Enabling Comprehensive Querying of Road and Civil Structure Data using Graph-based Methods

Ina Heise, (ina.heise@tum.de)
Chair of Computational Modeling and Simulation, TU Munich, Germany

Anne Göbels, (goebels@dc.rwth-aachen.de)
Design Computation, RWTH Aachen University, Germany

André Borrmann, (andre.borrmann@tum.de)
Chair of Computational Modeling and Simulation, TU Munich, Germany

Jakob Beetz, (beetz@dc.rwth-aachen.de)
Design Computation, RWTH Aachen University, Germany

Abstract
Road infrastructure is a complex system of heterogeneous subsystems that influence each other. Due to their different characteristics, the individual subsystems *road* and *bridge* are often considered separately during their operation. Thus, effects resulting from the interaction between the subsystems can only be analyzed manually and with a high curatorial effort for specific assets. Therefore, the management of the individual subsystems can benefit from a comprehensive, overarching data management of all subsystems. This paper presents how to enable cross-domain analysis of roads and bridges using graph-based representation of legacy data. The approach permits the consideration of specific granularities and focuses primarily on mapping the different spatial reference systems, enabling the implementation of location-based queries across different subsystems. To validate our approach, we demonstrate the application based on real-world data from the German infrastructure administration.
Keywords: Roads, Bridges, Infrastructure, Graph-based methods, Cross-system analysis

1. Introduction
Road infrastructure plays an essential role in society, both socially and economically. Its maintenance requires extensive measures, the necessity and effectiveness of which depend on the condition of the individual subsystems in the road infrastructure system and how they interact. The quality of maintenance can benefit significantly from a reliable information base that supports the choice of appropriate measures and decisions. Providing such an information hub requires representing the road infrastructure as an overall system with various interacting subsystems, such as roads and bridges.
The complexity of this task lies not only in integrating the various subsystems but also in the different levels of detail occurring. However, integrating higher granularities would significantly expand the variety of effects to be analyzed. For instance, automatically recording gradients during road inspections makes it possible to identify low points

where water may collect. Bridge components – or areas of them – under such low points are particularly vulnerable to water damage. To comprehensively analyze these interrelationships on a large scale, it is essential to devise a semi-automatic superimposition of road inclination with component areas of bridges.

In practice, however, different road infrastructure elements are usually managed in separate, unconnected systems operated by different stakeholder organizations with varying information system infrastructures. A data exchange consists, if at all, of a (one-time) duplication of data from one system to the other, resulting in duplicate data management and potential contradictions (Weise et al., 2018). Due to the resulting data silos, comprehensive analyses are only possible on a small scale and with a lot of manual effort. As a result, the individual subsystems are currently mostly considered in an isolated fashion, and the interactions between the subsystems are not regarded further.

To open up these data silos, previous approaches have already been devised to establish graph-based representations of legacy bridge and road data. Since a proprietary program is used for managing the road data whose data structure is not publicly documented, a generic approach was chosen, resulting in schema-less representation as a Labeled Property Graph (LPG). The bridge data management system, on the other hand, is based on a national, public data model that – transformed into an ontology – enables the bridge data to be represented as a schema-based Resource Description Graph (RDF).

This paper presents a combined consideration of the two graph representations to enable comprehensive analyses of legacy data available for different parts of the infrastructure at varying levels of detail. The approach is based on mapping the different spatial reference systems, enabling the implementation of location-based queries to both graphs and providing the results (spatially) superimposed.

The concept is validated using pre-existing German infrastructure data. The data, containing information on over 43,000 structures and 40,000 km of roads, is (partly) transformed into graph representations to showcase two practical applications of a comprehensive analysis of road and bridge data. The first application enhances bridge management system data with information from road condition surveys, while the second identifies bridge areas prone to water damage using inclination details from road data.

