ERJU FP2 WP27 Digital Register Specification, Development, and Implementation

D27.5 – Digital Register Object Catalogue for the Data Factory

Due date of deliverable: 15/05/2025

Actual submission date: 09/04/2025

Leader/Responsible of this Deliverable: Danny Kreyenberg, DB InfraGo

Reviewed: Y

Document status		
Revision	Date	Description
01	24/02/2025	First Draft for review within WP27
02	09/04/2025	Version after processing WP27 review comments
03	19/02/2026	Version after processing JU Review comments
04	26/05/2026	Final version after processing JU Review comments

Project funded from the European Union's Horizon Europe research and innovation programme		
Dissemination Level		
PU	Public	X
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
Thomas Th Renner	DB InfraGO	Author
Magdalena Pech	DB InfraGO	Contributor
Harish Narayanan	Nextrail	Contributor
Lubor Bazant	AZD	Reviewer
Susana Herranz de Andrés	CEDEX	Reviewer

EXECUTIVE SUMMARY

The integration of Digital Register's static infrastructure data with Data Factory annotated scenes offers significant advancements in AI training for railway operations. The Digital Register provides topology and 3D static infrastructure data, detailing static infrastructure objects such as track axes, platforms, catenary poles, and signs. The Data Factory, on the other hand, generates annotated multi-sensor data focused on dynamic objects like people, animals, bicycles, and smoke, supporting AI models for sensor-based perception and obstacle detection.

By merging these datasets, AI training gains improved contextual understanding, allowing for more realistic scenario modelling and better differentiation between static and dynamic objects. This integration enhances scene identification, validation accuracy, and object correlation, ultimately improving AI-driven railway safety systems.

As part of Task 27.3, "Digital Register Implementation for Data Factory," we explored the feasibility and advantages of combining static infrastructure data with Data Factory annotated scenes at Havelländische Eisenbahn (HVLE) [2] in Berlin from the Automated Train project [1]. The delivery highlights how this combination can enrich datasets, enabling more robust AI training and advancing railway automation. The focus was on demonstrating the potential benefits of such integration for enhancing AI-driven perception and safety systems in rail environments

ABBREVIATIONS AND ACRONYMS

DR	Digital Register
ERJU	Europe’s Rail Joint Undertaking
EUG	ERTMS Users Group
FA	Flagship Area
FP	Flagship Project
IP	Innovation Pillar
R2DATO	Rail to Digital and Automated Train Operation
RCA	Reference CCS Architecture
WP	Work Package
AI	Artificial Intelligence
HVLE	Havelländische Eisenbahn

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The Digital Register is providing static infrastructure data for digitalized railway operations. The data provided can be divided into topology and 3D static infrastructure data. The latter includes highly precise localized static infrastructure objects such as track axes, platforms, catenary poles, and signs.

A Data Factory is an enabler for systematic provision of large amounts of data mainly for the training of Artificial Intelligence (AI) software. For instance, AI training of the train sensor-based perception and obstacle detection systems observing the railway environment. Therefore, a Data Factory provides scenes of annotated multi-sensor data. Compared to static infrastructure data, the focus of the annotated scenes is more on dynamic objects that can have an impact on railway operations. Such annotated objects include persons, animals, bicycles, scooters, poles, flame or smoke.

Integrating Digital Register's static infrastructure data with the annotated scenes from the Data Factory provides new opportunities for creating more advanced datasets for AI training. By merging static infrastructure data with dynamic object annotations, this approach has the potential to deliver more detailed and accurate training data, supporting the development of more dependable AI models for the railway industry.

Identify and Enrich Relevant Scenes for AI Training

The integration of static infrastructure data into the Data Factory enhances the contextual understanding of annotated scenes. For instance, determining whether an annotated person is standing on a platform or on a track becomes more precise with the inclusion of Infrastructure data. By this, AI models are trained with more realistic scenarios, improving their ability to handle complex railway environments.

Static infrastructure data as Ground Truth or Secondary Validation Source

Static infrastructure data serves as a reliable ground truth or secondary validation source for annotated objects in the Data Factory. For example, verifying whether an annotated pole corresponds to a catenary pole is streamlined with infrastructure data. This cross-referencing capability minimizes annotation errors and ensures that the AI systems being trained can differentiate between similar looking objects with confidence.

