CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0032

FROM DATA DIVERSITY TO NORMS: STANDARDIZATION STRATEGIES FOR CONSTRUCTION LOGISTICS

Christoph Mönicks, Sabrina Puslat, Bert Leerkamp
University of Wuppertal, Wuppertal, Germany

Abstract

The construction industry is one of the least digitized industries. According to this, there are a large number of different digital solutions for controlling and managing construction logistics. For this reason, it is currently not possible to carry out an overarching data analysis to identify optimization measures in construction logistics. As part of this study, datasets from five construction logistics companies were analysed and compared. The datasets contain information on registered supply and disposal transports, including e.g. arrival times, material volumes and transport vehicles. Aim of the study is to standardize the data structures and collection processes.

Structural analysis is used to identify similarities and differences between the datasets. The focus is on mandatory fields, input types (e.g. free text, numerical values, drop-down menus), and attribute categorization (e.g., material type, size, quantity, packaging). The study also examined restrictions underlying the recorded attributes. The evaluation of data quality showed that while formal accuracy is high, the content quality varies significantly. For example, the 'material type' field is consistently filled, but implausible entries restricts the usability.

In addition, process analysis were conducted to assess data collection methods. The results show that the data entry processes differ depending on the provider, which is partly due to business models, but also to a lack of standardization. Variations exist in the order of data collection, stored attributes, and validation measures. Standardizing these elements through a DIN guideline could significantly improve the data quality. Finally, recommendations are given for standardized data structures and the collection process. A standardized approach would not only enhance scientific data evaluation but also optimize logistics processes in the construction sector and enable accurate traffic forecasting for urban and transport planning.

Keywords: construction logistics, data quality, data standardization, logistics data, optimization potential.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

To ensure a data flow between different stakeholders unified data standards within construction industry are an essential part to contribute to connected mobility and logistics. Digitalization in the construction industry is currently lagging behind, resulting in increasing conflicts in logistics processes [1]. Furthermore, from both planning and academic perspective, the lack of data makes it nearly impossible to properly assess the impact of construction transports. [2].

The research project, STArLOG - Data standards for construction logistics, co-funded by the European Union and the Ministry of the Environment, Nature and Transport of the State of North Rhine-Westphalia, tackles the missing data standardization and aims to develop a data standard for managing and controlling construction logistics. In future, the results will enable uniform data exchange between stakeholders on construction sites and form the basis for optimized, sustainable management of construction logistics and also enables further research. [3]

This paper focuses on data quality and data collection processes employed by different construction logistics companies. Each company operates its own logistics portal to manage the supply and disposal transports of their respective construction sites First, the partners' logistics data is analysed to identify similarities and differences and to assess data quality. The process of data collection by the partners is also examined in order to identify deficits and potential. As a result, data requirements of stakeholders

Corresponding author email address: moenicks@uni-wuppertal.de

from practice and research are defined in order to integrate them into the data standard. For example, no attributes are currently recorded to estimate the environmental impact of construction logistics. These include e.g. vehicle emission classes, distances travelled or origin-destination relationships.

2. Methodology

In the following, we describe the data provided by all partner companies. The sample sizes of the datasets vary between approximately 4,300 and 26,000 transports. Each project partner supplied data from 4 to 23 construction sites, sourced from their respective logistics portals. The datasets include detailed information on transport activities related to the supply and disposal of construction sites, such as arrival times, material types, material quantities, and vehicle types associated with each transport. In total, we examined data from 54 construction projects, comprising approximately 66,000 individual transports. .

First, a structural analysis of the datasets was conducted to identify similarities and differences between the data structures. The focus of this analysis was on mandatory fields, input types (e.g., free text fields, drop-down menus), and the categorization of key attributes (e.g., material types, quantities, vehicle types). In addition to the structural analysis, further topics were captured through a survey. In addition, the following topics were recorded via a survey:

- Orientation of the project partner's company. (main activities of the companies, provision of software vs. construction logistics process planning).
- Data collection processes and verification of the data.
- Challenges and added value of data collection.
- Detailed description of the recorded attributes.

