CETRA^{2026} 27-30 May 2026, Bol, Island Brač, Croatia
9th International Conference on Road and Rail Infrastructure

AUTOMATED BIM MODELS GENERATION IN RAILWAY INFRASTRUCTURE

João Ventura[1], Francisco Andrade[1], Rui Gavina[1], Ricardo Santos[1, 2], Diogo Ribeiro[1, 2]
[1] iBuilt, School of Engineering, Polytechnic of Porto, Porto, Portugal
[2]CONSTRUCT - Faculty of Engineering, University of Porto, Porto, Portugal

Abstract

The decarbonization of the railway infrastructure is a complex process that requires more than material innovation, requiring intelligent processes that incorporate sustainability and circular economy principles from the earliest design stages. This goal increases the need for interoperable digital baselines for existing railway corridors. Advances in geospatial data processing and semantic classification have opened new opportunities for automating the generation of openBIM baseline information models in railway infrastructures. This paper presents an automated workflow developed within the NRPCES-EU project that transforms open geospatial data into alignment-centered IFC 4.3 railway models. The structuring of these models follow established buildingSMART standards, ensuring consistency and interoperability across platforms. The resulting IFC provides a standard-compliant asset inventory baseline and a consistent reference backbone for linking external datasets. The proposed methodology includes parametric reconstruction and semantic enrichment by associating buildingSMART Data Dictionary concepts and Product Data Template-aligned property structures. The workflow is demonstrated on the Leixões line, an existing railway corridor in Porto, Portugal. The results indicate that IFC 4.3 baseline models can be generated with reduced manual effort, providing a scalable starting point for asset inventory and for linking external datasets within lifecycle-oriented digital workflows.

Keywords: railway infrastructure, digital modelling automation, geospatial data processing, IFC 4.3

1 Introduction

The railway infrastructure plays a central role in contemporary strategies for sustainable and low-carbon mobility. As a long-life, asset-intensive system, the railway sector faces increasing pressure to reduce environmental impacts while ensuring safety, reliability and cost-effective operation across the asset lifecycle. Addressing these challenges requires data-ready digital representations for lifecycle-oriented decision-making, particularly maintenance and performance management [1]. Digital twins have emerged as a key enabler in the railway domain, linking data-enriched virtual representations of physical assets with lifecycle monitoring. Recent research corroborates the increasing application of digital twins in railways for condition monitoring, predictive maintenance, operational optimization and safety enhancement [2]. In current practice, many railway digital twin implementations rely on manually created Building Information Modelling (BIM) models, which limits scalability for geographically extensive linear assets [3]. Consequently, digital twin initiatives frequently prioritize sensor integration and data analysis, implicitly assuming the existence of reliable base models, which is often not the case in existing railway corridors.

DOI: https://doi.org/10.5592/CO/CETRA.2026.1811

For large-scale railway projects, however, the primary sources of geometric and asset-related information are typically geospatial datasets, including track centerlines and terrain models. Interoperability between Geographic Information System (GIS) and BIM remains challenging: GIS supports corridor-scale spatial data, whereas BIM provides structured models for lifecycle use. Connecting these two domains remains a complex task, and manual translation from GIS to BIM workflows is time-consuming, error prone and difficult to scale, particularly for railway corridors, as evidenced by recent BIM-GIS integration [4]. Within this interoperability gap, the representation of railway geometry, particularly horizontal and vertical alignments, constitutes a fundamental foundation. Several studies have addressed the reconstruction of railway alignments from surveyed or GIS-derived centerline data, proposing automated constrained optimization methods capable of recovering the underlying geometric structure with high positional accuracy [5]. In parallel, geometry-driven approaches directly fit analytical or parametric representations to measured railway data, demonstrating consistently good performance for recreating existing alignments [6]. Despite their geometric robustness, these approaches typically operate independently of standardized information models and do not explicitly address the representation of reconstructed alignments within openBIM schemas, such as Industry Foundation Classes (IFC), limiting their direct application in railway digital twin environments.

