 Scene Profile: Plain line (SNCF)

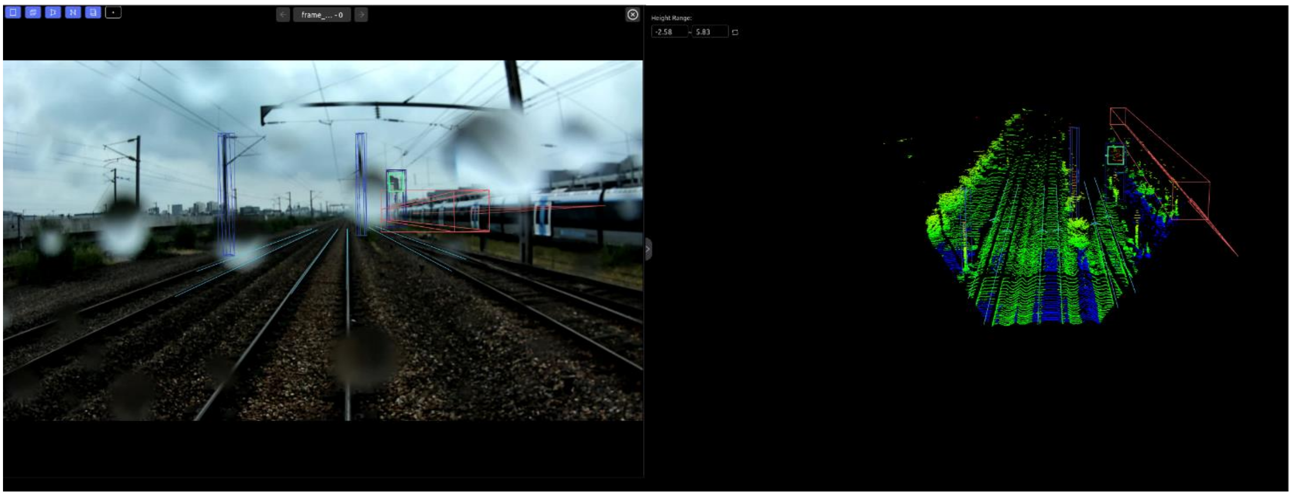

Figure 11 Annotated Scene: Plain line

**Sequence ID:**
**Duration:** 35 seconds
**Sensor Modalities:** RGB (low_light), LiDAR
**Environment:** Daytime, rain drops on the lens

#### 8.1.8.1 Sensor Views and Object Visibility

| Sensor | Annotation Types | Detected Object Classes |
|---|---|---|
| **Low-Light RGB Camera** | Bbox, Bbox3D, Poly2D | catenary pole, track, train |
| **LiDAR** | Cuboid, Seg3D | catenary pole, track, train |

#### 8.1.8.2 Technical Highlights

- **Sensor Insights:**
  - *RGB (Low-light):* High-resolution imagery provides detailed annotations of both dynamic actors and railway infrastructure.
  - *LiDAR:* Accurate cuboid and segmentation data for all structural elements and passengers, enhancing 3D modeling of the scene.
- **Infrastructure Complexity:**
  - The scene includes numerous parallel tracks, switches, and signalling elements, creating a complex but well-annotated sensor environment.

## 8.2 Summary of Annotated Scenes and Selection Rationale

To support the development and evaluation of perception systems for autonomous railway operation, a comprehensive set of scenes was annotated across multiple sensor modalities, including RGB, infrared, and LiDAR. The selection comprises sequences contributed by **Deutsche Bahn (DB)** and **SNCF**, each chosen to reflect operationally relevant, perceptually challenging, and environmentally diverse situations.

Together, these scenes capture a wide spectrum of use cases—from routine platform observations to complex operational dynamics and adverse weather effects—ensuring the resulting dataset provides a robust foundation for training and benchmarking machine learning models.

### 8.2.1 Scenes Provided by Deutsche Bahn

- **Morning Platform View**: A low-density daytime scenario used for calibration and baseline evaluation. The environment is well lit, with limited crowd presence and clear visibility of infrastructure elements.
- **Night Station Scene**: Captured under artificial illumination at night, this sequence tests model robustness under poor visibility and includes annotated trains, tracks, and signaling infrastructure.
- **Crowd at Platform**: A high-density scene with a stationary train and numerous individuals along the platform. It supports detailed modeling of person detection, occlusion, and social grouping.
- **Incoming Train with Passengers at Platform**: A dynamic scenario showing a train arriving while passengers wait. This scene targets perception challenges associated with movement, crowd dynamics, and transitional events.

