Towards a Common Digital Space: Proposing a Schema for Spatially Linking Heterogeneous Resources

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

Oliver Schulz, (schulz@dc.rwth-aachen.de)
Desgin Computation, RWTH Aachen University, Germany

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

Abstract
The spatial alignment of digital resources is necessary for most processes in the life cycle of an asset in the built environment. However, there is currently no established means to explicitly describe the spatial relationships inherent in the resources in a machine-readable and vendor-neutral way. Therefore, the paper presents a schema for superimposing heterogeneous data in digital spaces. The schema defines four main space types inherent in the Architecture, Engineering and Construction industry: Global, Asset, Document and Entity Space. Global Space represents the asset's physical location, while the Asset Space is the digital replica of the built object superimposing all related building resources. Each resource defines its own Document Space, which in itself contains one to many Entity Spaces. An Entity Space is a coherent unit of information, such as a section or a model. Moreover, the schema provides a means to express the relationships between the different space types. To verify the approach, we demonstrate the application with a compact bridge data set. This work should serve as a preliminary step towards automating the spatial linking of heterogeneous resources in the built environment.
Keywords: Spatial Linking, Heterogeneous Data, Common Data Environment

1. Introduction
The design, construction, and operation of buildings or infrastructure assets involve many heterogeneous resources covering different aspects of the assets and using different representation formats. Often, the data is distributed among various stakeholders with different provenances.
Every process in an asset's life cycle utilises many complementary, heterogeneous documents that are spatially related. However, this spatial relationship is often not explicit, so it must be re-established in time-consuming processes. This is only possible if extensive knowledge of the asset is available.
The implicit spatial interconnections between the resources are only derivable for domain experts superimposing the data on the respective asset. However, the amount of

data and the digital and computer-supported processes require implementing a machine-readable way of expressing and querying spatial interrelations.
Even though mathematical solutions – such as transformation matrices, coordinate reference systems, etc. – exist to describe spatial references, the application of these approaches for organising heterogeneous asset documentation files spatially in the Architecture, Engineering and Construction (AEC) domain requires further research.
These solutions are widely adopted in other domains, such as Geographic Information Systems (GIS) or Computer Vision (CV). However, in contrast to the precise values in GIS, the spatial references in existing asset data are often vague and do not always require exact coordinates described in a matrix. In some cases, this is not even possible.
Furthermore, documents in AEC follow an inherent spatial pattern that derives from a spatial connection to the built asset. For example, section plans represent a fixed projection of the building, implicitly superimposing their content with floor plans and models. Besides their inherent spatial relation to the asset, entities such as floor plans, sections, and geometry can also have a spatial relation to the document they are stored in. For example, a section is contained in the left part of a plan, which combines different views.
To represent these spatial relations in a machine-readable format, this paper proposes a metadata schema. This schema defines the various types of space associated with AEC projects and assets, allowing their spatial relationships to be expressed and providing different methods for articulating these relationships. The schema development is based on findings identified by investigating heterogeneous bridge data sets containing approx. 1,500 resources.
We demonstrate the application of our approach on a subset of a bridge data set and discuss the potential for automatically creating spatial links using the presented schema. This work is part of ongoing research and will be further validated in future work.

2. Related Work

The following section provides an overview of state-of-the-art research and technologies for data organisation and gives insights into existing (research) approaches for spatially referencing and querying asset data of the AEC domain.

