ENABLING COMPREHENSIVE QUERYING OF ENVIRONMENTAL CONTEXT AND ROAD INFRASTRUCTURE DATA USING A GRAPH-BASED APPROACH

Ina Heise, André Borrmann
Chair of Computing in Civil and Building Engineering, TUM Georg Nemetschek Institute,
Technical University Munich, Germany

Abstract

Enabling predictive maintenance (PM) of road infrastructure by understanding condition developments requires accounting for internal and external influences. However, the corresponding data are highly heterogeneous and distributed, limiting comprehensive large-scale analysis. To bridge this gap, an existing Digital Twin (DT) approach is enhanced with environmental context data. Knowledge graphs enable scalable querying across heterogeneous subsystems, while a federated database addresses data distribution. The framework is validated using real-world data from Germany. Linking historical climate records with bridge condition development demonstrates comprehensive querying. By enabling scalable cross-system data association, the proposed DT framework enhances the practical applicability of existing PM methods.

Introduction

PM is crucial for ensuring the resilience and sustainability of road infrastructure. Its practical implementation is facilitated by informed decision-making, which requires not only a comprehensive understanding of the current state of the system but also an understanding of trends and tendencies in future conditions.

The challenge of comprehensive state representation arises from the system's complexity, driven by its size, heterogeneity, and the various relationships that span multiple scales. Furthermore, when analyzing trends and predicting the development of conditions, considering the system's interactions with its environmental context becomes crucial, as it allows integration of the climate conditions to which the system was exposed. When dealing with complex real-world systems, the DT concept depicts a recognized method. Applying DTs involves using a digital representation of the physical system, along with its various states, to apply computer-based methods. However, the complexity of road infrastructure and its environmental context is also apparent in the existing data, which is distributed across distinct systems managed and updated independently by different institutions and reflects heterogeneity in its underlying data structures. Thus, creating a database suitable for investigating the development of road infrastructure conditions in conjunction with external influences, such as climate, requires associating data with considerable manual effort, which strongly conflicts with the size of the system under consideration. A comprehensive DT that, in a scalable manner, captures road infrastructure alongside its climate exposures and relationships and facilitates automated data association, overcomes this limitation.

Contributions

Studies examining environmental impacts on road infrastructure deterioration primarily rely on a research database that already provides condition and climate data in an integrated manner. While these studies demonstrate the importance of considering environmental context, they neglect the data heterogeneity and distribution prevalent in road infrastructure management practice. Integration of heterogeneous maintenance data, particularly sensor data, is a key topic in research on road infrastructure DTs. Previous work presents frameworks that address the data-integration challenge using graph-based approaches. However, existing approaches focus on individual assets and neglect scalability to entire road infrastructure systems, which is essential for investigating correlations between environmental data and condition development. Other studies on graph-based DT frameworks focus on scalable capture of road infrastructure and its internal relationships, but neglect the system's interactions with the environmental context. Accordingly, a gap remains between the need for large-scale evaluation of road infrastructure maintenance data alongside its environmental contexts and the existing DT approaches that focus either on time-series data integration or scalability. Accordingly, we build on existing approaches for the scalable representation of road infrastructure with its interactions and enhance them by integrating interface relationships with environmental context data, particularly georeferenced time-series climate data. Thereby, not only is the integration of highly heterogeneous data structures at scale accounted for, but also the effects arising from differences in update cycles, spatial structuring approaches, and data distribution across various systems. Furthermore, scalability factoring requires an efficient approach to automatically establish links between time-series data resources and road infrastructure elements during the DT setup. By leveraging characteristics of road infrastructure, we facilitate inference of these relations, thereby significantly reducing the number of required geospatial calculations. The proposed concepts are validated on complete real-world datasets from German road infrastructure management and climate recordings. Thus, the ability to handle real-system sizes while acknowledging the data-inherent heterogeneity and distribution is proven.

Related Works

Environmental Context in Infrastructure Deterioration

Considering climatic influences is crucial when investigating the development of road infrastructure conditions. Providing a broad perspective, Sias et al. (2025) conduct a literature review of the effects of climate change on road infrastructure.

Zhang et al. (2025) examine the effects of rising temperatures on asphalt pavement aging. Focusing on more specific mechanisms, Zhou et al. (2020) analyze the effects of freeze-thaw cycles on concrete pavements, while Chen et al. (2023) examines the pavement roughness' dependency on the influences of temperature, moisture, rainfall, and freeze-thaw cycles. Using a data-driven approach, Piryonesi and El-Diraby (2021) integrate temperature ranges, perspiration, and freeze-thaw cycles in their pavement condition prediction. Lamei et al. (2025) investigate climate predictors for bridge component deterioration. It becomes evident that considering the climatic conditions to which a road infrastructure system is, was, and will be exposed is essential in understanding deterioration. This aligns with the Long-Term Pavement Performance (LTPP) program's (Federal Highway Administration, US Department of Transportation (2017), Schwartz et al. (2015)) goal of supporting research on the deterioration of road infrastructure by providing a comprehensive database. The LTPP database contains data on test road sections and bridges in North America and Canada, including condition records and accumulated climate data per test road section. It is apparent that virtually all studies investigating climate effects on deterioration on a larger scale (such as Zhang et al. (2025), Chen et al. (2023), Piryonesi and El-Diraby (2021), Lamei et al. (2025)) build on data from this LTPP database.