In addition, interviews were conducted with the partners in order to clarify discrepancies and gain a detailed insight into the logistics systems.

The combination of these three methods - standardized survey, qualitative interviews and the data analysis - formed the basis for a well-founded evaluation of the construction logistics process data (Figure 1).

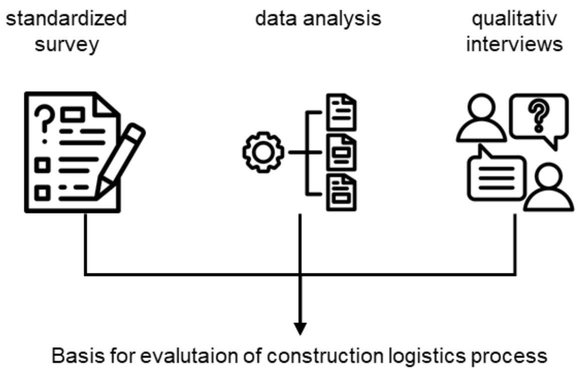

Figure 1: Methodological approach of the analysis

3. Data structures and data quality

3.1. Analysis of data structures

The detailed analysis of the five partners' data structures revealed a fundamental distinction between two main categories – construction site information and transport information –, which are described in the following.

<u>Construction site information</u>

Contains key information about the construction project, such as project title, location, construction volume, gross floor area and other project-specific attributes. The construction site information was differentiated into two sub-categories (see Figure 2, blue):

1. Information about the construction project: This includes basic information such as location, name of the construction project or gross floor area.
2. Configuration in the booking portal: This information includes available resources such as cranes or storage areas, layout plan of the construction site, opening hours, etc. The visibility and level of detail of this information can be customized.

Transport information

Transport information refers to planned transports for the respective construction sites. The subcontractors usually enter this information in the logistics portal for each transport. It contains information such as material type, material quantity, vehicle types, etc. A categorization was developed for the systematic analysis of the information. This consists of eight main groups (see Figure 2, green):

1. General information (e.g. fee-text comments)
2. Vehicle (e.g. type of the vehicle used)
3. Delivery times (e.g. scheduled arrival time and expected duration of stay)
4. Resources (e.g. required technical equipment such as cranes or lifts)
5. Suppliers (e.g. company information of the delivery company)
6. Recipient/trade (e.g. data of the client or subcontractor)
7. Material (e.g. type and quantity)
8. Transport direction (e.g. delivery or collection)

