Digital Twin for the Whole Life Cycle of Transportation Infrastructure: Digital Characterization Technology and Database Construction

Paiyong Zeng
PowerChina Huadong Engineering Corporation Limited, Hangzhou, China

Abstract:Aiming at the core pain points in the whole life cycle management of transportation infrastructure, such as data silos, decision-making lag, and insufficient collaboration between models and data, this paper proposes a four-dimensional digital characterization system based on "form-state-property-trend". A standardized characterization data list and rule system covering multiple facility types including roads, bridges and tunnels, and running through the whole stages of planning, design, construction and operation & maintenance is constructed. The multi-source heterogeneous data collaborative interconnection technology and distributed characterization database platform are innovatively developed, which break through the limitations of traditional data management modes and realize a new paradigm of digital management featuring "digital-model linkage, spatio-temporal integration and full-cycle traceability". By assigning weights and quantifying the completeness of characterization with an index system, core evaluation criteria are put forward, including the description integrity of not less than 95% and the business process coverage of not less than 90%, providing technical support for the high-quality construction of the digital twin base of transportation infrastructure. This technical system can effectively eliminate information barriers, improve the refinement level and decision-making efficiency of whole life cycle management, and provide an underlying technical paradigm for the digital transformation of smart city transportation.

Keywords: Transportation Infrastructure; Digital Twin; Digital Characterization; Four-Dimensional Framework; Multi-Source Data Fusion; Characterization Database

1. Introduction

1.1 Research Background

With the in-depth advancement of smart city construction and transportation digital transformation, the whole life cycle management of transportation infrastructure is confronted with multiple challenges such as complex business processes, heterogeneous data sources, and large spatio-temporal spans. In traditional management modes, problems including inconsistent model granularity across various stages, non-unified data coding, and difficult collaboration of multi-source information lead to a prominent data silo phenomenon. As a result, full life cycle data cannot be effectively circulated and traced, which restricts the realization of refined modeling and intelligent decision-making. As the core support for digital-physical integration and virtual-real interaction, digital twin technology provides a new path to address the above problems by constructing virtual mirrors of physical entities. However, current digital twin applications still lack a full-dimensional and dynamic digital characterization method for transportation infrastructure. Issues such as loose correlation between models and data, and absence of data governance standards lead to insufficient construction quality and application efficiency of the digital twin base. Therefore, researching systematic digital characterization technology and standardized database construction schemes has become the key to promoting the upgrading of full life cycle digital management of transportation infrastructure.

1.2 Research Status

Scholars at home and abroad have carried out relevant research on the digitalization of transportation infrastructure. In terms of digital characterization, existing studies mostly focus on geometric modeling or attribute description in a single stage, such as BIM-based model construction in the design phase and GIS-based spatial information characterization. However, they lack full-dimensional coverage of the static attributes-dynamic states-future trends throughout the whole life cycle. In terms of data management, existing databases mostly adopt a single storage architecture, which is difficult to meet the storage and interaction requirements of multi-source heterogeneous data (geometric data, time-series data, business data, etc.). Moreover, the low degree of data standardization results in difficulties in cross-stage and cross-system data collaboration. In terms of the evaluation system, no quantitative evaluation index for digital characterization completeness has been established, making it hard to guarantee the construction quality of the digital twin base. In summary, existing research lacks an integrated solution covering characterization framework-data standard-database architecture-evaluation system, which cannot meet the application requirements of digital twins for the whole life cycle of transportation infrastructure.

1.3 Research Content and Innovation Points

This paper conducts research on the full life cycle digital characterization and database construction of transportation infrastructure, with the main contents including: (1) constructing a four-dimensional digital characterization framework based on form-state-property-trend; (2) establishing a standardized characterization data list and rule system; (3) developing multi-source heterogeneous data collaborative interconnection technology; (4) designing and developing a distributed characterization database platform; (5) establishing a quantitative evaluation index system for digital characterization completeness.

The core innovation points are as follows:

The four-dimensional characterization framework of form-state-property-trend is proposed, which breaks through the limitations of traditional single-dimensional characterization. It realizes full-dimensional coverage from static geometric description to dynamic state perception, and from current situation assessment to future trend prediction, thus constructing a complete closed-loop digital expression system for the whole life cycle.

A multi-source heterogeneous data collaborative interconnection technology is innovated. Through key technologies such as a unified coding system, separate storage of models and data, and spatio-temporal data fusion, it solves the problem of standardized integration of different format models (RVT, IFC, DGN, etc.) and multi-type data (geometric data, time-series data, business data), realizing cross-stage and cross-system data interconnection and intercommunication.