Although the LTPP database is a key resource for developing deterioration models, its static storage of aggregated climate data with fixed assignments to road sections introduces limitations. In particular, the lack of flexibility restricts the development of new deterioration models and the application of models that are not aligned with the accumulated metrics provided. In addition, data maintenance and integration of additional data sources are cumbersome, uncertainty quantification is rendered impossible, and adaptation to nonstationary climate conditions is not feasible. Moreover, this centralized data structure does not reflect real-world asset management practices, where road infrastructure and environmental data are typically maintained by separate institutions using distinct systems, each with heterogeneous underlying data structures.

Integrated Infrastructure DTs

Broo and Schooling (2023) highlights DT's capabilities as an effective tool for improving infrastructure systems maintenance through extensive data analysis, but also outlines existing challenges in integrating complex systems and data.

Knowledge graphs (KG) facilitate the integration of heterogeneous datasets by representing elements as nodes connected by edges that capture their relationships as stated in Hogan et al. (2022). Various graph models exist, including the Resource Description Framework (RDF) and Labeled Property Graph (LPG) model. While RDF benefits from unique identifiers and a clearly defined schema through ontologies, the LPG model offers greater flexibility and a more compact structure by allowing properties to be assigned to nodes and edges. While both approaches have their respective advantages and disadvantages, they are, in principle, suitable, as also reflected in existing works.

Hagedorn et al. (2023) presents a graph-based DT framework that integrates heterogeneous maintenance data for roads and bridges, building on a web-based implementation of the Information Container for Linked Document Delivery (ICDD). In Hagedorn et al. (2024), the focus shifts from maintenance to construction, but the ICDD approach is further refined by integrating dynamic sensor data. Although the ICDD approach enables linking RDF graphs with other time-series datasets, thereby facilitating data integration, its reliance on a file-based architecture requires a centralized data storage system. This centralized architecture contrasts with the distributed data management practices, especially when external data sources, such as weather data, are to be included. Furthermore, the largely static structure not only limits adaptability to changing requirements but also increases maintenance effort and reduces scalability in complex, evolving systems.

Li et al. (2024) examine another area of infrastructure systems with their graph-based DT approach for pipes and tunnels. By using LPGs for implementation, the approach benefits from the flexibility and a more compact graph structure. Li et al. (2024) approach involves considering environmental data in the form of measured ground settlements and, by capturing interactions, their effects on the considered structures. However, the approach explicitly stores interaction-describing relationships and operates on a single graph that represents the entire system. Thus, once again, current infrastructure management practices are neglected, just as the system's ability to adapt for integrating additional subsystems, interactions, or data resources.

Ramonell et al. (2023) integrates IoT data with infrastructure IFC models by proposing an LPG-based microservice architecture for asset maintenance. This approach reveals a limitation that also applies to the approaches proposed by Li et al. (2024), Hagedorn et al. (2023), and Hagedorn et al. (2024): In all cases, the integration of sensor data is examined only for individual assets. The scalability of the proposed DT approaches to an entire infrastructure system, including the associated number of assets and the corresponding number of sensors, remains unexamined. Additionally, when incorporating environmental data, such as climate data, this information often spans a broad area, making it relevant to multiple assets. As a result, adopting existing methods that rely on data integration via explicit component-level links is highly inefficient. In addition, existing approaches also emphasize streaming sensor data. Although this provides a comprehensive view of a component's actual condition, it neglects large-scale evaluation of historical time series, as compared to, e.g., documented condition changes in the context of correlation analysis. Thus, characteristics of road infrastructure systems, such as size, heterogeneity, and the time spans of interest, are not addressed.

Scalable Road Infrastructure DTs

When it comes to the scalable use of knowledge graphs, Purohit et al. (2020) reports improved scalability for LPGs and suggests mapping RDF graphs onto LPGs to combine the efficiency of LPGs with the knowledge provided by the underlying ontologies of RDF graphs.