Enhanced Dynamic and Static Object Correlation

Integrating static infrastructure data bridges the gap between static infrastructure and dynamic object annotations. This correlation is essential for scenarios where the interaction between moving objects (e.g., bicycles, scooters, or animals) and static infrastructure (e.g., platforms or tracks) impacts railway operations. AI models trained on such enriched datasets can better predict and mitigate potential hazards.

The objective of Task 27.3, "Digital Register Implementation for Data Factory," is to showcase the opportunities and applications of integrating static infrastructure data with annotated scenes from the Data Factory. To demonstrate this potential, static infrastructure data were created for selected Data Factory scenes. Both the static infrastructure data and annotated scenes were independently collected and generated on railway premises and tracks owned by the Havelländische Eisenbahn (HVLE) [2] in Berlin as part of the Automated Train project [1].

2 COMPARISON OF DATA FACTORY AND STATIC INFRASTRUCTURE DATA

Following table shows a list of annotated objects of the Open Sensor Data for Rail 2023 [3], which is used as a reference of annotated objects in the Data Factory[4]. For each element in the table, it is written down if it's part of the Automated Train[1] static infrastructure data object catalogue.

Table 1: Comparison of data factory and static infrastructure data

Data Factory Annotations	Part of the static infrastructure data
Person	No
Crowd	No
Train	No
Wagons	No
Bicycle	No
Group of bicycles	No
Wheelchair	No
Motorcycle	No
Road Vehicle	No
Animal	No
Group of Animals	No
Drag shoe	No
Track	Yes
Transition	No
Switch	Yes
Catenary Pole	Yes
Signal Pole	Yes
Signal	Yes
Signal Bridge	Yes
Buffer Stop	Yes
Flame	No
Smoke	No

Automated Train [1] static infrastructure data objects are characterized by a high degree of accuracy and structural detail. In general, each object is modelled as a composite of multiple components. Each component features a geometric primitive describing its spatial extent.

In the following sections, the formats of such objects are described. Furthermore, we focus on objects that also occur as annotations in the Data Factory.

Table 2: Specification of the static infrastructure data object: Track

<table>
<tr><td>Object</td><td>Track (i.e. Rails)</td></tr>
<tr><td>Definition</td><td>Rails are linear load-bearing and guiding elements that are arranged in pairs and parallel to each other at a certain distance of the track gauge.</td></tr>
<tr><td>Components</td><td>
<table>
<tr><td>Component</td><td>Definition</td><td>Relation</td><td>Geometry</td></tr>
<tr><td>Main Rail</td><td>A line giving the inner upper edge of the left and the right rail - given as a pair.

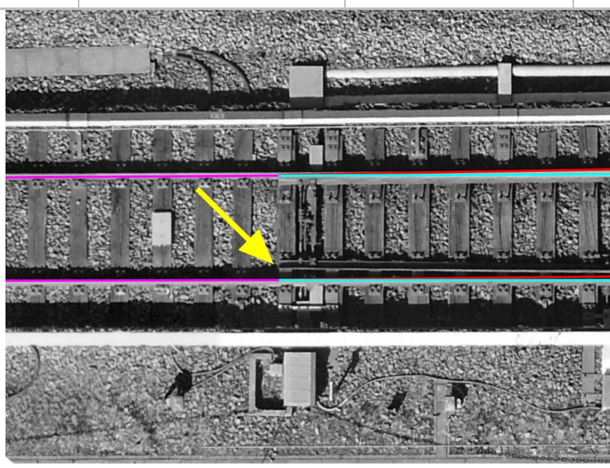
Rails are only split at switches.</td><td>2..2</td><td>Polyline</td></tr>
<tr><td>Center Line</td><td>The Centerline is defined as the line calculated from the middle of the inner upper edge of the left and the right rail.</td><td>0..1</td><td>Polyline</td></tr>
</table>
</td></tr>
<tr><td>Sample Image</td><td></td></tr>
</table>

Table 3: Specification of the static infrastructure data object: Catenary Pole

Object	**Catenary Pole**
Definition	A catenary pole is a vertical support structure that holds the overhead wire used for the electrification of rail lines.