2. Related Research

Various research approaches have already investigated cross-system analyses and data management methods for the infrastructure sector, which is a well-known problem in integrating Building Information Modeling (BIM) and Geographic Information System (GIS) data. The challenges lie in dealing with the different scalings, transforming the relative coordinate system of the BIM data and the absolute coordinate system of the GIS data, mapping different levels of detail, and transferring semantic data Zhu and Wu, 2022. Similar challenges arise when combining bridge data based on a relative spatial reference system with an absolute approach for referencing used in road data.

The CEDR Interlink project (Luiten et al., 2018) was a central project for transnational communication and data exchange in the road sector. It aimed to establish interoperability between the European National Road Authorities (NRAs) by offering a central Object Type Library (OTL) to which national data models should be mapped. The method is based on the Linked Data approach, which can connect heterogeneous data models.

In the ”okstraOWL” project presented in Beetz et al. (2018), a methodology to link the German object catalog for the road and traffic data (OKSTRA) to other (international) information models was developed. To this end, the OKSTRA data model was converted into an ontology and mapped to the Dutch OTL (RWS-OTL). The approach was tested on the instance level by conducting cross-model queries and was evaluated as a contribution

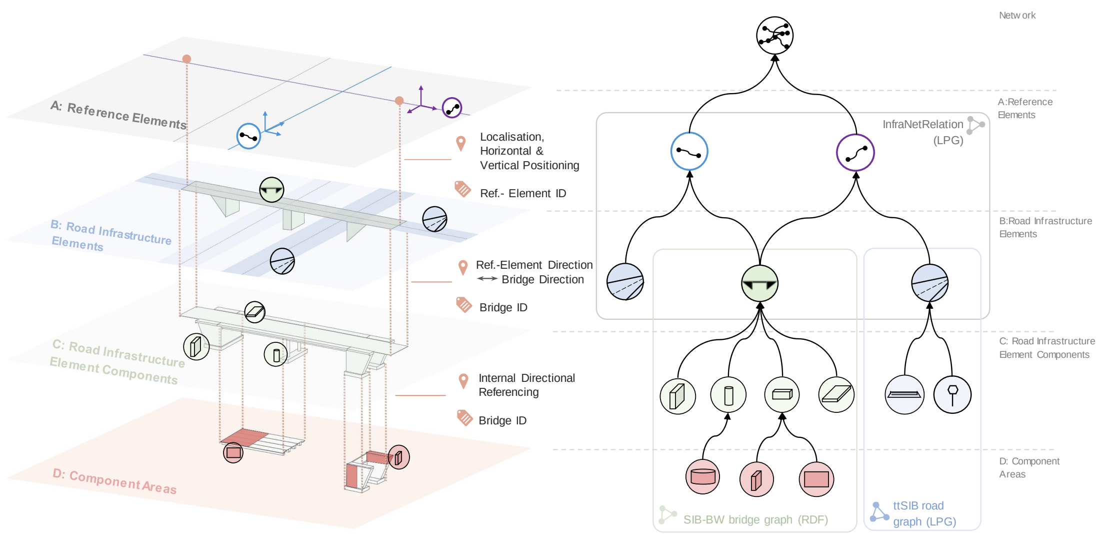

Figure 1: Assignment of road infrastructure elements to reference elements

to a pan-European data model stemming from INTERLINK.
Tulke et al. (2016) proposed an infrastructure information system (ISIS) that should optimize information allocation and is based on the existing German Bridge Management System (SIB-Bauwerke). They emphasize the importance of including road/route information in a bridge maintenance system to enable bridge condition analyses of entire route sections. In their approach, however, they only visualize the location of the bridges on a digital map based on their coordinates. However, the authors also state that a connection to road databases is recommended in the future.
All the work presented demonstrates the need for mapping between the various data (models) and scalings in the infrastructure sector and the facilitation of comprehensive analyses. Our work contributes to enabling this networking and is also rooted in graph-based methods, which have demonstrated their benefits in the Interlink and OkstraOWL projects.