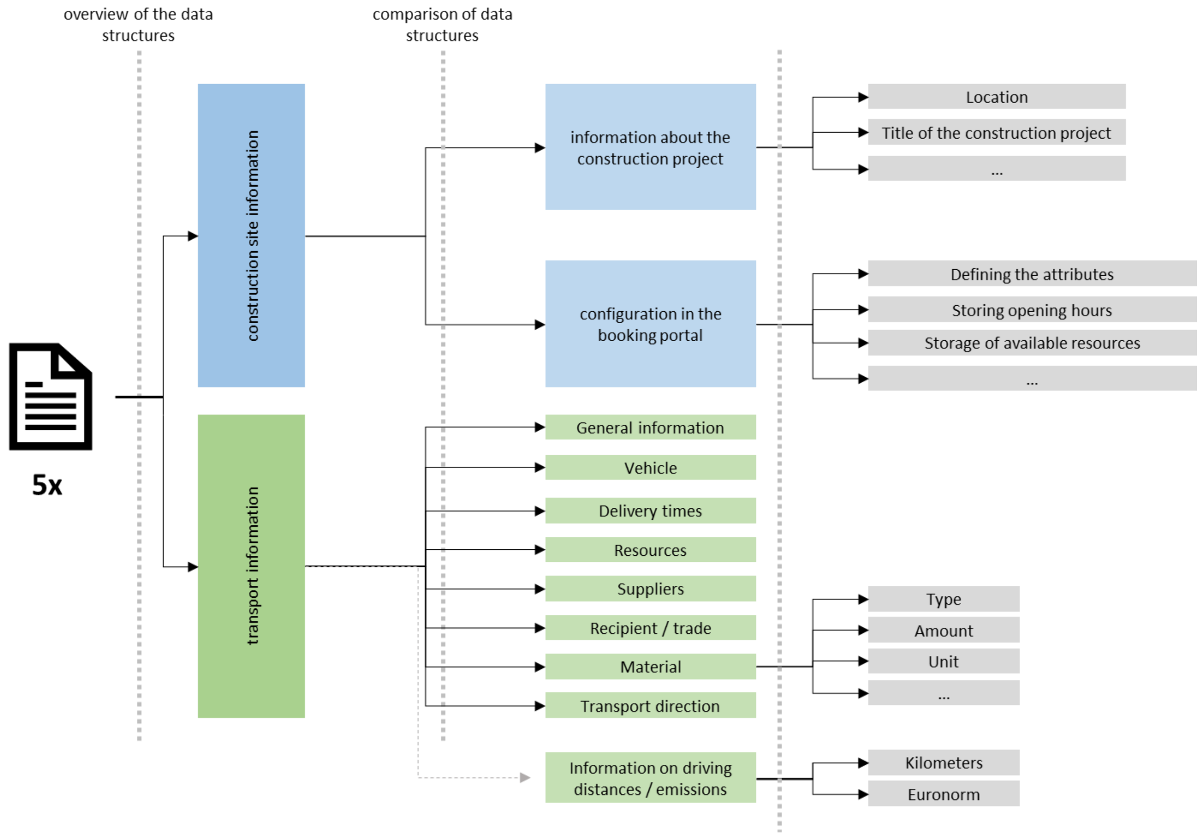

Figure 2: Data structures in construction logistics

The revealed categorization provide a good foundation for standardization. The data already provides a sufficiently accurate basis of information about the construction project and focuses the logistics processes on construction sites. The data structure is clear and practical to use. For example from the client's perspective, the most important information in a transport booking are the delivery time and the required resources (e.g. crane). This applies to the specific systems of the providers, which function well in themselves, but pose challenges for users due to their individuality. Environmental attributes and traffic-generating parameters are currently not reflected in the data structures. For example, information on transport distances as well as the transport emissions are currently missing [4]. Furthermore, the recorded attributes cannot be evaluated automatically due to the large number of different input formats [4].

There is potential for improvement through standardization, which could significantly enhance data usability. Standardizations that are uniformly implemented in all systems enable automated analysis of big data and increase user-friendliness. Subcontractors would no longer need to familiarize themselves with new structures and processes for each system. Instead, they could apply familiar patterns.

The categorizations and the insights gained from analysis of the systems enabled a structured and comparable recording of all attributes contained within the logistics portals. In general, the project partners record the same attributes. However, there are significant differences in the level of detail. While some systems capture detailed information about the building under construction (e.g. number of floors, number of employees, construction volume), others limit their data to a few basic attributes. There are also differences in the level of detail in the transport information. For example, some systems only record the material type and material quantity. In other systems, additional information such as the exact packing size is recorded. It should be noted that the more detailed attributes are often incompletely.

3.2. Analysis of the data types

The analysis examined which entries must be provided by users (mandatory fields) and which are available as optional additions. Furthermore, a distinction was made between data types such as numerical values, free-text input and enumeration. This analysis forms a basis for the definition of data standards and enables a well-founded understanding of the logistics systems.