Thus, existing DT approaches that address the scalable integration of heterogeneous data build upon LPGs. Heise et al. (2025a) introduce a DT-framework for the scalable evaluation of spatial relationships between heterogeneous road infrastructure elements. Heise et al. (2024) extend this approach to capture spatial relationships at multiple scales, and in Heise et al. (2025b), the framework is further enhanced to capture temporal relationships, enabling the investigation of dependencies between condition developments as time-dependent variables. However, while these works enable data analysis that accounts for internal interactions within road infrastructure, external influences, such as climate, remain unaddressed, although existing research underscores their significant impact on deterioration processes.

A gap remains in DT approaches that can also efficiently and scalably integrate long-term observation data at the corresponding frequency, especially since this data naturally lacks infrastructure-specific characteristics and is most efficiently stored in relational (time-series) databases and maintained independently. Thus, we adopt the DT framework from Heise et al. (2025a) for internal relationships in road infrastructure and enhance it by incorporating interface relationships with environmental context data.

Method

The core of capturing the complex relationships and interactions in road infrastructure lies in applying the System-of-Systems (SoS) concept initially introduced by Ackoff (1971). Relationships within road infrastructure and between road infrastructure elements and the environmental context are treated as distinct subsystems, as illustrated in Figure 1 on the left. Elements represented in multiple subsystems serve as connection points that link the subsystems to an overall SoS.

Reflecting this subdivision of the overall system in the architecture of the digital representation not only accounts for the distributed management and storage of data resulting from the actual distribution of responsibilities but also enables the selection of the most suitable database management system (DBMS) type. The right side of Figure 1 depicts the implementation of the SoS concept in a DT framework, incorporating digital representations of road infrastructure and its relationships as *InfraTwin*, and climate data as a distinct *Climate Twin* in *Layer I*. The cross-system evaluation of the subsystems builds on identifying representations of the same elements across subsystems, and implementation is achieved in *Layer II* via a federated database.

Layer I: Subsystems

The *Climate Twin* is assumed to be an existing external system that provides time-series climate data associated with a georeferenced element (typically the weather station where the data were recorded or the area to which the recorded values apply). It is common practice to use relational DBMSs for time series data. As an example for such a system, we consider the *Climate Data Center (CDC)*[1] maintained by the German Meteorological Service (Deutscher Wetterdienst).

The *InfraTwin* comprises further subsystems and builds on the concept Heise et al. (2025a). These subsystems capture both elements of road infrastructure and their conditions, as well as their internal relationships in the *Internal Relationship Graph*. The framework of Heise et al. (2025a) is enhanced by an additional subsystem comprising interface relationships to external elements in the *Interface Relation Graph*. Each *InfraTwin* subsystem is modelled as a knowledge graph and stored in a distinct graph database instance. The modeling of internal relationships within the *InfraTwin* follows the approach presented in Heise and Borrmann (2024) and exploits linear reference elements as a spatial structuring system. Assets and reference elements, stemming from the road network, are represented as nodes. Modeling their relationships implements the principle of absolute linear referencing. To capture the relationships between road infrastructure elements and the environmental context, this spatial structuring approach is leveraged, as it already accounts for road infrastructure-specific characteristics.

Rather than explicitly linking every asset to its environment, the geospatial context elements are mapped onto the linear reference elements as illustrated in Figure 2. Accordingly, it is explicitly modelled which parts of the road network belong to which environmental context. Environmental contexts are represented as areas on a linear reference element with specific start and end points, defined by distance values. Implementing this absolute linear referencing in the graph structure follows Heise and Borrmann (2024) in a simplified manner: environmental data and linear reference elements are represented as nodes, and the relationship between them is captured by connecting edges, as detailed in Figure 2 on the right.