3. Background to the subsystem approaches

The integration of subsystems and their relationships follows a hierarchical approach, with each subsystem being incorporated into the overall system by assigning them to the next higher super-system. This method results in a tree structure, as illustrated in figure 1, where the spatial relations between elements within a system can be determined by explicitly defining their relationships to the higher-level system. Every Subsystem is represented by a distinct graph as shown in figure 1. The approaches used within the different subsystem graphs are briefly explained below.

3.1. Spatial reference system of the road infrastructure data

Linear reference elements are used for spatial structuring. Accordingly, relationships between the respective infrastructure element and the reference element are modeled explicitly as localization in the longitudinal direction and positioning in the transverse direction, as illustrated in figure 2 and described in Heise and Borrmann, 2024.
When specifying the localization information, the relationship between the road infrastructure element and the reference element is modeled based on the principles of abso-

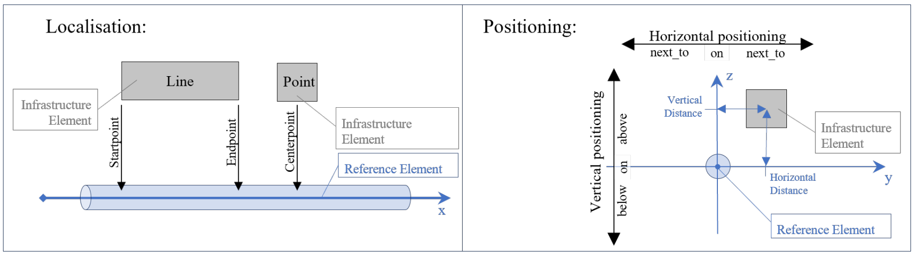

Figure 2: Assignment of road infrastructure elements to reference elements

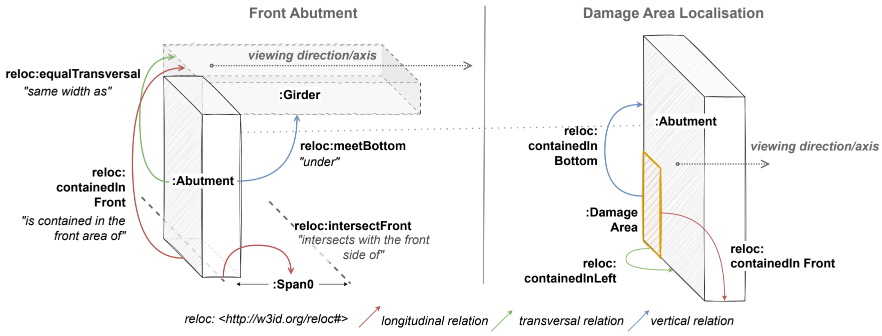

Figure 3: Spatial referencing of bridge components and damage damage areas

lute linear referencing as per ISO 19148.

Positioning describes the relation of the infrastructure element to the reference element in the transverse direction. Further differentiation is made between horizontal and vertical positioning and an optional distance, as shown in Figure 2.

All infrastructure elements that require linking are represented as nodes in the LPG called *InfraNetRelation* (Figure 1). This graph also contains nodes for representing reference elements, as well as localization and positioning elements. The infrastructure elements are connected to their corresponding reference elements through the localization and positioning nodes, which define the spatial relationships between the infrastructure element and the reference element. This graph is stored in an LPG database instance.

Using linear reference elements for modeling spatial relationships aligns well with the Anweisung Straßeninformationsbank (ASB), a federal regulation that sets out the data requirements for German infrastructure. The ASB mandates that all road infrastructure datasets include a reference to the ASB network established throughout Germany. To apply the approach to German infrastructure data, the sections ("Abschnitte") and branches ("Äste") in the ASB network can be used as reference elements. However, the implementation of ASB network references varies from system to system, as do the systems used for spatial structuring within the subsystems.