Figure 3 illustrates the distribution of mandatory and optional fields across the attribute categories. It becomes evident that delivery time specifications are consistently defined as mandatory in the systems. This can be attributed to the fact that scheduling of deliveries is a key control parameter in construction logistics processes and is important for avoiding. The vehicle type is mandatory in all systems. In contrast, the license plate number is generally treated as optional, as this information is often not available at the time of booking. In the resources category, information on access points or loading zones is generally considered mandatory, as it is essential for logistic. The recording of additional resources such as loading aids (e.g. lifts) is often treated as optional. In the materials category, attributes such as material type and quantity are required. The dimensions of the load (e.g. length, width, height) are often requested on an optional basis. The large number of optional fields for suppliers is due to the fact that subcontractors cannot specify a supplier in some systems. In addition, the specific supplier is usually not yet known at the time of booking, so that usually only the name can be entered.

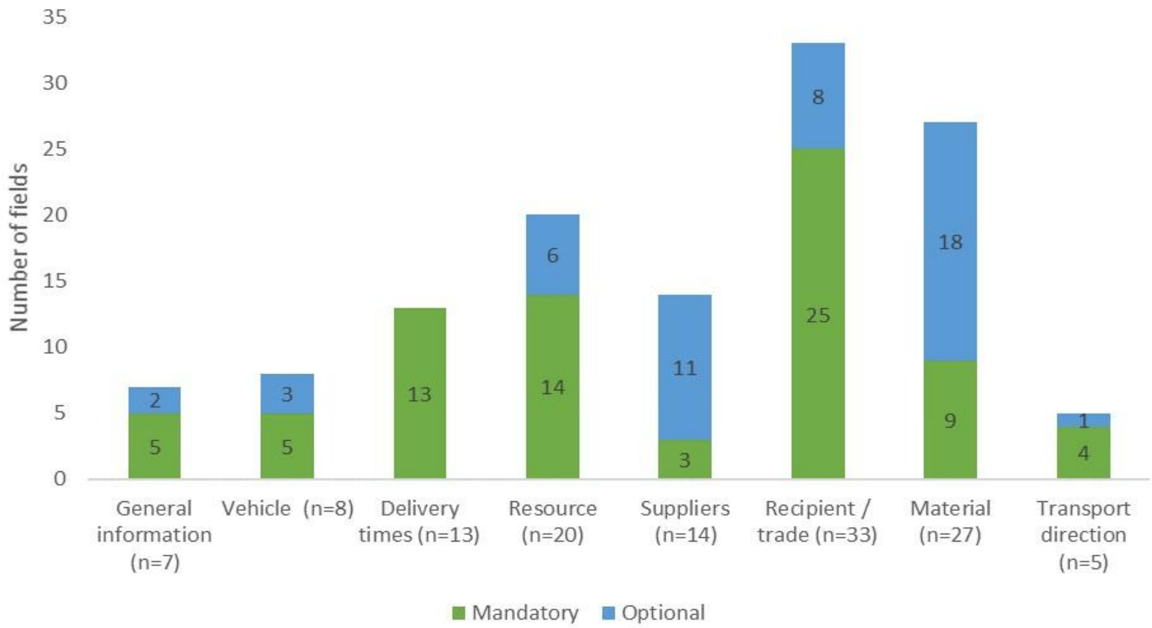

Figure 3: Proportion of mandatory entries for transport information (n = 127)

Many data fields are designed as free-text fields (approx. 43%). While this input method allows high flexibility, it also increases the risk of inconsistent, ambiguous, or erroneous entries. As a result, this often leads to unusable data. Approximately 28% of the attributes are captured via predefined selection fields (dropdown menus). This structured input format enables the selection from fixed options, improves data evaluability and reduce errors. Time and numerical values (17% and 12%) are applied to attributes that require a clear and machine-readable structure. Delivery times are typically entered as timestamps with time zone, while the weight and quantity of the load are accepted as numerical values.

3.3. Assessment of data quality

Puslat et al. (2024) already examined the data quality from logistics portals of two construction logistics companies. In the following, the data quality of the five datasets is assessed based on the approach of Puslat et al. (2024) [4]. The assessment of the data quality was based on the 15 IQ dimensions of information quality according to Rohweder (2021) [5]. Due to privacy and the heterogeneity of the systems, the datasets were evaluated individually on the basis of several analysis, e.g. the distribution of vehicle types, delivery times, and material diversity.