Each DB scene includes extensive multimodal annotations, cross-sensor object continuity, and infrastructure-rich views to cover a wide range of technical and contextual requirements.

### 8.2.2 Scenes Provided by SNCF

SNCF contributed a complementary set of scenes, with a particular emphasis on **rain-related visual disturbances**, such as **raindrops on the camera lens**, **reduced contrast**, and **surface reflections**. These artefacts were deliberately retained, as they reflect common real-world conditions that perception systems must be able to handle reliably.

- **Platform with Customers**: A daytime sequence with light rain and scattered individuals on the platform. Despite minor lens artefacts, critical infrastructure and dynamic entities remain visible and well annotated.
- **Shunting Yard Arrival**: Featuring a stopped train in a yard environment, the scene includes annotated switches and multiple track segments. The presence of rain and complex track geometry adds to its technical value.
- **Crossing Train**: This brief sequence depicts a train crossing on an adjacent track while rain affects visibility. The combination of occlusion and reduced image clarity makes it a meaningful inclusion for robustness testing.
- **Plain Line**: A longer sequence showing mainline tracks under steady rain conditions. The scene includes distant infrastructure, parked trains, and several parallel track segments, supporting long-range detection tasks.

These SNCF scenes extend the dataset's environmental realism by exposing perception models to imperfect visual inputs. All recordings are annotated using the same multimodal scheme as DB, ensuring consistency and cross-comparability.

### 8.2.3 Summary

The combined scene selection reflects a deliberate balance between structured, controlled recordings and organically complex, weather-affected environments. Each scenario contributes uniquely in terms of object diversity, spatial arrangement, illumination conditions, and contextual relevance. This enables the dataset to serve both as a training basis for model development and as a benchmarking reference across varied use cases in autonomous railway perception.

The annotated scenes described above form the empirical backbone for the perception-related tasks addressed in subsequent stages of the R2DATO project. Their diversity—in terms of sensor coverage, environmental conditions, and operational complexity—ensures that they are suitable not only for targeted algorithm development but also for systematic evaluation of model performance.

In the following chapter, we examine the concrete applications of this dataset within the **machine learning pipeline**, with particular emphasis on how the annotated data supports **training, validation**, and **performance analysis** of perception models developed in the framework of Deliverable 7.5.

# 9 CONCLUSIONS

Deliverable 7.4 of the R2DATO project constitutes a critical building block within the overarching Data Factory initiative under Work Package 7. Its central aim—to jointly collect and annotate real-world sensor data for the development of perception systems in autonomous railway operation—has been fulfilled with a high degree of technical rigor and collaborative efficiency. This deliverable represents a foundational step toward enabling scalable, interoperable, and resilient AI-based perception solutions in the European railway sector.

**Assessment of Objectives and Contributions**

The key objectives defined at the outset of D7.4 have been fully addressed and operationalised:

- **Joint field data collection** was successfully executed by Deutsche Bahn and SNCF in various operational environments. Using multi-sensor platforms (RGB, infrared, LiDAR), the partners recorded diverse railway scenarios, ranging from dense platform crowds and station arrivals to shunting yards and open-track conditions. Special attention was given to scenes exhibiting environmental complexities, such as night-time visibility limitations and weather-induced artefacts like raindrops on lenses.
- **Annotation of sensor data** was carried out using a combination of manual and semi-automated methods, conforming strictly to the ASAM Open Label data standard. The annotated objects cover a broad spectrum, including persons, trains, infrastructure elements (signals, catenary poles, switches), and dynamic scene components. Annotations were produced consistently across 2D and 3D modalities and verified through quality assurance protocols based on defined KPIs.
- **Interoperability and integration** were ensured through adherence to common data formats and shared annotation semantics, allowing seamless use in downstream components of the R2DATO Data Factory, including the training of AI perception models in D7.5 and the public release of a representative dataset in D7.6.

By transforming raw, real-world sensor data into structured, annotated, and technically validated datasets, D7.4 has created a reusable asset that adds long-term value to research, development, and deployment activities within and beyond the R2DATO framework.

**Key Insights and Significance**

One of the most significant achievements of D7.4 lies in the **realism and variability** of the data it provides. Unlike many existing datasets in the mobility domain—which often rely heavily on synthetic data or curated conditions—these deliverable embraces operational imperfections. The deliberate inclusion of weather artefacts, lens occlusions, and visually degraded scenarios extends the dataset’s utility for training models that must operate under real-world uncertainty.

Furthermore, the deliverable demonstrates that **cross-organisational collaboration** in data acquisition and annotation is feasible when supported by harmonised methodologies, shared infrastructure, and iterative validation. DB and SNCF managed to jointly develop a coherent dataset despite differences in national infrastructure, rolling stock, and data collection conditions.