In Germany, the proprietary system ttSIB is the primary tool for managing road data. The data it holds can be accessed through a WebFeatureService. The main challenge lies in the lack of publicly available documentation on the data structure. As a result, a generic approach was taken, where all available feature types were queried. The objects were automatically extracted from each feature type and converted into LPG nodes labeled with the respective feature type. Additionally, references to objects contained in other feature types were interpreted as links between the nodes. This approach allowed the ttSIB data to be represented as a graph despite the unknown data structure (see figure 1 *ttSIB road graph*).

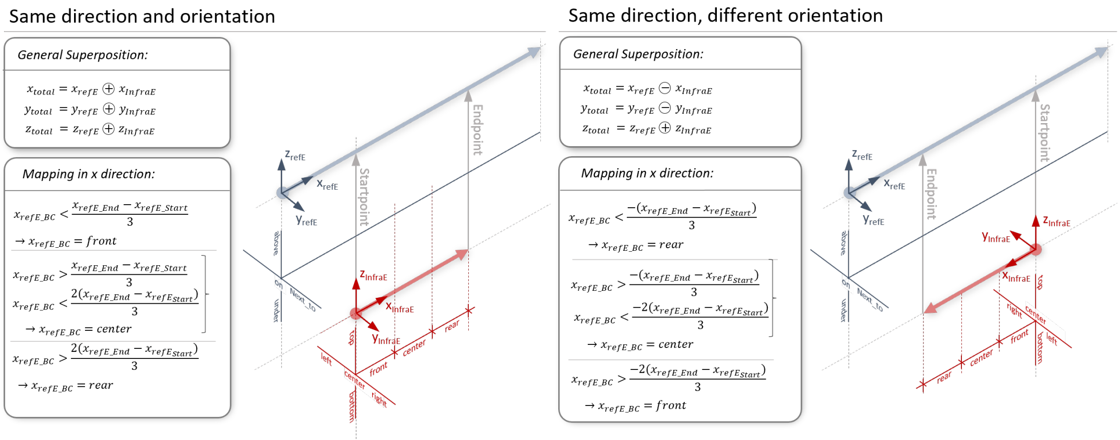

Figure 4: Mapping of the reference systems

3.2. Spatial reference system of the bridge data

The bridge information is originally stored in the relational database SIB-Bauwerke, containing administration, construction, and maintenance data, structured by the national data model for bridge maintenance data (ASB-ING). As the information system and the documentation process do not support a three-dimensional model representation of the bridge, the spatial assignment of information to a specific location is implemented by textual directional location indicators referring to the bridge's direction.

The bridge direction serves as the super-ordinate viewing direction for documenting component locations or damaged area positions, enabling the definition of "the front abutment" or "a damaged area at the front of the left railing". Also, the numbering of the spans follows this direction, starting with "Span 0" at the front of the bridge. The direction is defined in relation to the direction of the corresponding reference element of the road network (the same or reversed direction) and with an additional description referring to cities or cardinal directions (e.g., "from Cologne to Munich").

In previous work, a process was developed that converts the database entries for single bridges into RDF graphs (Göbels & Beetz, 2021) and transfers the implicit spatial information into explicit spatial relationships.

The Relative Location Ontology (RELOC) was developed to express the spatial relationships, which can reflect the relative directional-topological location description used in the existing documentation (Göbels & Beetz, 2024). Core elements of the RELOC Ontology are the directional terms, subdivided into three areas per axis: longitudinal axis: *Front, Center, Rear*, transversal axis: *Right, Center, Left*, vertical axis: *Top, Center, Bottom*. These directional terms are combined with the topological relations *Meet, Intersect, Contained in* and *Equal* to express detailed spatial arrangements. Figure 3 displays the resulting graph structure, expressing, for example, the spatial relationships of a "front abutment" to its neighboring components and the localization of a damaged area on a component. The underlying viewing direction is taken from the original bridge documentation and is part of the converted legacy bridge data graph.