2.1. Common Data Environments

Common Data Environments (CDEs) are web-based platforms that serve as central access points for all building project-related data (Preidel et al., 2021). While the overall concept of a CDE is defined in the ISO 19650 (ISO, 2018), a more fine-grained list of features and functionalities of CDEs is described in the German DIN SPEC 93191 (DIN, 2019). From a technical point of view, no specific details are provided for the CDEs. Nevertheless, the standards define CDEs as being based on the concept of containers, though the technical specifications of these containers are not defined.
In the AEC industry, various providers offer CDEs based on the standards. However, the technical implementations are mostly unknown and product-specific. In research, these containers are used in different ways, with two possibilities: a file-based implementation using the Information Container for linked Document Delivery (ICDD) (DIN Standards Committee Building and Civil Engineering, 2021) or a web-based implementation using the Linked Data Platform (LDP) (Mihindukulasooriya & Menday, 2015).
Approaches for CDEs applying ICDD (Senthilvel et al., 2020) or LDP (Werbrouck et al., 2023) exist. While both approaches can store and link documents and provide meta-information about their resources, the lack of explicit spatial information to establish the spatial relationship between the documents is evident. A first attempt to address this issue has been made in Schulz et al. (2023), where a scene graph structure was proposed

for CDEs based on Linked Data. The spatial connection of the different files is managed with a parent-child relationship between the different resources.

2.2. Spatial concepts in BIM

In the AEC industry, spatial references are often implicitly created between the various heterogeneous documents for existing buildings but are not archived explicitly in the long term. For example, in Building Information Modeling (BIM) processes, the data from the building surveys is used, which is recorded as pictures, plans, point clouds, etc. In some cases, this data is also referenced directly in the BIM authoring software to serve as a basis for the modelling process (Petzold & Rechenberg, 2021). This means that the spatial overlay is often present in the authoring software but not transferred to the CDE, which means it is not archived for later use.

In BIM projects, the superimposition of the federated BIM models is solved via model coordination. This can be achieved by exchanging a coordination file that defines the project origin for all planners (Schäferhoff et al., 2021) before the various stakeholders start constructing the models, thus ensuring a spatial superimposition of all models.

Furthermore, the plans are generated from the BIM and represent a spatial 2-dimensional abstraction of the model within the authoring software, which is thus spatially and geometrically linked to the 3D model. However, the spatial link is usually lost when these plans are exported from the application as documents.

Regarding BIM's spatial searchability, approaches such as QL4BIM (Daum & Borrmann, 2014) and BIMSPARQL (Zhang et al., 2018) provide mechanisms for spatial queries. While QL4BIM implements its own query language, which is based on the point-set theory (Daum & Borrmann, 2014), BIMSPARQL is an extension for the W3C recommendation of the SPARQL Query Language for RDF (SPARQL)[1]. It has a thematic relation to the Open Geospatial Consortium (OGC) GeoSPARQL[2] extension for SPARQL, which allows spatial queries to be made in a geospatial context. The main difference between BIMSPARQL and GeoSPARQL in spatial queries is that GeoSPARQL only allows queries in 2D space (longitude and latitude) and uses different Spatial Reference Systems. In contrast, BIMSPARQL only queries in a 3-dimensional Cartesian space. However, the spatial queries of both approaches, QL4BIM and BIMSPARQL, focus on querying elements and concepts inside the BIM or IFC model and not on heterogeneous representations of the buildings and their spatial extents.

Integrating spatial information expressed in textual location descriptions into BIM models is investigated in Göbels et al. (2023). The textual documentation of damaged area locations during bridge inspection is processed to identify the affected model component and to create a geometric representation of the damage. The process works with rule-based transformations of location terms, such as *top*, *left*, or *front*, interpreting each directional area of a component (e.g., "the front part") as one-third of its total extension in this direction. However, the approach converts the textual location data into a geometric representation. It only implements semantic links between the inspection data and the BIM model and does not focus on expressing the spatial transformation itself.