The development of the IFC 4.3 schema, aligned with ISO 16739-1 [7], represents a major step towards standardized and interoperable data exchange for railway infrastructure. Nevertheless, the practical adoption of IFC 4.3 in railway projects remains limited. Early studies have demonstrated the feasibility of encoding GIS-derived centerline as IFC alignment entities through semi-manual, case study-driven workflows, highlighting limitations in automation, scalability and semantic consistency [8]. These challenges are compounded by the lack of harmonized semantic definitions for railway assets and properties. Without standardized semantic resources, IFC-based railway models often suffer from inconsistent classification and incomplete property attribution. The buildingSMART Data Dictionary (bSDD) provides consistent concept definitions and property semantics, while Product Data Templates (PDTs) define template-based property sets and value structures for information exchange [9, 10]. Together, both improve suitability of IFC-based railway models for lifecycle-oriented analyses and digital twin applications. This gap highlights the need for automated and standard-compliant workflows capable of bridging geospatial data, railway geometry reconstruction and IFC-based information modelling.

2 Methodology for automated IFC 4.3 generation from GIS data

This section presents an overview of a framework for the automated generation of IFC 4.3 railway information models from geospatial data. The proposed methodology integrates geospatial data extraction, alignment analysis, parametric reconstruction, and semantic enrichment, based on bSDD and PDTs, enabling the structured transformation of GIS inputs into a IFC 4.3 baseline. As presented in figure 1, the framework can be divided into four main stages: i) open-access data collection, ii) parametric railway alignment reconstruction, iii) IFC 4.3 automatic generation, and iv) geometric and semantic enrichment.

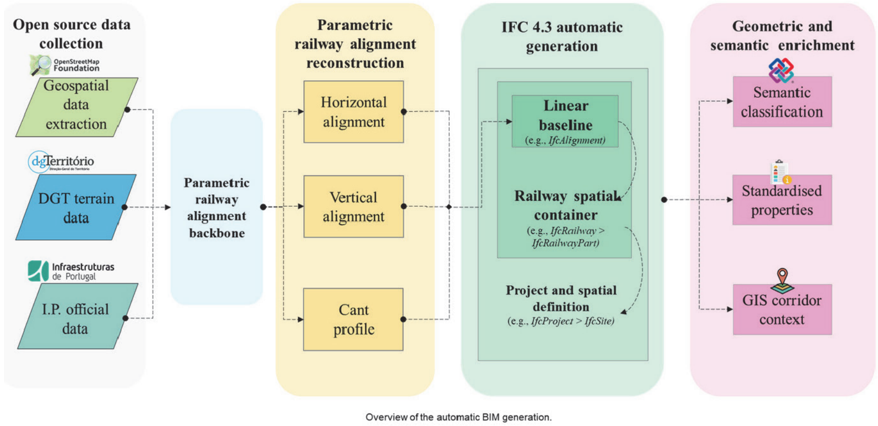

Figure 1 Overview of the methodology for automated railway IFC 4.3 generation

2.1 Open-access data collection

Geospatial data extraction is performed through a standalone automated pipeline. This strategy integrates open-access railway line data based on official reference datasets from Infraestruturas de Portugal (IP), the railway infrastructure manager, and national topographic information provided by Direção Geral do Território (DGT), the Portuguese public authority responsible for cartography and geospatial data. The purpose of this integration is to ensure geometric consistency, traceability, and reproducibility from the earliest stage of the workflow. The geometry of the railway corridor is automatically retrieved from OpenStreetMap (OSM) through automated spatial and semantic queries to the Overpass API. Rather than relying on manual procedures, the methodology identifies relevant entities and extracts the railway track geometry together with selected referenced features within the surrounding corridor. The extracted railway geometry is then transformed into a single, continuous centerline, representing the track axis. Depending on the scope of the analysis, the centerline may represent the full extent of the railway network or may be limited to a user-defined section based on station references or chainage limits. This step ensures that the resulting railway geometry defines a coherent railway alignment suitable for subsequent parametric reconstruction (section 2.2).

2.2 Parametric railway alignment reconstruction

The railway alignment is the primary geometric descriptor of the corridor and provides the backbone for IFC-based modelling. In this framework, a GIS-derived centerline polyline is transformed into a continuous, station-based parametric alignment that can be deterministically regenerated and mapped to standard railway representations. The reconstructed alignment comprises the horizontal layout, a longitudinal profile, and (where available) cant information. The proposed procedure analyses the centerline geometry and reconstructs a parametric sequence of standard alignment elements (tangents, circular arcs and transition curves). This is achieved through a process that proposes candidate alignment spans and refines them during parametric conversion to ensure consistent element sequencing. The resulting parametric layout supports deterministic regeneration of the alignment and direct mapping to IFC 4.3 alignment entities (figure 2).