For further processing and comprehensive querying, the converted SIB-Bauwerke RDF graph of each bridge data set is stored as a named graph in an RDF graph storage system (triple store).

4. Mapping Concept

To effectively query data across systems pertaining to diverse infrastructure elements and varying levels of detail, it is crucial to establish the capability to connect identical elements represented in the different graphs.

4.1. Object identification across systems

Distinctive identifiers are employed to link road infrastructure elements in the *InfraNetRelation* graph with related road objects in the *ttsib road* or the associated *SIB-BW bridge* graph. The method of implementing these identifiers varies based on the graph model used:
In the LPGs, the identifiers are upheld as attributes at the respective nodes. A property uniqueness and a property existence constraint are established for the corresponding attribute to ensure their uniqueness and presence for all nodes with the corresponding label. In the RDF graph storage system, each bridge data set is stored as an individual named graph with a unique Uniform Resource Identifier (URI). For convenient querying, the bridge ID is assigned to the respective graph in the metadata graph of the RDF storage. These IDs form the basis for comprehensive analyses by enabling the connection of information distributed across different systems to a specific road infrastructure element.

4.2. Mapping of the spatial reference systems

Since spatial structuring in infrastructure is often based on reference systems, aligning the different systems involves mapping their respective reference systems. Challenges stem from potential variations in the orientations of these reference systems in relation to each other, as illustrated in Figure 4, which displays two distinct possible configurations.
In the first scenario, the reference systems of the infrastructure element align with the overarching road infrastructure reference system in terms of orientation and direction. This constellation applies to the majority of road infrastructure elements. When an element referenced within an infrastructure element is superimposed with the referencing in the road infrastructure level, it results in an addition.
In the second scenario, the orientations of the reference systems are aligned, but their directions vary. This occurs in the context of bridges when the stationing direction of the internal reference system does not align with the direction of the reference element on which the bridge is located. Consequently, the signs change during superposition calculation, as illustrated in Figure 4.
Additionally, discrepancies in accuracy across the various reference systems must be considered. Different reference systems may employ distinct methods for describing spatial relationships. In this particular scenario, the reference system for the road infrastructure utilizes a quantitative approach to pinpoint an infrastructure element, contrasting with the qualitative description of spatial relationships in the longitudinal direction within the bridge's spatial reference system. Therefore, overlaying in the longitudinal direction also involves interpreting the bridge's qualitative location description into quantitative location descriptions within the road infrastructure system and vice versa.
In addition, it is necessary to establish a mapping of the terms used for the qualitative description of locations in both reference systems. As depicted in Figure 4, in the transverse direction, the term *on* in the road infrastructure reference system corresponds to *center* in the bridge reference system. Similarly, *above* is associated with *top*, while *under* corresponds to *bottom*. A combination of qualitative and quantitative location descriptions is utilized for the y-direction of the road infrastructure reference system. Specifically, *next_to* with a positive distance value is to be linked with *right*, and with a negative distance value to *left*.

The standard analysis procedure for spatial relationships between different infrastructure subsystems follows a multi-stage process. The road infrastructure systems with potentially relevant spatial relationships for the specific use case are identified initially. This can be based on specific road infrastructure elements (such as a particular bridge) or a specific constellation of infrastructure elements (such as all roads and bridges within an overlapping area on the same reference element with specific positioning information). Once the relevant subsystems for the use case have been identified, additional information can be retrieved within the subsystem-specific graphs using unique identifiers. If spatial relationships need to be considered when analyzing this subsystem-specific information, the positioning information is then superimposed.

5. Prototypical implementation concept

For the technical realization of the proposed concept, it is necessary to query data across different graph databases and graph database systems.