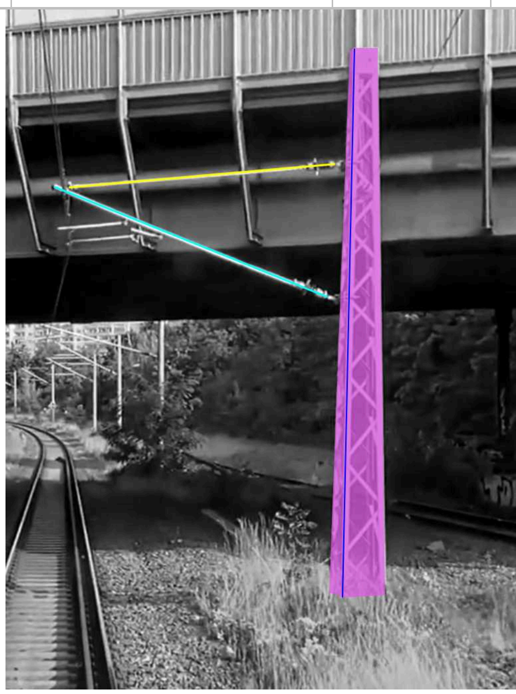
Components	Component	Definition	Relation	Geometry
	Main Pole	The main vertical structural element used to support various railway infrastructure components, such as signals, overhead lines, or bars, by elevating them to the required height for visibility or functionality.	1:1	Line - if it is a has a small diameter Quadrilateral- if it is a flat pole with a significant length Pyramid Frustum - if it is a massive pole with a significant extend in all dimensions Conical Frustum - if it is a round shape with significant extend in all dimensions
	Extension	The solid structures holding the wire of a catenary pole are given as extensions.	0:N	Line Pyramid Frustum
Sample Image				

Table 4: Specification of the static infrastructure data object: Signal Pole

Object	**Signal Pole**

<table>
<tr><td>Definition</td><td>A signal pole is a vertical structure that supports signaling equipment, such as lights, boards, or other devices, used to communicate instructions to train operators.</td></tr>
<tr><td>Components</td><td>
<table>
<tr><th>Component</th><th>Definition</th><th>Relation</th><th>Geometry</th></tr>
<tr><td>Pole</td><td>The usually vertical structural element used to support or attach to various components.</td><td>1:1</td><td>Line - if it is a has a small diameter

Quadrilateral - if it is a flat pole with a significant length

Pyramid Frustum - if it is a massive pole with a significant extend in all dimensions

Conical Frustum - if it is a round shape with significant extend in all dimensions</td></tr>
<tr><td>Bar</td><td>A horizontal or angled structural element extending from a pole, used to support and position components such as signals, overhead lines, or barriers in the appropriate location for operational effectiveness and visibility.</td><td>0:1</td><td>Line - if it is a has a small diameter

Pyramid Frustum - if it is a massive bar with a significant extend in all dimensions</td></tr>
<tr><td>Signal</td><td>A device mounted on the pole that provides visual indications - such as stop, proceed, or caution - to control and manage railway traffic.</td><td>1:N</td><td>Polygon - if it is mainly flat Pyramidal Frustum - if it is a massive structure</td></tr>
</table>
</td></tr>
</table>