[1] SPARQL Query Language for RDF: https://www.w3.org/TR/rdf-sparql-query/ accessed: 28.05.2024

[2] OGC GeoSPARQL - A Geographic Query Language for RDF Data: https://docs.ogc.org/is/22-047r1/22-047r1.html accessed: 28.05.2024

Table 1: Details about the two bridge data sets

Category	Dataset 1	Dataset 2
Bridge built in	1949 [1890, 2005]	2002
Bridge Length	351 m	57 m
Nr. Floor/Overview plans	73	4
Nr. Section plans	159	11
Nr. formwork, reinforcement, tendon plans	204	17
Nr. Detail plans	81	23
Nr. Models	2 [35 part models]	1
Nr. Point clouds	5	1
Nr. Pictures	646	4
Nr. Inspection/Damage Pictures	251	45

3. Method

Before developing a schema for spatial relationships in heterogeneous documents for bridge assets, the content of two different bridge data sets, consisting of multiple plans, pictures, models, and point clouds, was analysed. The content of the data sets is depicted in Table 1.

The first data set is from a bridge built in 1949. It replaced a destroyed bridge built in 1890, and in 2005, a general overhaul of this bridge took place. As part of the research project, which also includes this paper, up-to-date point clouds and 3D models were created. Thus, data from over 100 years of documentation are available, representing a broad spectrum of analogue and digital documentation methods and file formats.

The second data set is from a highway bridge built in 2002. It represents a more standard application use case with less complex documentation and history.

After analysing the content, the findings F1 to F8 were identified.

Finding 1 (F1): *Each data set is part of a project concerning a built asset. All documents in the data sets relate to space since they represent an aspect of the asset located in physical space.*

Finding 2 (F2): *All the documents in the data sets represent the real assets. Each document focuses on a different aspect or domain. As a result, two documents can depict the same physical location of the asset but contain different information. Depending on the phase and purpose, the documents are adjusted in their scale or level of detail.*

Finding 3 (F3): *Spatial asset data is represented by different types of entities, such as pictures, technical drawings, models, and point clouds.*

Finding 4 (F4): *When spatially representing entities in a three-dimensional space, each entity can be described using a boundary type described by a volume, an area or a point.*

Finding 5 (F5): *Each entity has an implicit spatial relation to the asset. Specific types of entities have predetermined constraints on this spatial relation, e.g., a 2D section intersects the asset vertically. The relation of the entity to the asset space can be defined in different ways by referring to:*

- *an internal reference system (e.g., construction axes, section markers)*
- *geo-locations (e.g., cardinal points, city or landmark references)*
- *directional indications (e.g., the front abutment)*
- *a transformation matrix (e.g., an alignment of origins/reference points)*

The spatial relation is often vague, and one entity, like a standard construction detail, can have multiple spatial relations with the asset.

Finding 6 (F6): *2D construction plans can contain various views of the asset, referring to different locations or having different spatial relations to the asset, for example, a plan combining floor plans, sections and details.*

Finding 7 (F7): *2D construction plans have in common that they are usually orthogonal to either the horizontal- or the vertical axis. In the data sets of the bridges, it was also found that views and sections are often described as longitudinal or transversal, meaning they are orthogonal to the main or secondary axis of the bridge.*

Finding 8 (F8): *The information space, e.g. the depth of a section plan or a view, is often bigger than the space of the entity.*