Composite databases in neo4j enable data querying across different LPG database instances. These databases do not store any data themselves. Instead, they hold references, referred to as aliases, to their constituent databases. Whenever there is a need to query data across multiple database instances, the CYPHER query is directed to the composite database, which then accesses the data of the individual components using the aliases. The bridge data graphs are stored in the RDF graph storage system (Apache Jena Fuseki Triplestore) as named graphs. Using the SPARQL query language[1], it is possible to query all contained graphs at once or filter for graphs fulfilling specific query patterns (e.g., having a specific identifier, name, or content).

A python processing script serves as middleware to facilitate a connection between the triplestore and the composite database. This script sends separated queries to the appropriate database servers based on the specific use case to be addressed (using CYPHER for queries to the neo4j instance or SPARQL for queries to the Triplestore). Subsequently, the server's response is further processed, if needed, to generate additional queries and/or superimpose the positioning information of the subsystems. Ultimately, the obtained information can be visualized in a user-friendly manner.

6. Proof of Concept

To validate the presented concept, real-world data from German infrastructure management is utilized. To apply the proposed concept, they are transformed into graphs through methodologies described in section 2. Moreover, a composite database was defined, enabling a comprehensive querying of the *ttSIB road* and the *InfraNetRelation* LPG graphs by the defined aliases. In this setup, the following use cases are implemented as a case study.

6.1. Comprehensive condition assessment of roadways on bridges

The carriageway on a bridge serves as both a structural component inspected during structural inspections, as well as a surface inspected in regular road inspections every four years (Villaret et al., 2015). While structural inspections only document damage as it occurs, standard characteristic values are determined during road inspections to assess the overall condition of the road surface. Consequently, two different descriptions of the road surface exist in two non-connected systems. In this use case, the information about carriageways on bridges from various systems is analyzed in a linked manner.

To do this, the roadway conditions documented for the bridges are first queried using SPARQL queries as seen in figure 5. For this purpose, damages to the respective bridges

[1] https://www.w3.org/TR/sparql11-query/

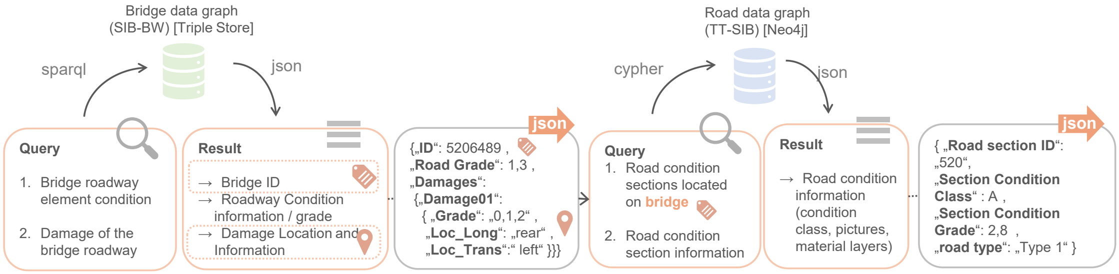

Figure 5: Comprehensive condition assessment of roadways on bridges

associated with the carriageway as a component are queried. The bridge ID and the damage entries found are returned. Based on the bridge ID, the road condition sections located on the respective bridge are then determined in the road infrastructure element reference element relation graph via a CYPHER query to the composite database. In the second step, the information available for this road condition section is extracted.

6.2. Correlation of road inclination and water-related bridge element damages

The German bridge data management system lacks information about the gradient of the bridges stored within it. Therefore, it is impossible to discern the low points of the carriageway in the vicinity of the bridge using only the bridge data. Nevertheless, identifying low points on bridges is crucial since these areas are particularly prone to water damage resulting from sealing damage. Consequently, the objective is to identify bridges that may have a low point and then analyze the occurrence of water damage for the components within the potential low point area.

As illustrated in figure 6, the road data is queried first to identify contiguous pairs of inclination objects, where the longitudinal inclination is negative in the first object and positive in the second object. This inclination change indicates a low point in the gradient in the area of these pairs.