From a technological perspective, the deliverable underscores the **importance of multimodality** in perception development. The coordinated use of RGB, infrared, and LiDAR sensors—together with consistent object IDs and temporal alignment—enables richer data fusion and opens pathways for advanced AI techniques such as sensor fusion learning, temporal tracking, and domain adaptation.

**Challenges, Limitations, and Lessons Learned**

Despite the success of the deliverable, several **practical and methodological challenges** were encountered during its implementation:

- **Environmental unpredictability** posed a challenge during recording campaigns. Factors such as lighting changes, weather variability, and background activity introduced noise into the data and occasionally affected annotation precision.
- The **manual annotation process** proved labor-intensive and time-consuming, particularly for complex 3D objects in LiDAR point clouds. While semi-automated tools were employed, the need for human validation remained substantial. This underscores the urgency of developing more scalable annotation pipelines in future efforts.
- **Sensor calibration and cross-modal alignment** required careful tuning and validation. Maintaining temporal and spatial consistency between modalities (e.g., synchronising LiDAR cuboids with RGB bounding boxes) demanded substantial technical effort and repeated calibration iterations.
- **Standard harmonisation** across partners was an initial hurdle. Differences in annotation tools, sensor setups, and interpretation of data schemas necessitated close coordination and consensus-building in the early phases of the work.

These challenges were addressed through biweekly consortium synchronisation meetings, joint validation sessions, and rigorous documentation practices. The lessons learned from these processes will serve as a valuable foundation for subsequent tasks in the R2DATO project and for other Europe’s Rail initiatives focused on data interoperability and standardisation.

**Open Points and Recommendations for Future Work**

While D7.4 can be considered **complete in its current scope**, several opportunities for further enhancement and follow-up activities have emerged:

- **Geographical extension** of the dataset could further improve its representativeness. Including recordings from additional railway operators—particularly from Central and Eastern Europe—would enhance generalisability and facilitate pan-European adoption of trained models.
- **Increased automation of annotation processes** is necessary to reduce human workload and improve scalability. This could include the integration of weak supervision, active learning, or annotation transfer from synthetic to real-world data.
- **Coupling perception data with control and decision-making systems** represents a logical next step. The real-world data collected in D7.4 could be used to simulate or validate perception-driven decision chains in autonomous train operation.
- **Feedback loops with AI model training** (under D7.5) should be formalised. Annotation refinements informed by model performance metrics could further improve the dataset’s quality and effectiveness.
- **Standardisation and regulation**: The use of the ASAM Open Label format positions the deliverable within a broader European standardisation effort. Future work should explore contributions to emerging standards and ensure compatibility with cross-sector initiatives in automated mobility.
- **Final Reflections**

Deliverable 7.4 demonstrates that the joint collection and annotation of high-quality, multi-sensor railway data is both technically feasible and operationally valuable. It not only fulfils the objectives set out in the Grant Agreement but also exceeds them by incorporating diverse and realistic conditions, fostering cross-border collaboration, and laying the groundwork for reusable, standardised datasets in the railway AI domain.

By translating raw sensor streams into structured, annotated data, this deliverable bridges the gap between observation and perception—making it a cornerstone of the R2DATO Data Factory and a key enabler for the next generation of autonomous train systems across Europe.

# REFERENCES

(2025, May 02). Retrieved from Automated Quality Assurance for Object Detection Datasets: docs.ultralytics.com

(2025, May 02). Retrieved from Automated QA, Review & Honeypots: docs.cvat.ai

(2025, May 02). Retrieved from encord-team/encord-active: The toolkit to test, validate, and ...: https://github.com/encord-team/encord-active

(2025, May 02). Retrieved from Data Collection and Annotation - Ultralytics YOLO Docs: docs.ultralytics.com

DB InfraGO AG. (2025, 05 02). Retrieved from RailLabel auf GitHub: https://github.com/DSD-DBS/raillabel

Huan, S. (2025, May 2). *Cooperative Holistic Scene Understanding: Unifying.* Retrieved from https://proceedings.neurips.cc/paper_files/paper/2018/file/82161242827b703e6acf9c726942a1e4-Paper.pdf#:~:text=2%C3%974%20is%20the%202D%20object,room%20layout%20LPHY%20%3D%201

Neumaier, P., & Dubiel, S. (2024). *D7.1 Requirements and specifications of the Data.* Berlin, München.

Neumaier, P., & Dubiel, S. (2024). *D7.3 Perform simulation with initial.* Berlin, München.