Utilizing DSM and SysML for Modeling Data Flows in Complex Networks – A Case Study on Autonomous Public Transportation

Christopher Langner, Yevgeni Paliyenko, Daniel Roth, Matthias Kreimeyer

University of Stuttgart, Germany

Abstract: Advancing communication functionalities of smart products result in a growing importance of data within modern systems. This requires a data-centric understanding of the system's communication dynamics to support the interdisciplinary design of such networks. Especially when dealing with diverse, complex stakeholder networks, the modeling approaches of DSM/Graph and SysML are proven concepts. This work investigates the suitability of using those models for the described data-related analysis of complex systems by applying them to the case study of autonomous public transportation. After extensive data gathering and model building, numerous industry expert interviews are conducted. The general applicability of those modeling approaches for the given task is confirmed. The emphasized strengths of the models are the graph's capability of presenting a fast systems overview, the ability of a comprehensive representation of all dependencies in one model view through DSM, and the possibility of combining functional and communication-related dependencies in the SysML model.

Keywords: Dependency and Structure Modeling (DSM), Systems Modeling Language (SysML), Model-Based Systems Engineering (MBSE), Data Communication, Autonomous Public Transportation

1 Introduction and problem clarification

The evolution of information and communication technology (ICT) has facilitated unprecedented advancements in global communication by intensifying data exchange between stakeholders and systems. Resulting from these advancements, the availability and significance of data continue to expand, and so does the importance of approaches for enabling data communication and data management. Consequently, effective methodologies for describing, analyzing, and managing interactions within data exchange processes are needed. (Parra-Moyano et al., 2020)

Associated with this observation, product-service systems (PSS), cyber-physical systems (CPS), and data-driven design (DDD) are intensively discussed in the engineering disciplines. PSS describe the integration of physical products and non-physical services to generate an offering for the customer. CPS combine intelligent software components with physical elements, enabling real-time data acquisition and analysis. This data can then be used for numerous purposes, for example, for extracting customer requirements based on real field data instead of previously proven approaches. Resulting from the amount of available data along the whole product lifecycle, a new paradigm for design processes based on the usage of data is expected, which is DDD. These developments underscore the necessity for enterprises to intensify their ICT capabilities to enhance their value propositions in an increasingly digitalized economic landscape. (Machchhar et al., 2022)

However, the design and implementation of communication systems that facilitate data exchange and manage communication pathways in distributed, multi-stakeholder networks present significant challenges due to the inherent complexity of these networks. Such complexity arises from the heterogeneity of stakeholders, each possessing distinct interests and data requisites. This necessitates a comprehensive understanding of data flow dynamics and communication processes within the system (Rak et al., 2020). Accurate modeling of information flow within an organization or system is essential for enabling in-depth analysis, which, in turn, facilitates structural optimization to ensure efficient and seamless data transmission (Eppinger & Browning, 2012). The modeling concepts of Design Structure Matrix (DSM) and Systems Modeling Language (SysML) are two promising approaches for modeling data flows in a complex stakeholder network (Elakramine et al., 2022; Delsing et al., 2024; Pannunzio et al., 2023; Brovar et al., 2022).

DSM methods provide a systematic approach for analyzing interactions within complex systems. These methods utilize an n x n matrix representation, where n denotes the number of system elements, and the matrix entries capture their interdependencies. DSM approaches can be categorized into three primary types: product architecture DSM, which maps relationships between system components; process architecture DSM, which delineates dependencies between process steps; and organization architecture DSM, which characterizes interactions among individuals or stakeholders. Each DSM type is represented as a standalone matrix, while Multi-Domain Matrices (MDMs) integrate multiple DSMs to facilitate a comprehensive systems analysis across domains. Mathematical techniques, such as clustering and sequencing, enable the optimization of system structures, product architectures, or team compositions. DSM has proven to be very useful when analyzing large, complex stakeholder networks regarding their communication. (Eppinger & Browning, 2012)
Strongly related to DSM methods is Graph Theory, a mathematical framework for analyzing networks and the relationships between interconnected entities. In this approach, systems are represented as graphs, where nodes correspond to individual elements, and vertices denote the dependencies or interactions between them. This representation enables a structured analysis of the system by applying graph-theoretic techniques, such as path analysis, clustering, and centrality measures. (Diestel, 2024) Basically, it is just another way of visualizing the same dataset and mathematical relations as in

a DSM. Depending on the system to be analyzed, the visualization as either a DSM or a graph can be more beneficial. Especially when large systems are analyzed, graphs can provide a well-visualized overview. Therefore, it is important to consider both ways of visualization when discussing DSM applications. (Lindemann et al., 2009)