Next, the road infrastructure reference element graph is searched for bridges that overlap with this previously identified area of interest. To this end, the bridges that overlap in the longitudinal direction with the point of the inclination change are queried using their relationships to the same reference element. The respective bridge ID is returned with the localisation of the low lowpoint on the reference element.

Using the identified bridge IDs and the mapping of the different reference systems used (as shown in figure 4), the corresponding bridge graphs in the RDF graph storage system can be queried for water-related damages situated longitudinally in the vicinity of the low point and transversely beneath or at the sealing. Additionally, it is possible to retrieve data on areas of the bridge that are more prone to water damage.

7. Results and Discussion

The use cases showed, through comprehensive analysis, how to check the consistency and plausibility of data for the same elements maintained in two separate systems on a large scale. The first use case demonstrates retrieving status information from the bridge and road management systems for the roadway sections on bridges. This approach allows for a more comprehensive picture of the road condition and could also function as an early warning system. For instance, if the road conditions observed during regular road inspections notably deteriorate compared to those documented during the last bridge inspection, it could indicate some settlements or other changes in the bridge's structure.

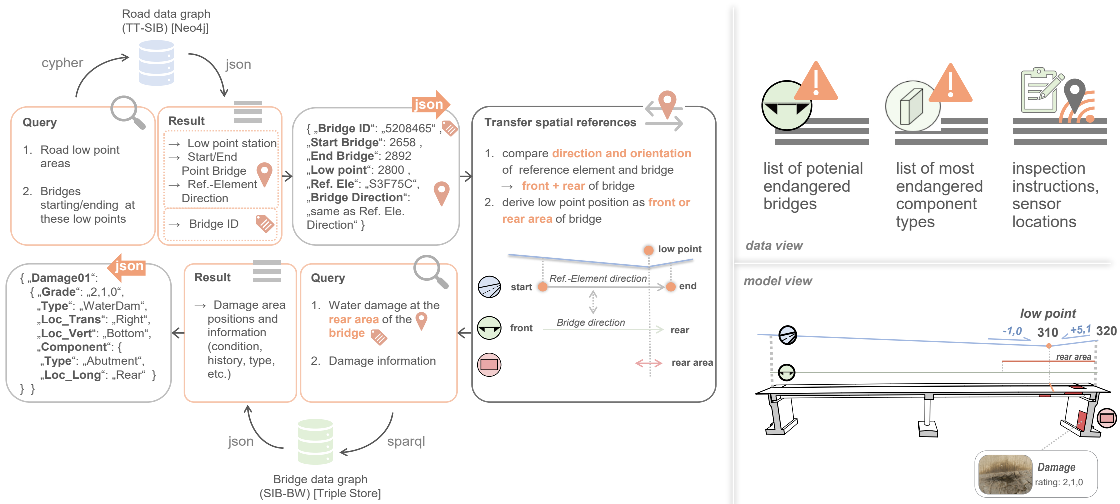

Figure 6: Correlation of road inclination and water-related bridge element damages

Furthermore, it is shown how the prerequisites for investigating a potential correlation between low points on roads and bridge element damage related to water in this particular area can be provided. The knowledge acquired could be applied in various ways. For instance, it could be integrated into a DT platform alongside other correlation analyses for bridge administrations. Additionally, the findings can raise awareness among bridge inspectors regarding specific damage to certain components or areas during inspections, thereby improving the likelihood of early damage detection. Moreover, the findings identify bridges and specific areas within those structures that are significantly more susceptible to water damage, which could aid in optimizing moisture sensor placements and usage. In addition to the proposed correlations, there are certainly other known and hidden correlations that could be explored through comprehensive data analysis of infrastructure systems.