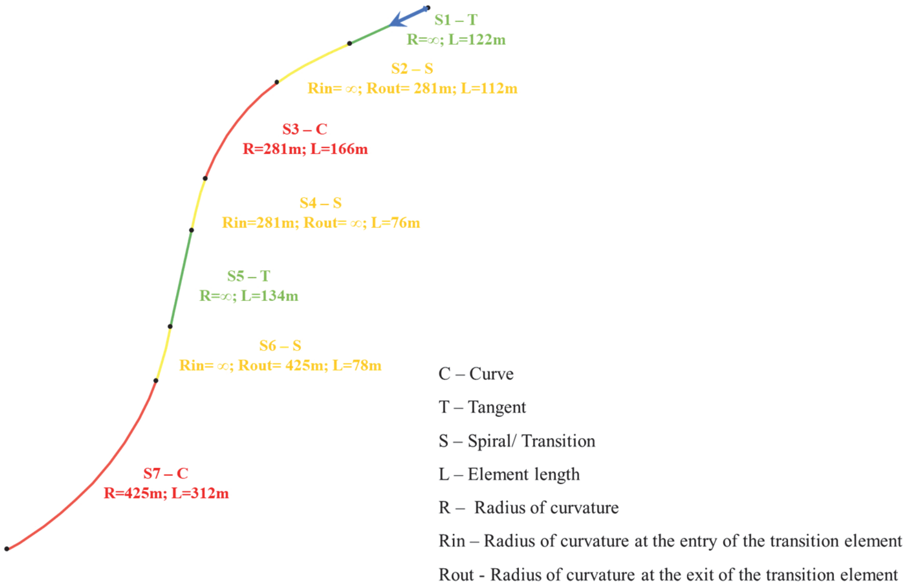

Figure 2 Example of alignment segmentation

2.3 IFC 4.3 automatic generation

The parametric alignment layout is automatically translated into an IFC 4.3 railway model through a dedicated generation module, implemented using IfcOpenShell [11]. This module instantiates the required IFC project context, spatial structure, and railway-specific entities to produce a standard-compliant and interoperable IFC 4.3 railway information model. The generation process starts by establishing the IFC project context, including unit settings and explicit georeferencing. In particular, the target coordinate reference system (CRS) is defined using IfcProjectedCRS, while the transformation between the model's local coordinates and the projected CRS is encoded through IfcMapConversion. A railway-oriented spatial structure is then created. At a higher level, the spatial hierarchy follows the standard IFC schema (e.g. IfcProject › IfcSite › IfcRailway/IfcRailwayPart).

The reconstructed alignment (section 2.2) is incorporated as an explicit IFC alignment definition (IfcAlignment). This alignment entity is decomposed into horizontal, vertical and cant components using the corresponding IFC alignment sub-entities, ensuring linear consistency with standard railway design representations. Beyond the alignment, the generation module produces a minimal yet robust IFC railway model that serves as an information model foundation. This process results in a 3D railway informative model derived directly from the alignment-centered IFC structure.

2.4 Geometric and semantic enrichment

The IFC 4.3 railway model is then enriched by adding standardized semantic information based on the bSDD and PDTs. GIS-derived enrichment layers can also be applied to include contextual corridor information. The enrichment is implemented as a dedicated post-processing step on the resulting IFC, ensuring that geometry reconstruction and information attribution remain reproducible and clearly separated.