Mandatory fields are approx. 90% complete. However, gaps were found in mandatory attributes, raising concerns about data completeness. Interviews clarified that the development of systems, such as the integration of new attributes, is a common cause of a lack of completeness. The new attributes are missing from older datasets, as they were not included at the time of initial data collection.

The analysis of the data reveals significant deficits, particularly regarding input formats. Free-text entries proved to be highly error-prone and resulted in a large variance in the labelling of identical content, such as material types. This complicates automated evaluations.

Although a vehicle type is recorded for each transport, the quality of this information is often questionable. Conspicuous are cases where the vehicle type did not match the material or quantity, as example a 'concrete mixer' recorded for a timber delivery. Ambiguities like this undermine the reliability of the data and make it difficult to determine which information is correct and which information is incorrect: the vehicle type, the material type, or both.

Enumeration do not always reflect real-world conditions, so there is a risk that formally correct but factually inaccurate entries are made. Some systems allow users to select options such as 'no information' or 'unknown'. For the 'vehicle type' attribute this option was used in more than 50% of cases in certain systems. This reduces the informative value and limits the usability of the data for analysis. As mentioned in Puslat et al. (2024) [4] these issues highlight the need for standardization and crosssystem learning to improve the quality of information collection.

As a result, the assessment is congruent with the results from Puslat et al. (2024). Of the 15 dimensions of Rohweder (2021), the five datasets fulfil six dimensions. These include in particular the purposedependent dimensions: timeliness, value-added, appropriate amount of data and relevancy. Furthermore, the data is objective and interpretable. The inherent (reputation, free of error, believability) and presentation-related dimensions (understandability, concise representation, consistent representation) and completeness are not fulfilled. As in Puslat et al. (2024) the category 'systemsupported' was excluded from the analysis. In Table 1 are the results of the assessment.

Table 1: Assessment of data quality and information content.

Category	**Presentation-related**	**Inherent**	**Purposedependent**	***System-supported***
Dimension	Understandability	Reputation	Timeliness	*Accessibility*
	Concise representation	Free of error	Value-added	*Ease of manipulation*
orange = dimension not fulfilled	Consistent representation	Objectivity	Completeness	
blue = dimension fulfilled	Interpretability	Believability	Appropriate amount of data Relevancy	
grey = dimension not considered				

4. Description of data collection processes

Interviews with the project partners provided detailed insights into the logistics portals and the data collection processes. The focus is on the structure and functionality of the portals. This included the layout of input screen, the sequence of requested information and the workflow, such as the further processing and controlling of the information. Figure 4 illustrates the entire process of data collection based on the analysis of the five systems, which is explained in the following.

Construction logistics companies coordinate the material flows and manage the logistics processes on large construction sites. Project-specific parameters are initially stored in the logistics portal. This includes, for example, the construction site location, gross floor area, access routes and specifications for the trades (e.g. working hours, availability of resources, loading zones). Dynamic adjustments are often not provided in the systems. All trades are informed that the transport registration system must be used. The subcontractors enter their material requirements in the portals. Site logistics manager then check the transport registrations. The subcontractors are requested to make corrections if the entry is incorrect or incomplete. If the transport request cannot be implemented logistically (e.g. no time capacities), the booking is rejected. If the transport request is approved, the material delivery can be carried out. This concludes the collection process and there are usually no further checks or corrections to the information.