[1] https://cdc.dwd.de/

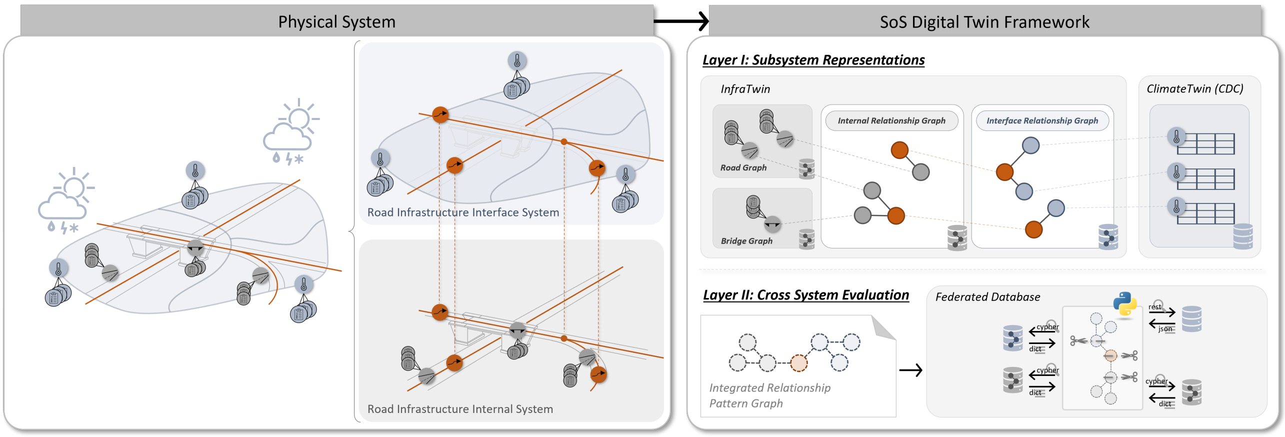

Figure 1: Representing Road Infrastructure with its Environmental Interactions as System-of-Systems Digital Twin Framework

Information on whether a road infrastructure element is located in a specific area can be derived by superimposing its relationships with the common reference element. Therefore, spatial relationships between environmental context and road infrastructure elements can be evaluated with the same precision as those between assets. This includes $n : n$ relationships between areas and assets, which may be important for structures spanning several kilometers, such as noise barrier walls, and their relationships to potentially multiple areas. In summary, the road network serves as the connection point between the two subsystems, which reduces not only the number of cross-system connections but also the number of elements to be represented in the *Interface Relationship Graph*, thereby efficiently reducing its size.

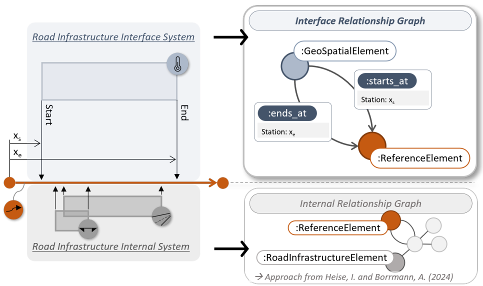

Figure 2: Graph Structures of the Relationship Graphs

Layer II: Comprehensive Evaluation

Applying the SoS principles and the corresponding subdivision into subsystems results in a distribution of information. Thus, elements with their corresponding data and various relationships are spread across several DBMSs. The cross-system relationship can be described as a pattern graph that spans the various subsystem graphs. Thus, identifying elements that exhibit a specific cross-system relationship comes down to a pattern-matching problem, searching for occurrences of this pattern graph in the overall system. The distributed storage of the subsystems turns the problem into a distributed pattern-matching problem, in which each pattern graph fragment corresponds to a single subsystem.

The implementation of distributed pattern matching forms *Layer II* in Figure 1. A federated database connects the distinct DBMSs of the different subsystems without holding any data itself. Instead, for each pattern graph fragment, corresponding queries are sent to the respective subsystem database endpoints. The order (as well as potential parallelization) of this query process is determined by the dependencies defined in the overall pattern graph. Additionally, aligning the query direction with the expected solution set sizes for each subsystem evaluation helps quickly exclude large portions of the system, thereby enhancing overall evaluation efficiency. Implementation of *Layer II* can be achieved through a script, e.g, using Python.

It is worth noting that the query language and the way data is accessed may vary across fragments, as the concept explicitly considers different DBMS types across subsystems. The depiction in *Layer II* of Figure 1 assumes that Neo4j is used as a graph DBMS, and accordingly, interactions involve sending Cypher queries, while the climate time series can be efficiently obtained via a REST API.

Creation of the Interface Relationship Graph

The *Interface Relationship Graph* comprises the relationships between the geospatial environmental context and the road network edges. Since these relationships are not yet present in the data, geospatial calculations are required. Assuming the road network's edge geometries are polylines, this calculation identifies which segments of those polylines fall within specific environmental areas. In principle, this analysis corresponds to geospatial querying already implemented in diverse interfaces and database engines, and it is also applied in existing approaches, such as those presented in the Related Works Section.

Yet, given the specific characteristics of road infrastructure systems, particularly the common sizes of real-world road networks, this poses a fundamental dilemma between geometric analysis accuracy and scalability. In addition, the number of polyline segments of a network edge is often unrelated to its spatial extent. On the contrary, network edges with a small spatial extent, often connecting other main axes, typically have a large number of polyline segments due to their high curvature. Therefore, considering inherent net characteristics when determining the spatial relationships between the road network and environmental areas, rather than computing them for every road network polyline segment, appears reasonable.

The proposed approach builds on the assumptions that (1) the geometry of the road network is significantly finer-grained than its environmental context areas. Thus, one network edge intersects at most one boundary between environment areas. And (2) only qualitative relationships between the road network and the environmental context areas are to be evaluated, and a tolerance of the geometric calculations of several meters is acceptable.

The approach consists of a two-step process: Firstly, the geospatial relationship between the net nodes and the areas is computed. Based on this information, the network edges themselves are considered second. During the analysis of the network edges, two cases are distinguished, aiming to use inference as far as possible rather than a geometric analysis. Figure 3 illustrates both cases.