A three-level distributed database architecture of application-service-storage is designed, which adapts to the storage characteristics of multi-source data. Technologies for graph-based storage of BIM models and spatialization of business data are innovated, realizing dynamic associated update of models and data, and improving the efficiency of data query and interaction.

A quantitative evaluation system for digital characterization is established, putting forward core indicators including description integrity $\geq$ 95% and business process coverage $\geq$ 90%. Through weight assignment and hierarchical evaluation methods, it provides an operable quantitative standard for the quality control of the digital twin base.

2. Digital Characterization System for Transportation Infrastructure

2.1 The "Form-State-Property-Trend" Four-Dimensional Characterization Framework

To achieve a comprehensive and dynamic digital depiction of transportation infrastructure, this paper proposes a four-dimensional digital characterization framework of "form-state-property-trend". The definition and connotation of each dimension are as follows:

Form: As the static spatial foundation of the digital model, it describes the physical morphology, spatial structure, and geometric characteristics of the facility, including dimensions, shape, spatial location, and topological relations. The modeling is implemented by combining the Boundary Representation (Brep) method with lightweight models, ensuring the accuracy and efficiency of geometric description.

Property: Reflecting the design intent and inherent capacity of the facility, it describes the intrinsic attributes, design parameters, and functional performance, covering core properties such as material characteristics, structural parameters, technical standards, and design indices. It provides basic data support for the performance analysis and safety assessment of the facility.

State: Embodiment of dynamic perception capability, it describes the operational status, health condition, and external environment of the facility at a specific time point, including real-time monitoring data such as traffic flow, pavement distresses, structural responses, and environmental parameters, realizing dynamic capture of the facility's operational state.

Trend: Supporting forward-looking decision-making, it describes the performance variation trend, degradation law, and future prediction of the facility over time, including Performance Condition Index (PCI), Bridge Condition Index (BCI), Tunnel Condition Index (TCI), degradation trend curves, remaining service life prediction, and risk assessment results, providing data support for maintenance decision-making and optimization.

This framework runs through the entire life cycle of infrastructure, achieving a complete digital expression from static to dynamic states and from current situation to future development, which lays a foundation for the formulation of subsequent data lists and rules.

2.2 Construction of Characterization Data List

Based on the four-dimensional characterization framework, this paper constructs a structured characterization data list covering multiple facility types and the entire life cycle stages. The list consists of 23 worksheets in total, which define key information of data items in detail, such as attribute name, data type, unit, and value range constraint. The core content is divided into four categories:

General Attributes: Defining full-life-cycle general attributes such as unique model unit identifier (unit_id), spatio-temporal datum, digital dimension pointer (form/state/property/trend), and metadata, ensuring the uniqueness and traceability of data.

Stage-Specific Attributes: Classified by facility type (road, bridge, tunnel, transportation hub) and life cycle stage (planning, design, construction, operation & maintenance), it lists the core data items of the "form-property-state" dimensions. Examples include basic project information and economic & technical indicators in the planning stage, structural parameters and technical standards in the design stage, schedule and quality document indexes in the construction stage, and real-time monitoring data in the operation & maintenance stage.

Performance Trend and Assessment: Defining the data of the "trend" dimension, including performance index calculation results, degradation trend model parameters, remaining service life prediction values, risk level assessment results, and maintenance decision recommendations, supporting forward-looking management.

Operation & Maintenance Process and Effect: Covering business data such as operation & maintenance work implementation records, process management data, and effect evaluation indexes, realizing digital traceability of the entire operation & maintenance process.

This data list provides a unified standard for data collection and standardization, serving as the foundation for eliminating data silos and achieving system interconnection.

2.3 Digital Characterization Rules

To ensure the standardized application of the data list and high-quality data governance, the following core characterization rules are formulated:

Classification of Characterization Objects: Clarifying the classification system of facility objects (roads, bridges, tunnels, auxiliary facilities) and business objects (planning, design, construction, operation & maintenance, etc.), realizing accurate identification and management of objects.

Data Modeling and Coding: Geometric data adopts a combined storage method of Brep, lightweight models, and graph-based storage. Attribute and business data are stored collaboratively in relational databases, time-series databases, and graph databases. A unified coding system is established, assigning a unique full-life-cycle identification code (unit_id) to each model unit, ensuring data uniqueness and interconnection.

Model-Data Mapping: Requiring one-to-one mapping between geometric model component identifiers and data codes. Business data is associated with model components through spatial information or codes, realizing "digital-model linkage" and ensuring accurate correspondence between models and data.