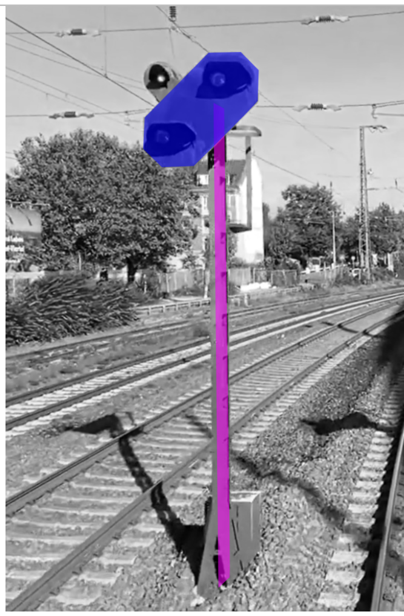
Sample Image	

Table 5: Specification of the static infrastructure data object: Buffer Stop

<table>
<tr><td>Object</td><td colspan="4">Buffer Stop</td></tr>
<tr><td>Definition</td><td colspan="4">A buffer stop is a safety device located at the end of a railroad track</td></tr>
<tr><td rowspan="2">Components</td><td>Component</td><td>Definition</td><td>Relation</td><td>Geometry</td></tr>
<tr><td>BoundingBox</td><td>A bounding box including all elements of the buffer stop.</td><td>1:1</td><td>Pyramid Frustum</td></tr>
<tr><td>Sample Image</td><td colspan="4"></td></tr>
</table>

It is important to emphasize that switches and signal bridges were not previously described, even though they are part of both initiatives. Switches in the static infrastructure data can be derived from the topology and track data object, while signal bridges do not occur within the premise selected for evaluation in Chapter 3.

In contrast to the detailed geometric breakdown found in the Digital Map, the Data Factory simplifies these representations for broader processing purposes. Within the Data Factory environment, most object of interests is abstracted and described solely as a Bounding Box, prioritizing spatial occupancy and other use cases such as object detection. Figure **1** shows an example scene in which a train is passing through a station. Signs and catenary poles are annotated with red bounding boxes, while people on the platform are annotated with blue bounding boxes. Rails are annotated with straight red lines.

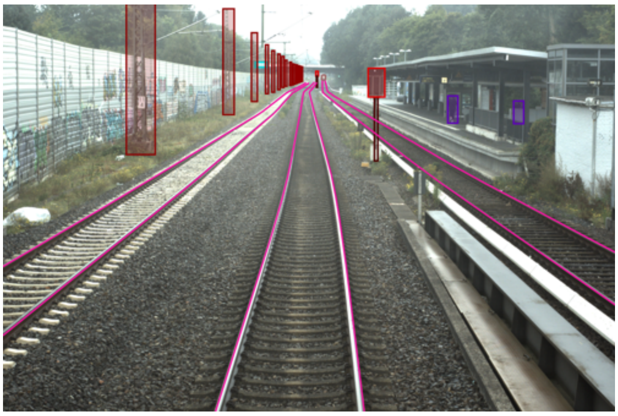


Figure 1: Example of a Data Factory scene of a train passing through a station [3]

3 RESULTS

We have created a total of nine data packets, each representing different scenes from the Data Factory. These data packets include all static infrastructure objects in the area where the train was operating in a scene. Figure 1 shows the location of the scenes, which are all located on the Havelländische Eisenbahn (HVLE) [2] premise in Berlin.