Quantitative descriptors are extracted from the generated geometry and attached to the model as information. These quantities include segment lengths, start/end chainage extents and basic geometric parameters derived from the reconstructed alignment, which are stored as property sets. This approach allows that geometry-derived quantities are directly accessible within the IFC model without modifying the geometry. Semantic enrichment complements geometry by assigning meaning and standardized semantic identity to target elements. Two mechanisms are applied: firstly, bSDD provides stable concept identifiers that are encoded using IFC-native classification mechanisms, supported by metadata for traceability, and secondly, PDT-based enrichment assigns standardized properties and measurable quantities to assets, separating type-level information (e.g. standard dimensions or specifications) from instance-specific properties (e.g. chainage range) and quantities (e.g. lengths or volumes). The addition of GIS-derived enrichment introduces corridor-level information into the model, such as operational points and other referenced features along the alignment. Key railway information, including level crossings, hazard zones and other location-dependent features, is corridor-based and non-geometric. Surrounding features (e.g. bridges, viaducts, tunnels, or water-lines) are typically available in GIS datasets and can be spatially related to the railway model. This referenced information is critical for digital twin applications, which require a coherent representation of railway assets and their surrounding context.

3 Results and discussion

The proposed framework was applied to the Leixões railway line (Porto, Portugal), a conventional railway corridor of approximately 19 km (figure 3a). The case study reflects a typical existing-line context where geospatial datasets are the primary information source and where scalable digital modelling is required, particularly following the line's recent re-opening to passenger traffic.

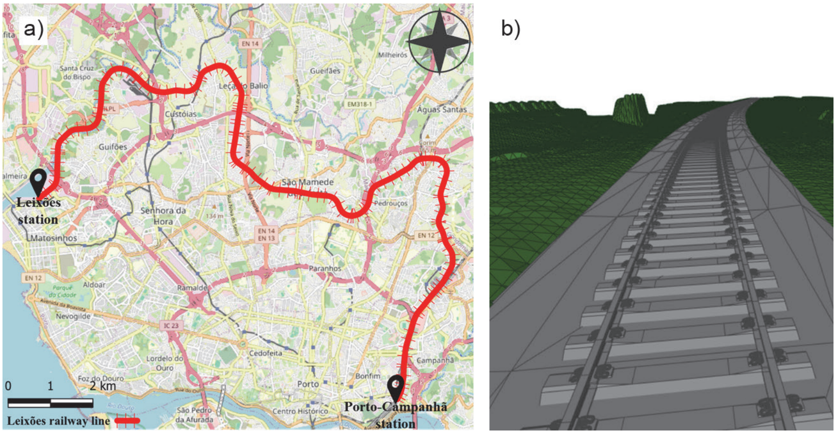

Figure 3 Leixões railway case study: a) geographical context and GIS-derived centerline used for alignment reconstruction; b) IFC 4.3 railway generation

The workflow generated a continuous, chainage-based parametric alignment compatible with IFC 4.3. The reconstructed alignment includes horizontal layout, a longitudinal profile supported by national terrain data, and cant information provided by the infrastructure manager. In addition, the workflow produced an IFC 4.3-compliant railway model in which the alignment is represented as an explicit IfcAlignment. This establishes a standard-compliant information model container in which the alignment serves as a linear reference for positioning and extending corridor-level information.

Following the generation of the baseline IFC model, the framework applies a post-processing enrichment stage to improve model usability of lifecycle-oriented workflows. The enrichment combines three information layers: i) chainage-based quantities, ii) standardized semantic classification and properties aligned with bSDD concepts and PDTs, and iii) GIS-derived corridor context. These additions transform the IFC into a baseline suitable for interoperability-focused workflows, enabling consistent asset inventory and corridor-level information management, and the structured integration of inspection and sensor observations along the corridor using chainage. By providing standardized properties and contextual information, the enriched model also supports lifecycle-oriented analyses, including maintenance planning and other corridor-scale decision-support uses within openBIM environments.

In general, the Leixões case study shows that geospatial data can be systematically transformed into a structured and semantically enriched IFC 4.3 railway model through and automated GIS-BIM workflow. The reconstructed parametric alignment demonstrates geometric consistency with the source GIS centerline, with mean deviations of approximately 1.0 m and maximum deviations on the order of 3 m, while the resulting model enables chainage-based positioning and structured representation of corridor elements (figure 3b). The alignment-centered representation enables the model to be incrementally extended with additional corridor elements within the same IFC structure. By bridging GIS-derived data and openBIM representations, the approach reduces reliance on manual BIM authoring and provides an open-standard foundation for long-term asset management and digital twin integration.