3.1. Space type schema

Based on the Findings, four main space types are defined: the *Global Space*, the *Asset Space*, the *Entity Space*, and the *Document Space* (see Figure 1). The *Global Space* is the actual physical space represented by a geodetic coordinate system.
The *Asset Space* represents the actual, physical asset located in the *Global Space*. As the asset is the central connecting element (F1) of all resources representing it, the *Asset Space* is the superordinate instance for the spatial superimposition of the *Entity Spaces*. The boundaries of the *Asset Space* have a spatial extension that encloses all elements belonging to the built asset (F4).
The Entity Space is a coherent unit of information covered by a single resource, such as a 3D model, a 2D drawing, or a picture (F3), representing its actual dimensions (F2) referring to the *Asset Space*. It can have a boundary represented as a volume, area, or point geometry (F4). The *Entity Space* has one (or more) spatial relation(s) to the *Asset Space* (F5). These can either be described using Natural Language (NL) or a transformation matrix (Figure 2). An *Entity Space* can also relate to another *Entity Space*, for example, if 2D plans are derived from a 3D model. Moreover, *Entity Spaces* can also represent concepts derived from descriptions or data sets that do not have a modelled geometry but describe an element in the asset.
The *Document Space* represents the extent of a document or file that contains *Entity Spaces*. Often, the *Document Space* uses a down-scaling process to represent the Entity Space (2D plans, pictures). A *Document Space* can represent different, disjointed *Entity Spaces*, such as a collection of sections and details on a construction plan. (F6)
The relation does not have to be explicitly defined if the *Entity Space* and its *Document Space* are the same. If a document displays multiple *Entity Spaces*, they each refer to a specific part of the *Document Space*.
The *Asset*, *Entity*, and *Document Space* need a defined orientation — i.e., the direction of their internal coordinate reference systems — to formulate their spatial interrelations.
In addition, it is assumed that the origin $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$ of every space is its centre of mass, except the space provides its own internal coordinate system.