On the other hand, SysML provides an extensible set of diagrams, including block definition diagrams, internal block diagrams, and activity diagrams, which are particularly useful for representing structural, behavioral, and data flow aspects of complex systems. This modeling capacity is critical in multi-stakeholder environments, such as PSS or CPS, where different domains (e.g., mechanical, electrical, software, service) must converge into a coherent system architecture. SysML enables the definition of system boundaries, interfaces, and interdependencies among subsystems and stakeholder roles, thereby offering a unified view across disciplinary perspectives (Delsing et al., 2024). SysML plays a pivotal role in stakeholder engagement and data communication throughout the systems engineering process. Its semi-formal visual notation acts as a boundary object, allowing domain experts and non-technical stakeholders alike to collaboratively reflect on system functionality, roles, and lifecycle implications. (Elakramine et al., 2022)

In the context of modeling data flows in a complex system, the deployment of autonomous buses within public transportation is a related example. Despite significant advancements in sensing technologies and control algorithms, the implementation of such systems continues to face substantial challenges. Among these, the absence of standardized regulatory frameworks and the ongoing need for empirical field data are widely recognized. However, equally critical is the lack of mature digital communication and data management infrastructures that are capable of supporting the operational, service-oriented, and organizational demands of autonomous mobility solutions. These infrastructures form the backbone for enabling safe, reliable, and adaptive system behavior, but are currently underdeveloped or fragmented across stakeholder boundaries. (Langner et al., 2024a)

Addressing this gap requires a thorough analysis of the data-driven dependencies and interrelations among the system's stakeholders and digital interfaces. Previous work has demonstrated the general DSM- and SysML-based modeling approaches to this domain. However, these studies have not yet yielded comprehensive, large-scale models that reveal the systemic interactions at play, nor have they systematically explored the respective strengths and limitations of these two modeling paradigms in supporting interdisciplinary design tasks (Langner et al., 2025b). To address this research gap, the present work investigates the suitability of DSM/Graph and SysML techniques for modeling data-centric communication networks in complex stakeholder constellations, specifically within the case study of autonomous public transportation. The objective is to identify how these modeling approaches can support the interdisciplinary understanding, system integration, and coordination necessary for the design of smart, service-integrated transportation systems. Thus, the central research question guiding this work is: *How suitable are the modeling techniques DSM/Graph and SysML for the interdisciplinary data-centric design of complex stakeholder networks?*

2 Approach to research

This research was conducted following the approach shown in Figure 1. The key to generating significant models is a well-defined, detailed dataset that is shared as the input for both the DSM/Graph and SysML modeling approaches. As described above, the data-centric modeling of the complex stakeholder network of autonomous public transportation is used as the guiding case study of this work. The dataset is merged from different sources, comprising relevant recommendations/guidelines (12 screened) from the Association of German Transport Companies (VDV), existing literature, data models from partners of the research project on autonomous public transportation within which this research was conducted, and results from extensive data gathering workshops with industry experts. Divided into two series, a total of five industry workshops are conducted. The first workshop series' goal is to identify the relevant stakeholders in the ecosystem of the autonomous bus and capture an initial overview of the data needs of those stakeholders. In order to capture this information, the lifecycle of an autonomous bus is used as a guiding framework. The lifecycle according to Langner et al. (2024b) is presented to the industry experts, and for each phase, it is discussed who is active and which role they play. In a second step, the second workshop series is conducted to capture the required communication in detail. The identified stakeholders and systems (entities) from the first workshop series are used as the structuring framework. It is discussed which exact information is needed per entity to fulfill their role and where this information is coming from. This data is collected in the format as described in Section 3.1. For all workshops (duration of 3.5 h each), the participating companies and the fields of work of the participants are given in Figure 1. Based on the merged dataset, which is described in more detail in Section 3.1, both a DSM/Graph model and a SysML model are built separately and independently by the authors, but sharing the exact same data input. 16 industry expert interviews (1.5 h each) are conducted for both validating the dataset input and evaluating the strengths and weaknesses of both modeling approaches. The details of the interview participants are given in Figure 1. Five of the experts participated in the second workshop series as well. At first, the captured dataset is presented and checked for completeness. Secondly, both the DSM/Graph and SysML models are presented. Based on this, thirdly, the opinion on the models is discussed to identify the individual assessment of strengths and weaknesses from the point of view of the industry experts along the guiding questions presented in Section 3.4.