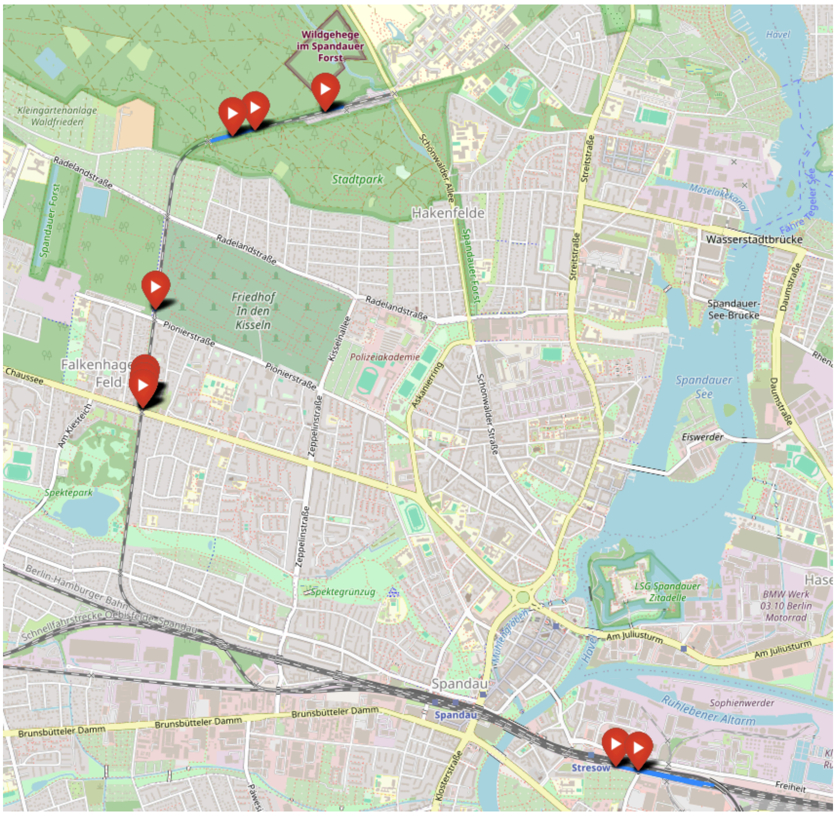


Figure 2: Measurement locations of the Data Factory scenes

Figure 2 illustrates an example of the static infrastructure data for one of the nine scenes. The data of the HVLE premise have been made fully available for this deliverable. It includes the static infrastructure data, including its JSON data schema, as well as the topology in the ERJU CCS/TMS data format. In addition, it includes the raw point cloud data used generating the infrastructure data. The inclusion of the point clouds is particularly valuable, as they provide comprehensive insights into the characteristics of the travelled track. This is highly beneficial for users working hands-on with both datasets.

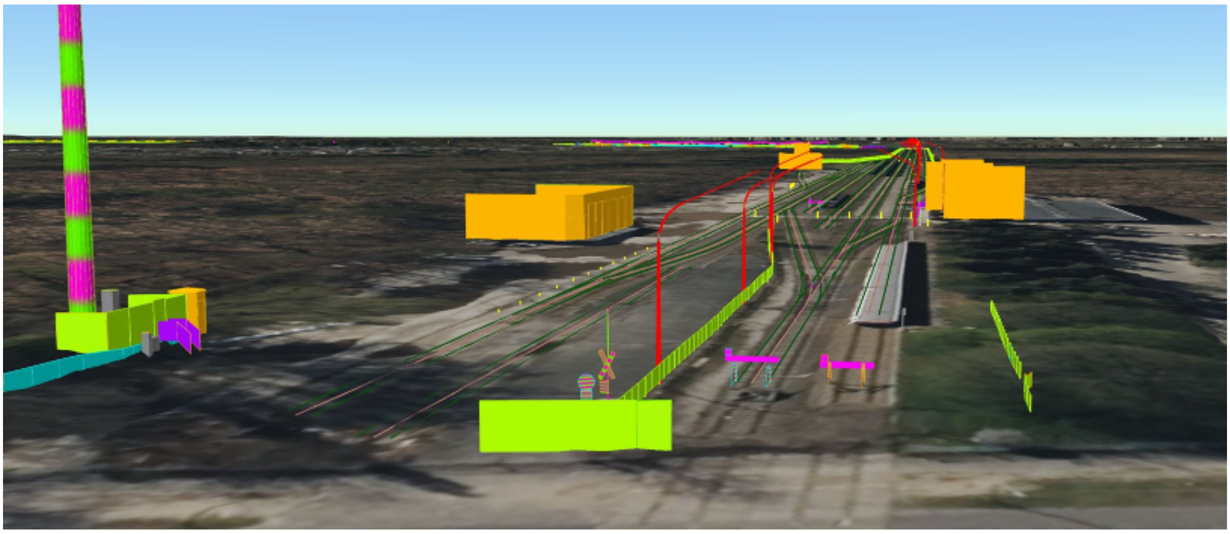


Figure 3: 3D static infrastructure data for one of the scenes from the Data Factory

4 REFERENCES

[1]. Automated Train, https://digitale-schiene-deutschland.de/en/projects/AutomatedTrain

[2]. Havelländische Eisenbahn (HVLE), https://www.hvle.de/ressourcen/infrastruktur.html

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

[4]. Data Factory, See Enabler 6.1 WP 7 GoA3/4 Data Factory specifications and Implementation in https://projects.rail-research.europa.eu/eurail-fp2/technical-enablers/

[5]. Deutsches Zentrum für Schienenverkehrsforschung (DZSF), https://www.dzsf.bund.de/