2.4 Data Weight Assignment and Indicator System

To quantitatively evaluate the completeness and effectiveness of digital characterization, this study establishes a supporting data weight assignment scheme and calculation method for key indicators.

Weight Assignment Scheme: According to the importance of data items for safety and decision-making, they are classified into four levels: core weight (4), important weight (3), basic weight (2), and auxiliary weight (1), which provides a basis for data quality assessment.

2.4.1 Calculation Framework for "Description Integrity"

(1) Definition of Core Concepts

Description integrity refers to the conformity degree between the data actually aggregated in the digital twin base and the applicable data items defined in the Whole Life Cycle Digital Characterization Data List.

Implications of the 95% Indicator: For a digital twin built for a specific facility (e.g., a certain bridge or a certain road section), if its calculated description integrity is ≥ 95%, the digital characterization of the facility is deemed highly complete at the data level, which can support intelligent analysis and decision-making.

(2) Preconditions for Calculation

Comprehensive Data List: A complete and standardized data list (i.e., the 23 aforementioned worksheets) has been established.

Clear Weight System: A weight reflecting the importance of each data item has been assigned.

Explicit Life Cycle Stage: The current life cycle stage of the evaluation object has been clearly defined.

Data Accessibility: The data in the digital twin base can be programmatically accessed and verified.

(3) Weight Assignment

The weight assignment is determined according to the Detailed Table of Weight Allocation.

(4) Calculation Steps and Formula

Calculation of Description Integrity for a Single Facility

Step 1: Define the Evaluation Context

Evaluation object: A specific model unit (e.g., unit_id = "Bridge_001")

Current life cycle stage: e.g., "Operation & Maintenance Stage"

Applicable data list: Determine the set of data worksheets to be evaluated according to the facility type and stage.

Example: For a bridge in the operation & maintenance stage, all "applicable" data items in the characterization data list need to be evaluated.

Step 2: Filter "Applicable" Data Items

Filter the data items that are meaningful and theoretically obtainable for the evaluation object from the list, and exclude the following:

Items inconsistent with the life cycle stage (e.g., inspection lot records of the construction stage are not applicable to bridges in the operation & maintenance stage)

Items inconsistent with the facility type (e.g., "cable force" of cable-stayed bridges is not applicable to girder bridges)

Items with technical or economic infeasibility (an exemption list needs to be defined in advance)

Step 3: Check the Completeness of Each Data Item One by One

For each applicable data item, check whether there are valid data records in the digital twin database and mark the status:

Figure 1.Valid Data Records in the Digital Twin Database

Data Status	Code	Judgment Criteria	Reduction Coefficient
Complete	C	Data exists, with correct format, reasonable value range and satisfied timeliness requirements	1.0
Partially Complete	P	Data exists but fails to meet some criteria (e.g., incomplete time coverage, insufficient precision)	0.5
Missing	M	Data does not exist or is completely invalid	0
Not Applicable	N/A	Judged as inapplicable according to the context	Excluded from the denominator

Step 4: Apply the Weighted Calculation Formula

$$\text{Description Integrity}=\frac{\sum_{i=1}^{n}(w_i \times s_i)}{\sum_{i=1}^{n} w_i}\times 100\%$$

Where:

n: Total number of applicable data items

wi : Weight of the i-th data item

si : Completeness coefficient of the i-th data item (C=1.0, P=0.5, M=0)

Step 5: Result Determination

≥95%: Description integrity meets the standard

80%–94%: Basically complete, requiring further improvement

<80%: Incomplete, requiring prioritized development

2.4.2 Calculation of Coverage for Over 90% of

Business Processes in Construction and Operation & Maintenance Stages

Core Calculation Principle:

Business Process Coverage=Total number of business processes÷Number of business processes fully or partially supported by the data list×100%

Judgment Standard: A business process is deemed covered if the core data items required for the process have corresponding entries in the data list.

Calculation Framework and Rules

(1) Calculation Basis

Business processes: 15 items in the construction stage + 15 items in the operation & maintenance stage = 30 items in total

Data list: 23 worksheets of characterization data

Core data requirements: 3–5 essential data categories for each business process

(2) Coverage Judgment Rules

Full Coverage: All core data requirements of a business process have corresponding entries in the list. Score: 1.0

Partial Coverage: More than 50% of the core data requirements of a business process have corresponding entries in the list. Score: 0.5

Non-coverage: Less than 50% of the core data requirements of a business process have corresponding entries in the list. Score: 0

(3) Calculation Steps

List 3 core data requirements for each business process based on their importance.