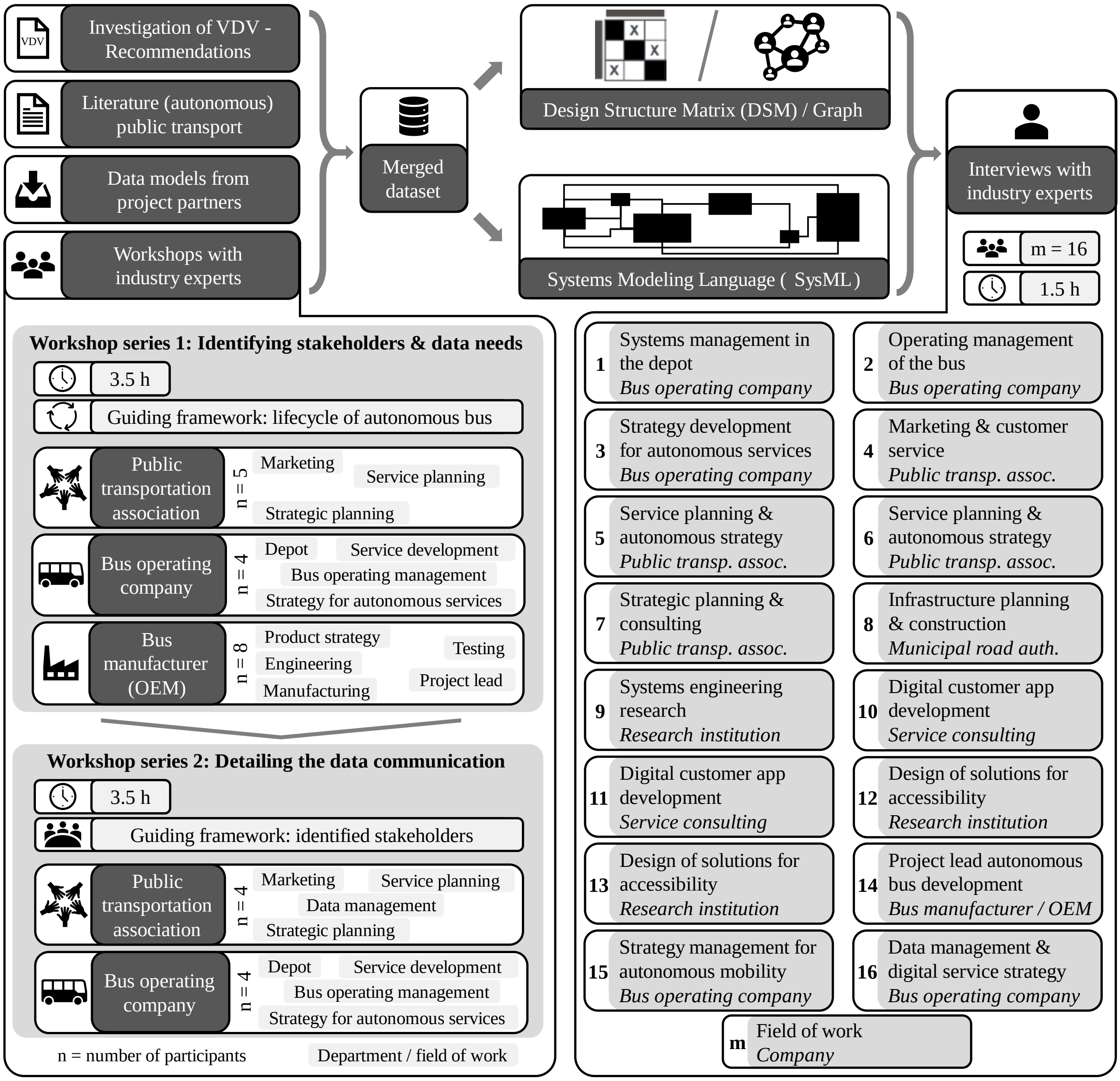

Figure 1. Research procedure

3 Case study: operating autonomous buses in public transportation

3.1 Acquired dataset

As described, a stakeholder-centric data capturing approach provides the dataset. It captures the information a stakeholder needs, which systems are involved in their work, and which information is exchanged between them. Resulting, a list of entities (both stakeholders and systems) and their directed dependencies (information flow and its direction) is defined, spanning 742 entries in total. An excerpt is shown in Table 1. Entities that have known subsystems are modeled accordingly, separating the parent system from the subsystem with a dot (e.g., Autonomous Vehicle.Vehicle systems).