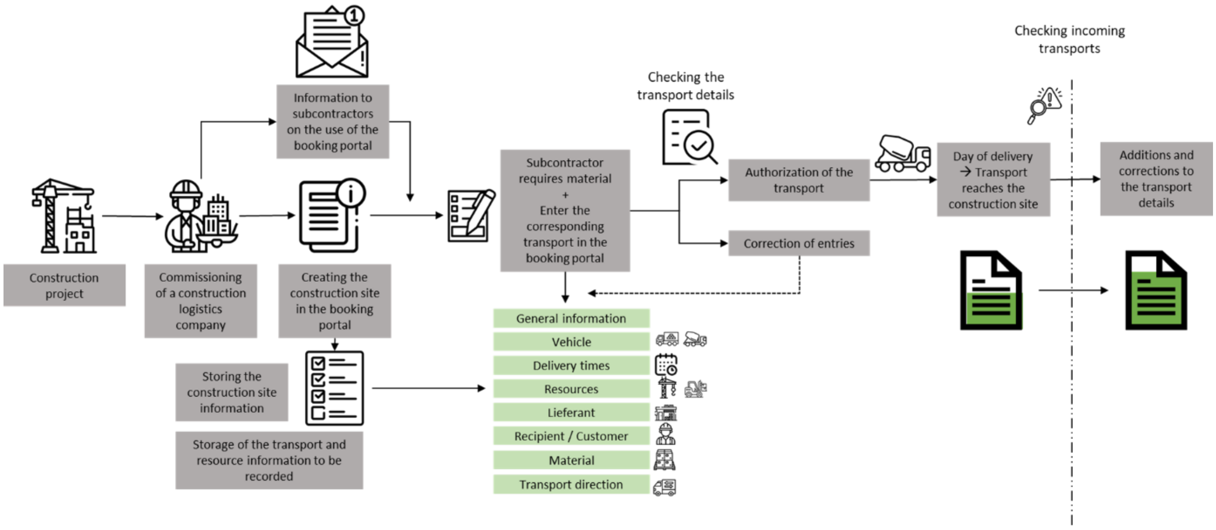

Figure 4: The data collection processes

Additionally, demo construction projects were used to simulate the data collection processes. It became clear that the systems offer high flexibility in data collection, which facilitates a detailed description of the transports. Nevertheless, sources of error were identified that can lead to implausible or incorrect data. These errors occur at the beginning of the process and continue into the later stages of data collection:

1. Incorrect or incomplete assignment of subcontractors to a trade.
2. Individual and inconsistent creation of construction projects.
3. Input errors during transport booking by subcontractors (e.g. incorrect vehicle or material).
4. Failure to book relevant resources, such as loading zones or cranes.
5. No verification of information when the transports arrive at the construction project, so that possible discrepancies go unnoticed.

These deficits highlight the need to improve data collection processes and integrate control mechanisms in order to improve data quality and increase usability.

5. Identification of data requirements

The analysis have revealed specific data requirements and potential for standardization. In order to improve the data quality and usability of the data, further requirements and potentials were identified on the basis of existing processes and datasets:

1. *Classification of construction projects*: A systematic classification of construction projects is essential for advanced analysis such as modelling material flows. A basic distinction between residential and non-residential buildings, along with a more detailed categorization (e.g. multifamily housing, hotels, commercial properties), should be standardized within the systems. Additional characteristics such as number of floors, gross floor area, or gross volume offer further analytical potential. The development of suitable building categories is a challenge, but can be addressed through collaboration with industry partners.
2. *Assignment to trades*: Each delivery should be assigned to a trade in the transport registration. This is a challenge due to the overlapping of trade activities. Furthermore, there is currently no complete definition of all building trades. A possible solution is a standardized input field to specify the target trade. This could be extended by linking the selection of trade with the delivered material, so that several materials in one delivery can be assigned to different trades. This would enable a clear assignment of each material.
3. *Standardization of transport-relevant attributes*: The definition of uniform vehicle types can help to harmonize the large number of entries across the systems, enabling automated analysis and cross-system findings. The recording of material types can also be standardized. The high error rate in free-text entry of the material can be reduced by implementing a two-step input process: First the selection of a general material category, followed by a free-text field to specify the exact material.
4. *Additional data requirements*: Currently, environmentally relevant information is not collected in the systems. The data collection should include parameters such as distance travelled and the vehicle's emission class for environmental assessment and logistic optimization. As an initial step, the origin of the transport can be recorded. Additionally, indicating the emission class of the delivery vehicle would enhance the environmental relevance of the data.
5. *Consistent input formats*: Imprecise input options such, as ‘No information’ or ‘unknown’ should be avoided. Instead, alternative entries should be allowed to achieve a higher data depth. In addition, the standardization of input formats in all systems is essential in order to ensure comparability and the interpretability of the data.
6. *Linking of attributes*: The system-based linking of inputs offers potential for improving data. For example, the selection of a crane could be linked to a specific access route or loading zone. Similarly, the choice of vehicle type should be linked to the material type or quantity to avoid implausible entries (e.g. overloading, incorrect vehicle type).
7. *Standardization of input processes:* A consistent design of the input mask in terms of sequence and structure of the requested information promotes user-friendliness and minimizes input errors. Such process harmonization makes it easier for subcontractors and suppliers to navigate in the systems and improves data quality.
8. *Control and tracking of transports*: The quality of the datasets is largely dependent on the controlling of the information. During the initial review by the logistics manager, any implausible data should be identified and corrected. In addition, the transport information should be checked when the vehicles arrive at the construction site. Unregistered transports must also be recorded in the system to ensure complete datasets.

In the STArLOG research project, data standards are being developed on the basis of these findings. These will then be validated on real construction projects in collaboration with the project partners. The aim is to improve the data quality and usability and to ensure that the data collection effort for users remains manageable. The results are analysed, adapted if necessary and finally documented in a DIN standard.

6. Conclusion

Standardizing construction logistics process data offers significant potential to enhance data quality and usability. Data requirements were identified in the existing information request and in the data collection processes of the project partner. Standardization is a promising approach to improving data quality, for example by standardising building, vehicle, material and trade categories. Overarching categories for these attributes should be defined and implemented in the systems. Therefore further research is needed to develop the necessary categorisations (e.g. for building categories, vehicle types and material types). The data collection processes also offer opportunities by standardizing input workflows and implementing control mechanisms. These measures can increase the information density and speed up the data entry process. In the future, this will lead to a higher data quality and reliable datasets, enabling more in-depth analysis.

Furthermore, for the first time, a DIN standard will be developed, enabling the improvement of process data and making it usable in practice and research. The DIN standard will document the identified and tested data standards and make them available to the construction industry.

References

[1] N. Brusselaers, H. Huang, C. Macharis, and K. Mommens, “A GPS-based approach to measure the environmental impact of construction-related HGV traffic on city level,” Environmental Impact Assessment Review, vol. 98, p. 106955, Jan. 2023, doi: 10.1016/j.eiar.2022.106955.

[2] N. Brusselaers et al., “Economic, social and environmental impact assessment for off-site construction logistics: the data availability issue,” IOP Conference Series: Earth and Environmental Science, vol. 588, no. 3, p. 032030, Nov. 2020, doi: 10.1088/17551315/588/3/032030.

[3] S. Puslat, C. Mönicks. “STArLOG – Datenstandards für die Baustellenlogistik.” GUT.Uni-Wuppertal.de. https://www.gut.uni-wuppertal.de/de/forschung/starlog/ (accessed Apr. 16, 2025)

[4] S. Puslat, F. Groß, and B. Leerkamp, “Challenges and Opportunities of Digitalization In Construction Logistics,” Proceedings of the Creative Construction Conference 2024, p. null-null, 2024, doi: 10.3311/ccc2024-005.

[5] J. P. Rohweder, G. Kasten, D. Malzahn, A. Piro, and J. Schmid, “Informationsqualität – Definitionen, Dimensionen und Begriffe,” Daten- und Informationsqualität, pp. 23–43, Nov. 2020, doi: 10.1007/978-3-658-30991-6_2.