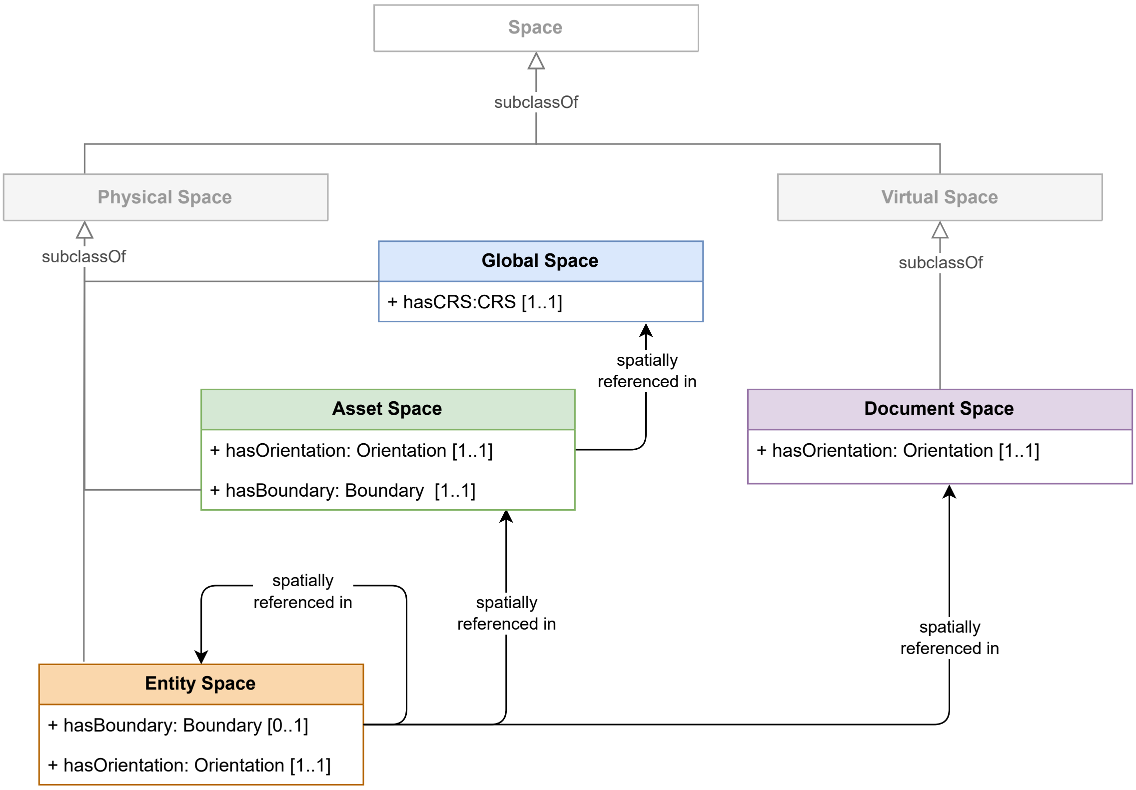

Figure 1: UML schema of the different identified space concepts in AEC

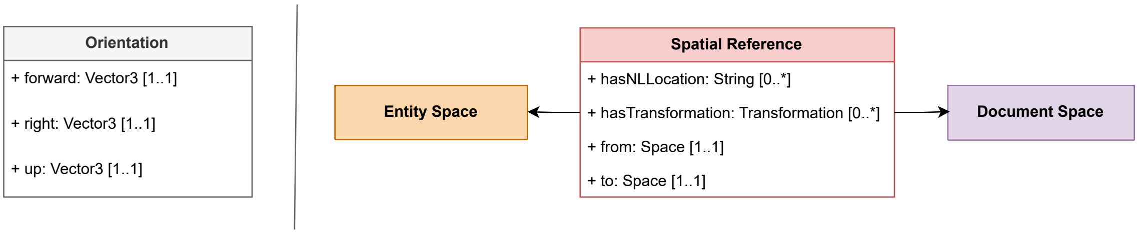

Figure 2: Definition of the *orientation* and *spatial reference* object

3.2. Definition of specific entity space types

The presented schema enables the explicit definition of spatial relations between an asset's heterogeneous resources. As the schema should serve as a first step towards a (semi-)automatic localisation and arrangement of resources, we further defined specific types of *Entity Spaces*. Each type sets particular constraints on the spatial relation to the asset, thus limiting the parameters, such as the axis or scales for the transformation.

The sub-classes of the *Entity Spaces* are defined based on their spatial extent and direction, resulting in the *Volume-*, *Area-*, and *Point Space*. The *Area Space* is further subdivided into *Vertical-* and *Horizontal Area Space*, which represent the projections of section- and floor plans.

The *Volume Space* – e.g., 3D and BIM models and point clouds – is usually created at true scale so that one meter in the model corresponds to one meter in the physical space, thus pre-defining the scaling parameters for the transformation.

A *Point Space* only requires the definition of a translation vector.

Automatic processing is specifically applicable for 2D documentation since spatial constraints can be obtained from the naming of the documents (F7). For example, a drawing labelled with *Cross Section* implies a parallel alignment of the plan with the vertical and

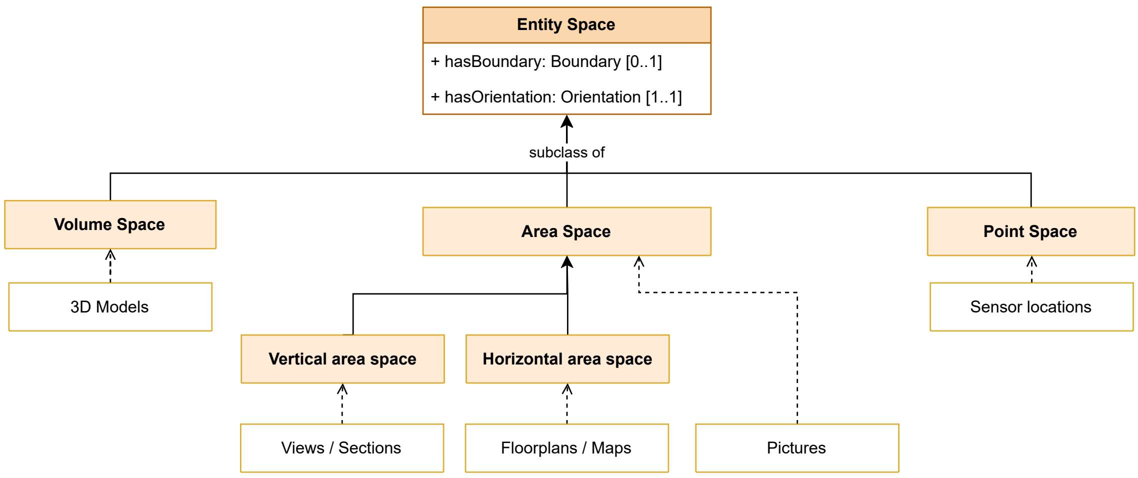

Figure 3: Subdivision of the entity space into defined subcategories.