In *Case I*, relationships between net edges and areas can be completely inferred by assuming that if two net nodes V_1 and V_2 are within an area A, this also applies to the network edge (and correspondingly the reference element) E connecting V_1 and V_2. Accordingly, the spatial relation required to create the *Interface Relation Graph* is determined using Equation 1. As the environmental area spans the entire reference element, the start station is 0, and the end station is the length of the reference element, respectively.

$$\frac{\mathrm{in}(V_1, A) \quad \mathrm{in}(V_2, A) \quad \mathrm{starts}(V_1, E) \quad \mathrm{ends}(V_2, E)}{\mathrm{in}(E, A)} \tag{1}$$

Case II covers V_1 and V_2 contained in different areas A_1 and A_2 and therefore involves the reference element intersecting a boundary. To localize the boundary P_{Bound} on the reference element, the individual points of the polyline $P_1, ..., P_n$ are checked for their containment in A_1. P_{Bound} corresponds to the first point identified outside A_1, and investigations stop if P_{Bound} is identified. The P_{Bound} station together with the length of the reference element provides the information to model the relationship between the two environmental areas and the reference element, as illustrated in Figure 3.

The result is the *Interface Relation Graph*, which captures the spatial relationships between environmental context areas and linear reference elements, thereby facilitating the comprehensive evaluation of climate and asset data.

Prototypical Real-World Implementation

To validate the practical applicability of the proposed framework, it is applied to a real road infrastructure system. Real-world data from German infrastructure management and publicly available climate data, both structured differently and originating from distinct sources, are used.

Setting Up the DT-Framework

The *InfraTwin* integrates data on bridges and the German road network, while the *ClimateTwin* is represented by the CDC. Bridge data in Germany is maintained in SIB-Bauwerke[2], a proprietary asset management system. SIB-Bauwerke builds on a relational database and supports only file-based data exchange. For the case study, we used a copy of the database for all Bavarian (a state of Germany) assets provided to us as part of a research project. The German road network is centrally managed and publicly provided by the Federal Highway Administration (Bundesanstalt für Straßenwesen (BASt)). For the case study, the data were obtained through the WFS interface[3]. The CDC provides several interfaces, including a REST API and an open file server. Since the REST API interface is still under construction and therefore does not yet provide access to the full data set, the HTTP-based open file server[4] was used, despite the resulting limitations in filtering options. The prototypical implementation of the concept employs Labeled Property Graphs (LPG) and Neo4j, a widely used graph DBMS, for their storage. The federated database is implemented in a Python script.

Internal Relationship Graphs Creation

To facilitate integration of the different legacy data resources, graph representations of the bridge data and the road network were created and stored in distinct graph database instances. Capturing the internal relationships follows the approach presented in Heise et al. (2025a). This involves creating the *Internal Relationship Graph*, which captures the spatial relationships between bridges and reference elements (as referred to in Heise et al. (2025a) as the *InfraNetRelation Graph*). Furthermore, graphs representing the net (*Net Graph*) and the bridge condition recordings (*Bridge Graph*), each stored in a distinct graph database, were generated.

Interface Relationship Graph Creation

In this case study, the *Net Graph* serves as a basis for creating the *Interface Relationship Graph*. According to the concept presented earlier, this involves assigning net nodes to their environmental contexts first. Since the climate data in the CDC is stored associated with weather stations possessing point coordinates, calculating the required information yielded a Voronoi-like decomposition via Euclidean nearest-neighbor assignment. To ensure computational efficiency, the distance calculation was performed using a vectorized approach. For each net node, distances to all weather stations were computed simultaneously, and the minimum value was selected, which avoided the construction of a full distance matrix. Implementation was achieved through a Python script using the NumPy library. The computed assignments of net nodes and weather sta-