Check the correspondence of each data item against the 23 worksheets.

Count the coverage status (full/partial/non-coverage) of each business process.

Calculate the coverage rate for each stage and the overall coverage rate.

3. Technical Scheme of the Characterization Database

3.1 Overall Architecture Design

The characterization database adopts a three-tier "application-service-storage" architecture, which achieves efficient integration, storage and service of multi-source heterogeneous data. The architecture is illustrated in Figure 1.

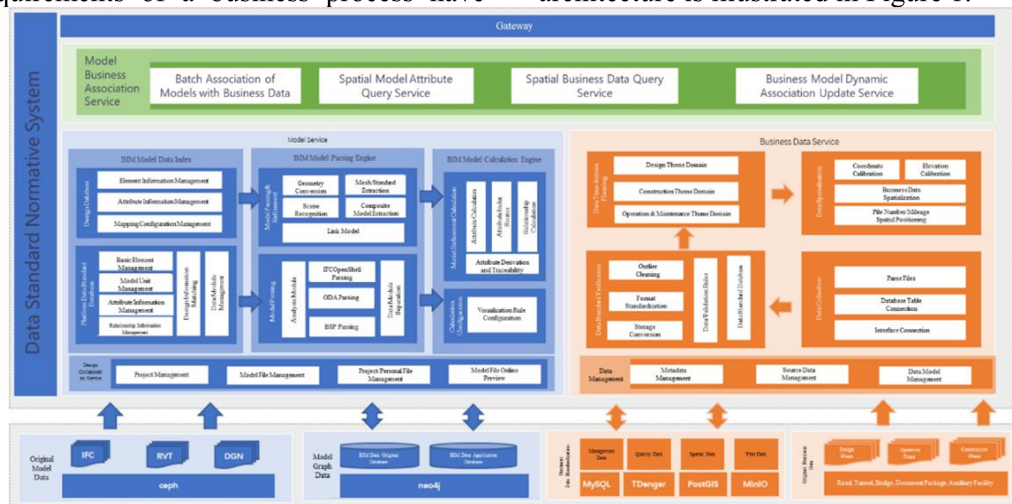

Figure 1. Overall Design of the Characterization Database

Application Layer: Provides four core services: batch association of models with business data, spatial model data query, spatial business data query, and dynamic association update of business models, realizing dynamic application and interaction of data.

Service Layer: Divided into model services and business data services:

Model services include functions such as design collaboration, BIM model parsing, and calculation engine.

Business data services cover modules like data collection, standardization verification, time-domain division, and data spatialization, realizing data cleaning and conversion.

Storage Layer: Adopts a collaborative storage scheme with multiple database types:

Raw model data is stored in Ceph object storage.

Graph-based model data is stored in the Neo4j graph database.

Standardized business data is stored in MySQL (relational data), TDengine (time-series data), and PostGIS (spatial data) according to its characteristics, realizing efficient data storage and access.

3.2 Core Technology Implementation

3.2.1 Multi-source Data Conversion and

Integration Technology

Cloud-native Parsing of BIM Models: Supports cloud-based parsing of models in multiple formats such as RVT, IFC, and DGN. Standardized extraction of model data is achieved through middleware including ODA and IFCOpenShell, thereby constructing a unified model knowledge graph.

Data Standardization Conversion: Based on characterization data standards, a dynamic middleware for data conversion and cleaning is developed to realize standardized processing of multi-source heterogeneous data. Consistency between model data and standards is ensured through conformance and compliance verification technologies.

Spatio-temporal Data Fusion: Fuses time-series data with BIM spatial data. With the BIM model as the base, data from different stages are associated by coordinates to achieve spatio-temporal collaboration of data. Through the business list linking function, automatic mapping between schedule, quality, cost lists and BIM model elements is realized.

3.2.2 BIM Model Graph-based Technology

Knowledge graph technology is innovatively adopted to store BIM model data, converting model geometric data, attribute data, and component relationships into a graph-based structure. Compared with traditional storage schemes, it has three major advantages: (1) Efficient complex relationship query capability, enabling rapid processing of dependency relationships between components; (2) Intuitive visual interaction, facilitating data exploration and analysis; (3) Supports flexible extraction of sub-models, enabling customized reorganization of model data according to business scenario requirements, and improving application efficiency and security.

3.2.3 Business Data Spatialization Technology

To address the lack of spatial information in business data, spatial tagging is adopted to endow characterization data with spatial attributes. Through coordinate mapping and ID mapping, the association between business data and BIM model spatial coordinates or spatial IDs is realized, enabling non-spatial data to have spatial query and analysis capabilities, and supporting spatial location-based retrieval and display of business data.