transversal axis of the asset and an orthogonal alignment to the longitudinal axis of the asset. Also, the alignment of the drawing to the internal reference system of the document is typically orthogonal.
Therefore, the rotation values can be easily predicted by categorising 2D plan resources into a specific sub-type of *Area Space*. To further automate the categorisation, transformation and alignment of the plan, content recognition processes can be used.

4. Proof of Concept

The schema is applied to a subset from bridge data set 1, which contains a 3D model and an overview plan superimposed in an *Asset Space*. The bridge's *Asset Space* is represented by a bounding box encompassing the entire space occupied by the bridge. In this *Asset Space*, the *Entity Space* of a partial bridge model is positioned, as well as the *Entity Space* of a cross-section extracted from the *Document Space* of a plan document. The cross-section is placed on the right centre of the plan and represents the regular superstructure of the bridge.
As we see the application in container-based environments, such as CDEs, the schema has been integrated into a Linked Data mock-up. The prefixes used in the mock-up do not represent an actual ontology since this work is rendered out of scope.
An abstract representation of the spatially arranged data set can be seen in Figure 4. Listing 1 presents a serialisation of the scene in Turtle[3]. The model's orientation is specified as a 3x3 matrix with a row-major order describing the forward, up and right vectors. The transformations consist of a 4x4 affine transformation matrices. For the Turtle serialisation, the notation of the matrices was chosen as a string, in which the numbers are listed in a row and separated by commas.

$$orientation = \begin{bmatrix} forward \\ right \\ up \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 1, 0, 0, 0, 1, 0, 0, 0, 1$$

[3]RDF 1.1 Turtle, W3C: https://www.w3.org/TR/turtle/ Accessed 08.05.2024

Listing 1: Snippet of the dataset serialised in Turtle to describe how the crosssection is derived from the plan and aligned with the bridge.

```
<Bridge> a space:Asset ;
   space:hasBoundary "175, 7.5, 16" ;
   space:hasOrientation "1, 0, 0, 0, -1, 0, 0, 0, 1" .

<Crosssection> a space:Entity ;
   space:hasBoundary "5.7, 8.3, 0" ;
   space:hasOrientation "-1, 0, 0, 0, 1, 0, 0, 0, 1" .

<SpatialReference1> a space:SpatialReference;
   space:NLLocation "The cross-section cuts through the bridge in the centre of the front segment of the bridge" ;
   space:hasTransformation "0, 0, -26.44, 45.44, 0, 26.44, 0, 0, 26.44, 0, 0, 8.2, 0, 0, 0, 1" ;
   space:from <Crosssection> ;
   space:to <Bridge> .

<Plan> a space:Document ;
   space:hasOrientation "1, 0, 0, 0, 1, 0, 0, 0, 1" .

<SpatialReference2> a space:SpatialReference ;
   space:NLLocation "The cross-section is in the centre right of the plan" ;
   space:hasTransformation "1, 0, 0, 0.21, 0, 1, 0, -0.65, 0, 0, 1, 0, 0, 0, 0, 1" ;
   space:from <Plan> ;
   space:to <Crosssection> .
```