[2] https://sib-bauwerke.de/

[3] https://inspire.bast.de/geoserver/bisstra.strasse/wfs?service=WFS

[4] https://opendata.dwd.de/

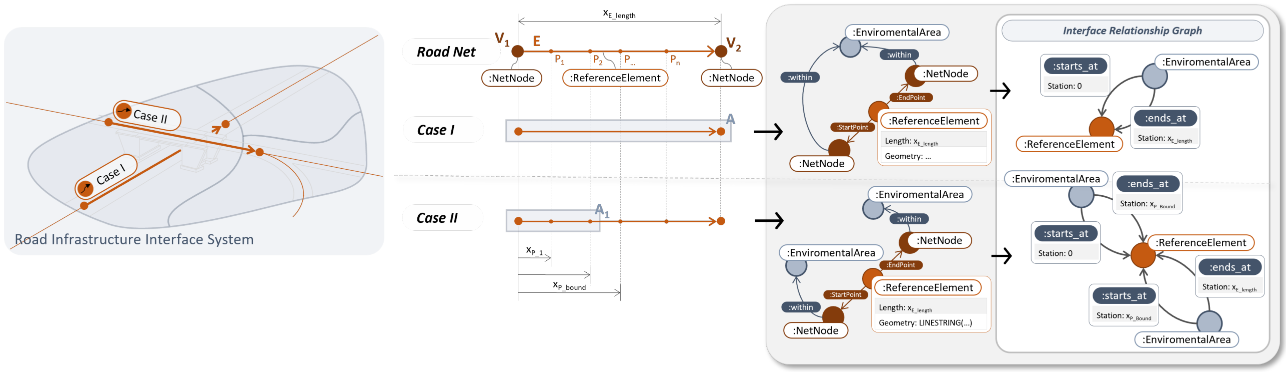

Figure 3: Interface Relationship Graph Creation

tions are fed into the *Net Graph*.
Based on this, the interface relationships between reference elements and weather stations satisfying the conditions of *Case I* could be inferred. The inference is implemented using a CYPHER query, which first extracts the corresponding reference element-weather station pairs (along with the length of the reference element and the respective IDs). Based on the created dictionary, the corresponding nodes and relations in the *Interface Relationship Graph* are generated.
The remaining reference elements fall under *Case II*. To determine x_{P_bound}, a CYPHER query directed to the *Net Graph* identifies the corresponding reference elements with associated weather stations linked to its start and end point. The IDs, along with their geometry (stored as WKT), are given back. The calculation of x_{P_bound} is then implemented in Python using the NumPy and Shapely libraries. The calculation results facilitate the enhancement of the *Interface Relationship Graph* with the remaining relationships between weather stations and reference elements.

The Example Use Case

Associating bridge conditions with frost-thaw cycles derived from daily temperature recordings serves as an example use case. The bridge conditions are contained in the *Bridge Graph* linked to the respective asset. Temperature recordings are provided in different resolutions in the CDC. Automatic association is facilitated by comprehensively evaluating the *Interface Relationships Graph* and *Internal Relationship Graph*, as illustrated in Figure 4. Firstly, the *Internal Relationship Graph* provides all Bridges in the dataset, along with their localization on a reference element. This information is used to subsequently query the *Interface Relationship Graph*. Using the reference elements' IDs, the weather station corresponding to each bridge is determined. The resulting dictionary contains bridge IDs assigned to weather station IDs.
For further querying, the capabilities of the provided interfaces were taken into account. If fetching temperature records for a specific time interval is possible, e.g., via a REST API, continuing the sequential process in which the condition ratings of the bridges were queried first, followed by the temperature data for the covered intervals, is beneficial. However, the CDC interface provides only a single text file containing all historical temperature data for each weather station, prohibiting the preselection of climate data based on bridge data availability. Accordingly, parallel data retrieval was implemented. The first subprocess extends the dictionary of weather station IDs and bridge IDs with a list of available condition grade recordings per bridge by interacting with the *Bridge Graph*. For each weather station ID, historical soil temperature records are retrieved from the CDC. Within the CDC, the ZIP archives corresponding to a specific weather station could be automatically identified and downloaded (via an HTTP request) using the employed naming convention. After downloading the ZIP archives, they are opened in memory, and the text file containing the actual temperature data is extracted and stored, using the corresponding weather station's ID as the file name.
Finally, investigating temporal relationships by implementing the approach presented in Heise et al. (2025b) enables associating frost-thaw cycles with bridge condition recordings, thereby facilitating further data analysis.

Results

The prototypical implementation demonstrated the applicability of the proposed framework to real-world data sets from Germany. The CDC dataset comprised data from 5.311 weather stations. The road network contains 36.565 network edges (and respectively reference elements). The relationship of 87% (31.916) of these reference elements to weather stations could be completely inferred as they fell under *Case I* (Figure 3). Only the remaining 4.649 reference elements required detailed analysis.
The *Internal Relationship Graph* contains 31.309 asset representations. Since we had access to asset data only from a part of Germany (Bavaria), but could use the road network for the entire country, this resulted in fewer assets as reference elements. Nevertheless, the reduction in explicitly modeled relationships by using the net as a connecting element is evident in the average number of assets per reference element, which is 3.3 and is expected to be even higher if road sections are to be integrated.
Implementing automated association of bridge condition