4 Conclusion

The automated generation of IFC 4.3 railway information models from geospatial data provides a practical response to the persistent gap between GIS and BIM workflows in existing railway corridors. The proposed framework establishes a standard-compliant workflow that transforms GIS-derived information into an interoperable IFC schema designed for reuse and incremental extension. The Leixões case study confirms that a reliable railway information baseline can be generated automatically from geospatial data, without manual BIM modelling, while remaining fully compatible with IFC 4.3. By combining parametric alignment reconstruction with semantic and GIS enrichment layers, grounded in bSDD concepts and PDT-aligned properties, the proposed approach enables consistent asset inventory, and lifecycle-oriented analyses in existing railway corridors. Crucially, the alignment-centered IFC provides a consistent reference backbone for linking external datasets (e.g. inspection records, sensor observations and sustainability indicators) to the correct locations along the alignment chainage.

Overall, the results indicate that automated generation of enrichment-ready IFC 4.3 models can significantly reduce reliance on manual BIM authoring, supporting scalable digitalization of existing railway assets. The contribution is not to replace detailed design BIM, but to provide a reproducible openBIM foundation that can be progressively extended with additional railway assets and data over time. Accordingly, the workflow offers a repeatable GIS-BIM pathway to produce open-standard railway models ready to feed asset management and digital twin applications. Future work will prioritize the inclusion of additional corridor assets (e.g. bridges, crossings and stations), broader validations under varying data quality and tighter integration with inspection/sensor data ingestion to support long-term asset monitoring and predictive maintenance.

Acknowledgement

This work was supported by the NRPCES project – “New Replacement Policy Considering Environmental Sustainability”, funded under the EUROGIA2030 Programme of the EUREKA Clusters Sustainability Call 2022 (SUS2022-039), with national co-financing from FEDER through the Norte 2030 program, under the scope of Portugal 2030 with national project number: 19217. The authors want to acknowledge UIDB/04708/2020 with DOI 10.54499/UIDB/04708/2020 and Programmatic Funding - UIDP/04708/2020 with DOI 10.54499/UIDP/04708/2020 of the CONSTRUCT - Instituto de I&D em Estruturas e Construções - funded by national funds through the FCT/MCTES (PIDDAC).

References

[1] Kaewunruen, S., Xu, N.: Digital Twin for Sustainability Evaluation of Railway Station Buildings, 4 (2018), DOI: 10.3389/fbuil.2018.00077

[2] Thompson, E.A., et al.: Revolutionizing railway systems: A systematic review of digital twin technologies, High-speed Railway, 3 (2025) 3, pp. 238-250, DOI: https://doi.org/10.1016/j.hspr.2025.05.005

[3] Kaewunruen, S., et al.: Digital twins for managing railway maintenance and resilience, 1 (2021) 91, DOI: 10.12688/openreseurope.13806.2

[4] Ciccone, A., et al.: Application of openBIM for the Management of Existing Railway Infrastructure: Case Study of the Cancello–Benevento Railway Line, Sustainability, 14 (2022), 2283, DOI: 10.3390/su14042283

[5] Pu, H., et al.: Recreating horizontal alignments of existing railways with a hybrid analytic and harmony search algorithm, Engineering Applications of Artificial Intelligence, 128 (2024), 107354, DOI: https://doi.org/10.1016/j.engappai.2023.107354

[6] Castro, A., et al.: A simple method for automatic recreation of railway horizontal alignments, Railway Engineering Science, 33 (2025) 1, pp. 62-78, DOI: 10.1007/s40534-024-00348-2

[7] ISO, ISO 16739-1:2024 - Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries, 2024.

[8] Garramone, M., Scaioni, M.: Ifcalignment for Raster-to-vector Gis Railway Centreline: A Case Study In The South of Italy, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4 (2022), pp. 39-45, DOI: 10.5194/isprs-archives-XLIII-B4-2022-39-2022

[9] Parracho, D.F.R., Martins, J.P.: idBIM4.0-National Library of BIM Objects: The Railway, DCE23-Doctoral Congress in Engineering, Porto, Portugal, June 2023.

[10] buildingSMART, I., buildingSMART Data Dictionary (bSDD), buildingSMART International, https://www.buildingsmart.org/users/services/buildingsmart-data-dictionary/

[11] Krijnen, T., Moult, D.: IfcOpenShell: The open source IFC toolkit and geometry engine, 2026.