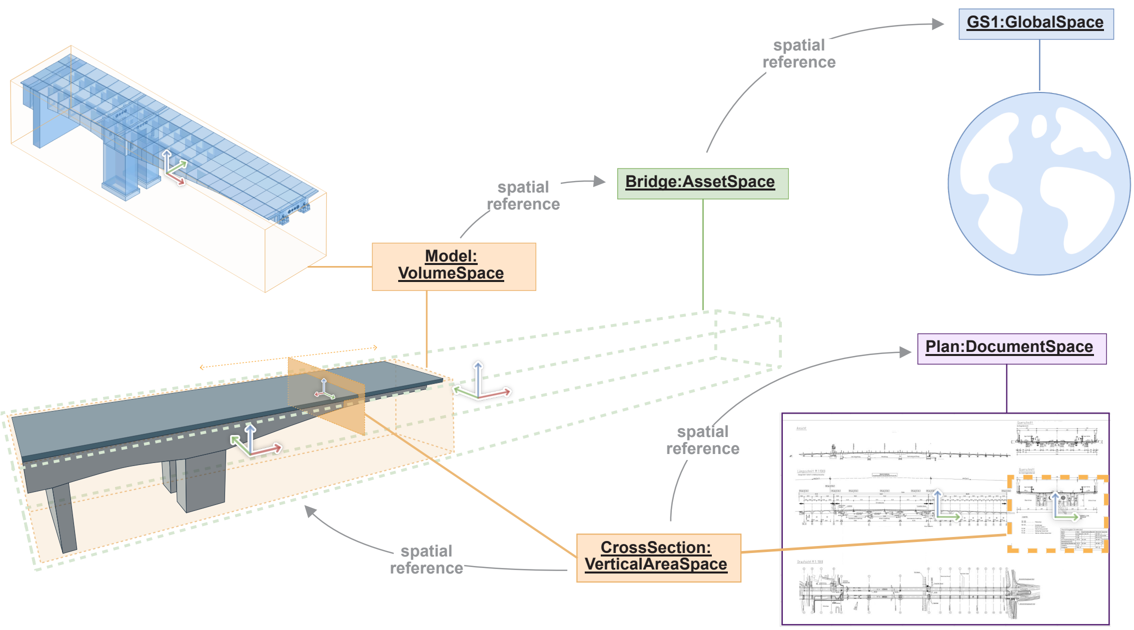

Figure 4: Abstract visualisation of different documents and their *Entity Spaces* from data set 1, spatially referenced in the *Asset Space*.

5. Discussion

Based on the Finding (F1 to F7), we identified the different types of space that are inherent in bridge construction projects and formalised them using a UML schema.

In a proof of concept, the schema was applied to a compact data set, describing the spatial relations of a partial 3D model and a cross-section to the *Asset Space*. The transformations were expressed by NL as well as 4x4 matrices. It has shown that both approaches complement each other since NL is well suited to express vague location descriptions

and is comprehensible for humans. In contrast, transformation matrices can represent precise coordinates in software applications and are machine-readable.
Although the schema could be applied to the bridge data, edge cases still need to be considered where *Entity Spaces* do not map into physical space, e.g., when using jogged sections or exploded-view drawings. To identify further limitations and validate the approach, we have to apply the schema on a larger scale in the future.
Furthermore, we have only analysed bridge data. Whether this schema can be transferred to other domains in the AEC industry must be further investigated. We suspect that assets such as buildings will pose further challenges, as bridges have an inherent, standardised spatial structure compared to other built assets.
To test the approach further with CDEs, we intend to integrate the schema with existing container concepts such as ICDD and LDP and investigate its compatibility with other metadata schemas.
Finally, the following steps involve creating an automated sorting of entities based on their classification using content recognition processes.

6. Conclusion

This work has created the basis for explicitly expressing the spatial knowledge implicit in bridge data as a schema. Thus, a step towards machine-readable spatial descriptions for heterogeneous data has been created.
In this work, the spatial arrangement of the data has been done manually. However, this work is an essential basis for automatically generating spatial links between heterogeneous data. This should make extensive data sets more accessible and, therefore, achieve a higher usability of existing data.
This is a step towards spatial CDEs and container-based systems that can be reused over the life cycle of the building, preserve spatial and non-spatial knowledge, and enable spatial queries in addition to the searchability of semantic knowledge.