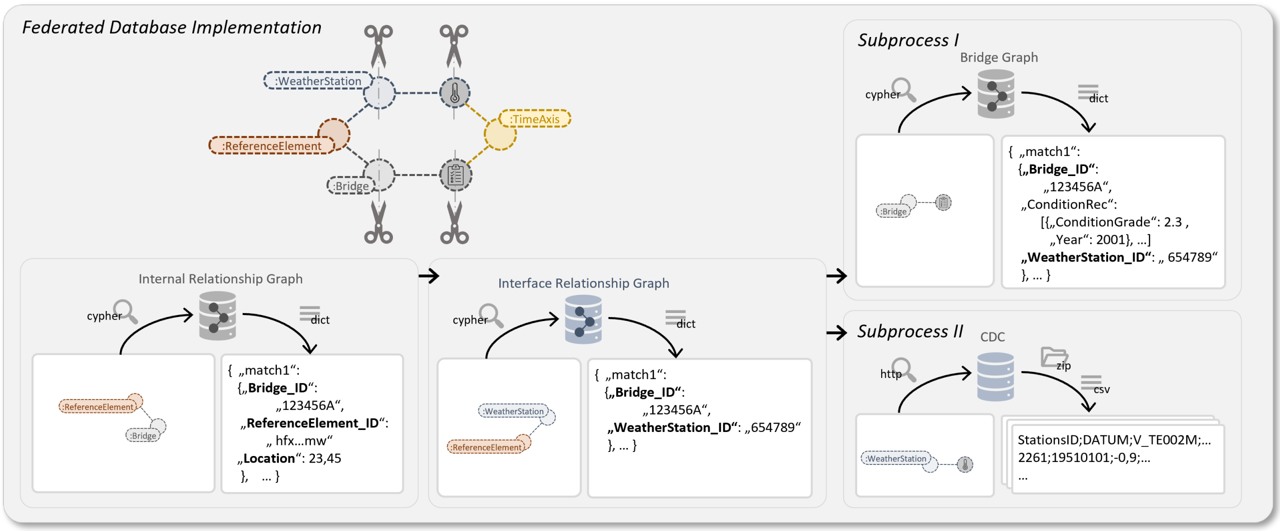

Figure 4: Prototypical Implementation of Layer (II): Federated Database for Comprehensive Evaluation

data with temperature recordings was successful. However, evaluating the temporal relationships proved a bottleneck due to the large number of temperature recordings.

Discussion

The prototypical implementation demonstrates the proposed concept's ability to comprehensively evaluate environmental context and road infrastructure data at a large scale, using temperature and bridge condition records. By considering complete real-world datasets, not only are realistic system sizes accounted for, but also the inherent distributed storage and data maintenance, as well as the significant heterogeneity of data structures involved. This paves the way for applying existing works on multifactorial deterioration models, which currently build on the central, homogenised LTPP data set, to a broader context, including infrastructure management systems currently in use.
By using the road network as a connection point between the two subsystems, a spatial structuring approach accounting for infrastructure-specific characteristics is employed. The approach's effectiveness results from reducing explicitly modeled relationships between road infrastructure and external elements. This minimizes not only the size of the *Interface Relationship Graph*, thereby improving performance during querying and creation, but also the number of cross-system connections that must be considered and maintained.
In addition, mapping environmental context areas to the reference elements enhances the accuracy of spatial relationship evaluation by accounting for the large longitudinal extent of the infrastructure assets. However, due to the very large difference in granularity when mapping climate data to the German road network, only 13% of the reference elements even cross a boundary between two environmental context areas. The effect can therefore be better exploited by examining relationships with environmental data of finer resolution, such as soil data.
The approach faces limitations in ensuring consistency across several subsystems. In particular, this affects changes to the road network or the available weather stations, which require updating the *Interface Relationship Graph* to ensure complete integration of the available data sources. Provided changes are known, the elements' IDs enable efficient updates. However, as the data sources involved are maintained completely independently, this knowledge most likely does not exist. In this case, there is a lack of efficient update strategies that prevent the entire subsystem graph from having to be checked for updates or completely regenerated. Although complete regeneration of the *Interface Relationship Graph* is not the optimal updating strategy, it still benefits from minimizing connection points between subsystems by using the road network instead of individual assets, and from deploying inference to avoid unnecessary geospatial calculations.
Furthermore, the case study revealed the limited effectiveness of the approach of Heise et al. (2025b) for analyzing temporal relationships when the data exhibit a very high degree of heterogeneity in update cycles. Integrating the concept of reference intervals proposed by Allen (1983) appears worth investigating.

Conclusion

The paper presents a comprehensive DT framework that captures road infrastructure and its environmental context. The integration is facilitated by applying SoS principles in the digital representation. A knowledge-graph-based approach addresses scalable semantic integration by capturing contextual relationships and enabling their flexible evaluation despite their inherent heterogeneity. The comprehensive evaluation, accounting for data distribution across heterogeneous DBMS (types), is implemented as a federated database. The applicability of the concept is proven by a prototypical implementation using complete real-world datasets from German infrastructure management and the German meteorological service (Deutscher Wetterdienst, DWD). The achieved data association enables investigating dependencies between condition development and temperature fluctuations. Thus, the presented approach paves the way for PM in infrastructure by leveraging existing data to better understand deterioration.