3.3 Database Application Scenarios

The characterization database supports four core application scenarios, realizing the full release of data value:

Batch Association of Models with Business Data: Through multi-rule configuration mapping, batch association between BIM models in different formats and business data is realized, such as linking of construction schedule processes and quality inspection lots, supporting reverse push and update of association relationships.

Spatial Model Attribute Query: Provides component-level spatial data query services, supporting real-time retrieval of information such as station numbers, mileage, GIS locations, and spatial relationships, and providing basic spatial data support for business systems.

Spatial Business Data Query: Based on the association relationship between models and business data, rapid query of business data such as construction progress, acceptance information, and equipment monitoring data is realized, supporting associated display in the BIM model.

Dynamic Update of Business Models: Supports dynamic association update between business data and BIM models. When design changes or business data modifications occur, the association relationships are automatically synchronized to ensure data real-time performance and consistency.

4. Conclusions and Outlook

4.1 Research Conclusions

Aiming at the core pain points in the whole life cycle digital management of transportation infrastructure, this paper proposes an integrated solution covering "characterization framework-data standard-database architecture-evaluation system". The main conclusions are as follows:

The constructed four-dimensional "form-state-property-trend" characterization framework realizes full-dimensional digital depiction of the static geometry, inherent properties, dynamic states, and future trends of transportation infrastructure. It breaks through the limitations of traditional single-dimensional characterization and provides a theoretical basis for full life cycle digital management.

The established standardized characterization data list and rule system cover multiple facility types and the entire life cycle stages, defining unified specifications for data collection and governance, and effectively solving the problems of data silos and inconsistent standards.

The developed distributed characterization database platform adopts multi-database collaborative storage and core technology innovation, realizing efficient integration, dynamic association, and rapid interaction of multi-source heterogeneous data, which supports the application requirements of "digital-model linkage and spatio-temporal integration".

The established quantitative evaluation system puts forward core indicators of description integrity ≥ 95% and business process coverage ≥ 90%, providing an operable quantitative standard for the quality control of the digital twin base.

4.2 Future Outlook

Future research can be further deepened in the following aspects: (1) Expand the adaptability of the characterization framework, study the differentiated characterization requirements of multiple transportation modes (highway, railway, rail transit), and improve cross-mode data standards; (2) Integrate artificial intelligence and machine learning technologies to enhance the intelligence level of performance trend prediction and maintenance decision-making; (3) Explore the in-depth integration of the digital twin base with other smart transportation systems (such as traffic signal control, emergency rescue) to expand application scenarios; (4) Research localized adaptation technologies to improve the independent controllability of the database platform, providing more solid technical support for the national transportation digitalization strategy.

Acknowledgments

This paper is supported by the National Key R&D Program of China (Grant No. 2022YFB2602101).

References

[1] HUANG J X. Research on Road Traffic Digital Twin System Based on Multivariate Composite Mapping[J]. Journal of Information Technology in Civil Engineering and Architecture, 2024, 16(5): 76-84. DOI: 10.16670/j.cnki.cn11-5823/tu.2024.05.08.

[2] WU C H, XU J D, FU Z Q, et al. Architecture, Key Technologies and Practical Cases of Highway Traffic Digital Twin System from the Perspective of Building a Powerful Transportation Country[J]. Journal of Transportation Research, 2023, 9(4): 1-16. DOI: 10.16503/j.cnki.2095-9931.2023.04.001.

[3] YANG G L, JIANG S M. Research on Multi-source Heterogeneous Data Fusion Method Based on ETL Technology[J]. Journal of Qilu University of Technology, 2024, 38(4): 64-70, 78. DOI: 10.16442/j.cnki.qlgydxxb.2024.04.003.

[4] LIANG H, FU D. Knowledge Graph Visual Fusion Method for Multi-source Heterogeneous Data[J]. Application of Electronic Technology, 2025, 51(1): 95-99. DOI: 10.16157/j.issn.0258-7998.233125.

[5] HE G T. Multi-source Electrical and Mechanical State Management of Expressways from the Perspective of Data Association[J]. China Transportation Informatization, 2024(11): 142-145. DOI: 10.13612/j.cnki.cntp.2024.11.041.

[6] LIAO Y T, TANG H, ZHU H. Design and Application of Integrated Intelligent Highway Management and Control Platform Based on Digital Twin[C]// Proceedings of the 17th China Intelligent Transportation Annual Conference. Beijing: Publishing House of Electronics Industry, 2022: 464-473.