Acknowledgements

This research is part of the Raumlink project funded by the German Research Foundation (DFG) – Project number 501812634. We would like to thank the SPP100 Plus project for providing us with the necessary data for this research.

References

Daum, S., & Borrmann, A. (2014). Processing of Topological BIM Queries using Boundary Representation Based Methods. *Advanced Engineering Informatics*, *28*(4), 272–286. https://doi.org/10.1016/j.aei.2014.06.001

DIN. (2019). DIN SPEC 91391-2: Common Data Environments (CDE) for BIM projects Function sets and open data exchange between platforms of different vendors Part 2: Open data exchange with Common Data Environments.

DIN Standards Committee Building and Civil Engineering. (2021, July). Information container for linked document delivery Exchange specification Part 2: Link types (ISO 21597-2:2020); German version EN ISO 21597-2:2020. https://doi.org/10.31030/3192763

Göbels, A., Rivadeneyra, F., & Beetz, J. (2023). Transfer of Implicit Semi-Formal Textual Location Descriptions in Three-Dimensional Model Contexts. *2023 European Conference on Computing in Construction - 40th International CIB W78 Conference*. https://doi.org/http://www.doi.org/10.35490/EC3.2023.268

ISO. (2018). ISO 19650-1:2018 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) - Information management using building information modelling - Part 1: Concepts and principles.

Mihindukulasooriya, N., & Menday, R. (2015, April). Linked Data Platform 1.0 Primer. https://www.w3.org/TR/ldp-primer/.

Petzold, F., & Rechenberg, B. (2021). BIM und Bauen im Bestand. In A. Borrmann, M. König, C. Koch, & J. Beetz (Eds.), *Building Information Modeling: Technologische Grundlagen und industrielle Praxis* (pp. 507–532). Springer Fachmedien. https://doi.org/10.1007/978-3-658-33361-4_26

Preidel, C., Borrmann, A., Exner, H., & König, M. (2021). Common Data Environment. In A. Borrmann, M. König, C. Koch, & J. Beetz (Eds.), *Building Information Modeling: Technologische Grundlagen und industrielle Praxis* (pp. 335–351). Springer Fachmedien. https://doi.org/10.1007/978-3-658-33361-4_16

Schäferhoff, G., Jäppelt, U., & Blume, T. (2021). BIM im Verkehrswasserbau. In A. Borrmann, M. König, C. Koch, & J. Beetz (Eds.), *Building Information Modeling: Technologische Grundlagen und industrielle Praxis* (pp. 687–700). Springer Fachmedien. https://doi.org/10.1007/978-3-658-33361-4_35

Schulz, O., Werbrouck, J., & Beetz, J. (2023). Towards Scene Graph Descriptions for Spatial Representations in the Built Environment. *Proceedings of the 30th EG-ICE International Workshop on Intelligent Computing in Engineering*.

Senthilvel, M., Oraskari, J., & Beetz, J. (2020). Common Data Environments for the Information Container for linked Document Delivery [https://ceur-ws.org/Vol-2636/10paper.pdf]. *Proceedings of the 8th Linked Data in Architecture and Construction Workshop*.

Werbrouck, J., Schulz, O., Oraskari, J., Mannens, E., Pauwels, P., & Beetz, J. (2023). A generic framework for federated CDEs applied to Issue Management. *Advanced Engineering Informatics*, *58*, 102136. https://doi.org/10.1016/j.aei.2023.102136

Zhang, C., Beetz, J., & de Vries, B. (2018). BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data (Á. Sicilia, P. Pauwels, L. Madrazo, M. Poveda Villalón, J. Euzenat, Á. Sicilia, P. Pauwels, L. Madrazo, M. Poveda-Villalón, & J. Euzenat, Eds.). *SW*, *9*(6), 829–855. https://doi.org/10.3233/SW-180297