The limitations of the current approach lie in the availability of road infrastructure data as graphs. To address this issue, segments of German road infrastructure data were automatically transformed into graphs. However, it is important to note that this data conversion occurred at a specific point in time and does not incorporate any subsequent updates made by the authorities to the original data. The identifiers utilized, which are also present in the legacy systems, enable the association of the graph representation with individual objects in the original data. Yet, maintaining consistency between the legacy data and its graph representation would presently necessitate a complete review of the legacy data for updates, which would be highly inefficient. An efficient method for ensuring consistency between infrastructure data in legacy systems and its graph representation would extend the approach presented and push it toward a digital twin.

8. Conclusion

This paper presents a method to enable comprehensive analysis of distributed infrastructure data using graph-based representations. We utilized bridge and road data resources with different data structures, converted into different graph representations. Our approach emphasizes the thorough examination of spatial relationships among infrastructure elements stored in diverse systems, specifically focusing on scalability to encompass entire networks rather than analyzing individual road infrastructure constellations.

There are opportunities to expand on the approach presented if additional infrastructure data using different reference systems, especially for spatial structuring, would be incorporated. On the one hand, this involves integrating systems from other countries, as was proposed in Beetz et al. (2018) by linking infrastructure data from the Netherlands and Germany. The approach outlined here can be merged with the data structure mapping presented in that paper. On the other hand, it is also possible to integrate systems that provide more detailed descriptions of spatial structures. For instance, extending the approach to the IFC data model is a viable option. Integrating such an approach holds promise, especially considering the imprecise location descriptions of components and damage in the data structures used in the legacy data, which complicates the derivation of precise damage and/or component positions. Therefore, applying the proposed method of linking heterogeneous data models of different granularity with different reference systems to solve problems of spatial querying in the context of BIM-GIS integration seems promising.

Acknowledgements

The presented research is part of the project "Digital twin for operating road infrastructure" funded by the Bavarian Ministry for Housing, Construction, and Transport and part of "Raumlink" funded by the DFG – Project number 501812634.

References

Beetz, J., Amann, J., & Borrmann, A. (2018). *Analysis of application possibilities of linked information (Linked Data) and ontologies and related technologies (Semantic Web) in the road sector* (Abschlussbericht). Bundesanstalt für Straßenwesen (BASt).

Göbels, A., & Beetz, J. (2021). Conversion of legacy domain models into ontologies for infrastructure maintenance. *Proceedings of the 9th Linked Data in Architecture and Construction Workshop - LDAC2021*, *Vol-3081*, 12.

Göbels, A., & Beetz, J. (2024). Relative Location Ontology: An ontological model for representing directional topological relationships between spatial entities in oriented space. *LDAC 2024 – 12th Linked Data in Architecture and Construction Workshop.*

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

Luiten, B., Böhms, M., Alsem, D., & O'Keeffe, A. (2018). Asset information management using Linked Data for the life-cycle of Roads. *Life Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision: Proceedings of the Sixth International Symposium on Life-Cycle Civil Engineering (IALCCE 2018)*, 7. https://doi.org/10.1201/9781315228914

Tulke, J., Schäfer, R., Brakowski, A., Braun, J.-D., & AG, H. S. (2016). *Intelligente Straßenverkehrsinfrastruktur durch 3D-Modelle und RFID-Tags* (tech. rep. No. B 130). Bundesanstalt für Straßenwesen (BASt). Bergisch Gladbach.

Villaret, S., Fohböse, B., Jähning, J., Karcher, C., Niessen, J., Buch, M. T., für Straßenwesen, B., & für Neue Wissenschaft, W. N. V. (2015). *Ermittlung von repräsentativen substanzwerten in homogenen abschnitten.*

Weise, M., Schmidt, I., & Hettwer, J. (2018). OKSTRA®, IFC und BIM: Neue Wege im Umgang mit Straßeninfrastrukturdaten. *7. OKSTRA-Symposium 15./16. Mai 2018 Bergisch Gladbach.*

Zhu, J., & Wu, P. (2022). BIM/GIS data integration from the perspective of information flow. *Automation in Construction*, *136*, 104166. https://doi.org/10.1016/j.autcon.2022.104166