Acknowledgements

The presented research has been funded by the TUM Georg Nemetschek Institute - AI for the Built World.

References

Ackoff, R. L. (1971). Towards a System of Systems Concepts. Management Science, 17:661–671.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. Communications of the ACM, 26:832–843.

Broo, D. G. and Schooling, J. (2023). Digital twins in infrastructure: definitions, current practices, challenges and strategies. International Journal of Construction Management, 23:1254–1263.

Chen, C., Deng, Y., Li, M., and Shi, X. (2023). Investigation of key climatic factors affecting asphalt pavement roughness in different climate regions. Transportation Research Part D: Transport and Environment, 122:103877.

Federal Highway Administration, US Department of Transportation (2017). The Long-Term Pavement Performance Program. Technical report.

Hagedorn, P., Liu, L., König, M., Hajdin, R., Blumenfeld, T., Stöckner, M., Billmaier, M., Grossauer, K., and Gavin, K. (2023). Bim-enabled infrastructure asset management using information containers and semantic web. Journal of Computing in Civil Engineering, 37.

Hagedorn, P., Petrova, E., and König, M. (2024). A system-of-systems approach for deploying containerized construction digital twins using linked data. Proceedings of the 20th conference of the International Society for Computing in Civil and Building Engineering (ICCCBE2024).

Heise, I. and Borrmann, A. (2024). infraspatialOT: An ontology for the representation of spatial relationships in road infrastructure. In LDAC 2024 - Linked Data in Architecture and Construction.

Heise, I., Esser, S., and Borrmann, A. (2025a). A graph-based systems-of-systems architecture enabling multi-scale Digital Twins for maintaining road infrastructure. Advanced Engineering Informatics, 68:103649.

Heise, I., Esser, S., and Borrmann, A. (2025b). Enabling Spatial and Temporal Correlation of Heterogeneous Infrastructure Data Using Knowledge Graphs. Proceedings of the 2025 European Conference of Computing in Construction (2025 EC³).

Heise, I., Göbels, A., Borrmann, A., and Beetz, J. (2024). Enabling Comprehensive Querying of Road and Civil Structure Data using Graph-based Methods. In Proceedings of the 41st International Conference of CIB W78.

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., and Zimmermann, A. (2022). Knowledge Graphs. ACM Computing Surveys, 54:1–37.

Lamei, T., Abdel-Rahim, A., and Ibrahim, A. (2025). Development of performance-based deterioration models for bridge components. Innovative Infrastructure Solutions, 10:230.

Li, T., Rui, Y., Zhu, H., Lu, L., and Li, X. (2024). Comprehensive digital twin for infrastructure: A novel ontology and graph-based modelling paradigm. Advanced Engineering Informatics, 62:102747.

Piryonesi, S. M. and El-Diraby, T. (2021). Climate change impact on infrastructure: A machine learning solution for predicting pavement condition index. Construction and Building Materials, 306:124905.

Purohit, S., Van, N., and Chin, G. (2020). Semantic Property Graph for Scalable Knowledge Graph Analytics. In 2020 IEEE International Conference on Big Data (Big Data), pages 2672–2677. IEEE.

Ramonell, C., Chacón, R., and Posada, H. (2023). Knowledge graph-based data integration system for digital twins of built assets. Automation in Construction, 156.

Schwartz, C., Elkins, G., Li, R., Visintine, B. A., Forman, B., Rada, G. R., and Groeger, J. L. (2015). Evaluation of LTPP Climatic Data for Use In Mechanistic-Empirical Pavement Design Guide(MEPDG) Calibration and other Pavement Analysis. Technical report.

Sias, J. E., Dave, E. V., Underwood, B. S., Bowers, B. F., Harvey, J. T., Henning, T. F. P., Tighe, S. L., Jacobs, J. M., Pregnolato, M., Qiao, Y., Mecray, E., Golalipour, A., Chamorro, A., and Hendrick, P. (2025). Climate change impacts on roadways. Nature Reviews Earth & Environment, 6:555–573.

Zhang, R., Sun, L., Qiao, Y., Sias, J. E., and Dave, E. V. (2025). Multidimensional comparative analysis of future climate change impacts on pavement infrastructure aging. Transportation Research Part D: Transport and Environment, 142:104702.

Zhou, C., Lan, G., Cao, P., Tang, C., Cao, Q., Xu, Y., and Feng, D. (2020). Impact of freeze-thaw environment on concrete materials in two-lift concrete pavement. Construction and Building Materials, 262:120070.