Table 1. Excerpt from the acquired dataset describing the information exchange between two entities

Source/Provider	Target/Consumer	Information
…	…	…
Autonomous Vehicle.Vehicle systems	Charging management system	Requested charging power per vehicle
Autonomous Vehicle.Vehicle systems	Charging management system	Battery type
Autonomous Vehicle.Vehicle systems	Charging management system	State of Charge
Autonomous Vehicle.OnBoard Unit (OBU)	Mechanic	Vehicle data according to VDV 238
…	…	…

Table 2 gives an overview of the included entities to outline the scope of the dataset. It consists of both stakeholders and systems. To structure it, the entities are arranged by activities or superordinate grouping clusters. It is important to note that this clustering does not influence the later building of the models, it just serves as a structuring for this overview of the entities. The models use the data list from Table 1 as an input, where the grouping from Table 2 is not included. Langner et al. (2025a) provide a detailed description of the entities in the network of autonomous public transportation.

Table 2. Included entities (without subsystems) in the dataset

Planning	**Operation in road traffic**	**Passenger information**
1. Strategic/network planning	1. (Autonomous) vehicle	1. Marketing / customer service
2. Rotation planning	2. (Driver)	2. Passenger information dispatcher
3. Duty planning	3. Passenger	3. Station displays
Disposition	4. Other traffic participants	4. Server for dynamic passenger information (DFI)
1. Vehicle disposition	5. Traffic manager	
2. Personnel disposition	6. Integrated on-board information system (IBIS)	5. Individual mobile apps
Depot		6. Real-time communication and assistance platform (EKAP)
1. Electric depot management system (E-BMS)	7. Automatic passenger counting system (AFZS)	
		Public databases
2. Charging management system	**Technology provider**	1. Central station register
3. Maintenance system	1. OEM	2. Platform for consistent electronic passenger information (DELFI)
4. Depot tracking system	2. Driving software provider	
5. Workshop management	3. Sensor technology provider	3. Mobilithek
6. Mechanic	**Service providers**	**Administration and authorities**
7. Cleaning staff	1. Transfer company (bus delivery)	1. Legislature/Lawmaker
Operational control	2. Transport research	2. Vehicle registration authority
1. Technical supervision	3. Insurance company	3. Road authority
2. Incident manager	4. V2X system provider	4. Road administration
3. Control center dispatcher	5. Map data provider	5. Approval and funding authority
4. Intermodal transport control system (ITCS)	6. Navigation service provider	6. Public transport authority
	7. Inspection agency (TÜV)	7. Standarization bodies
5. Event management system (EMS)	**Ticketing/Booking**	**Infrastructure**
Other operational aspects	1. Ticketing system online	1. Traffic infrastructure
1. IoT platform	2. Ticketing system offline (vending)	2. Station
2. Central data management	3. Tariff planning	3. Charging infrastructure
3. Operational financial planning	4. On-Demand offers booking center	4. Energy producer

3.2 Data communication model using DSM/Graph

As described in Figure 1, the merged dataset characterized in Section 3.1 is used as the data input to create a DSM and Graph model. The software used to build those models is *Lattix* for DSM and *Gephi* for graph. Both tools read the described Excel sheet from Table 1. The corresponding models are shown in Figure 2. It gives a systems overview. The DSM on the left side represents the system in a matrix, showing its entities on the first row and columns. Using the original DSM reading convention according to Eppinger & Browning (2012), all inputs an entity receives are placed in the rows, all outputs are represented in the columns. The graph model on the right-hand side of Figure 2 shows the same dataset. All entities are visualized as nodes, and in- and outputs are represented by vertices connecting the nodes. Since these vertices have a direction going from provider to receiver, this is a directed graph (Diestel, 2024). If multiple information are communicated between two entities, the size of the connecting vertex grows accordingly. In the DSM, this is represented by the number that is displayed in the connecting matrix cell, as seen in Figure 3. The numeric values arise from the Excel sheet from Table 1, e.g., the shown example leads to a value of three for the communication from the vehicle systems to the charging management system. In addition, the graph can increase the size of the entities visualized as the nodes based on the amount of information that is being passed through it, meaning the sum of the number of inputs and outputs. This helps to identify the important entities in terms of data communication. Unsurprisingly, the autonomous vehicle is a key entity in the system, which is highlighted in the DSM by a dotted mark for its row and column and in the graph by an arrow. While the graph's visualization helps to easily identify the important entities in the system, a strength of the DSM is the possibility to allow for a detailed investigation of selected dependencies between entities in the same view. By reading across the row, all inputs for one entity can be identified, and on the other hand, all outputs created by this entity by reading down the column, as it is exemplarily highlighted in Figure 2. If taking the perspective of a selected entity, it can easily be identified which interfaces are needed to connect with which entities and what information must be transferred through those interfaces.

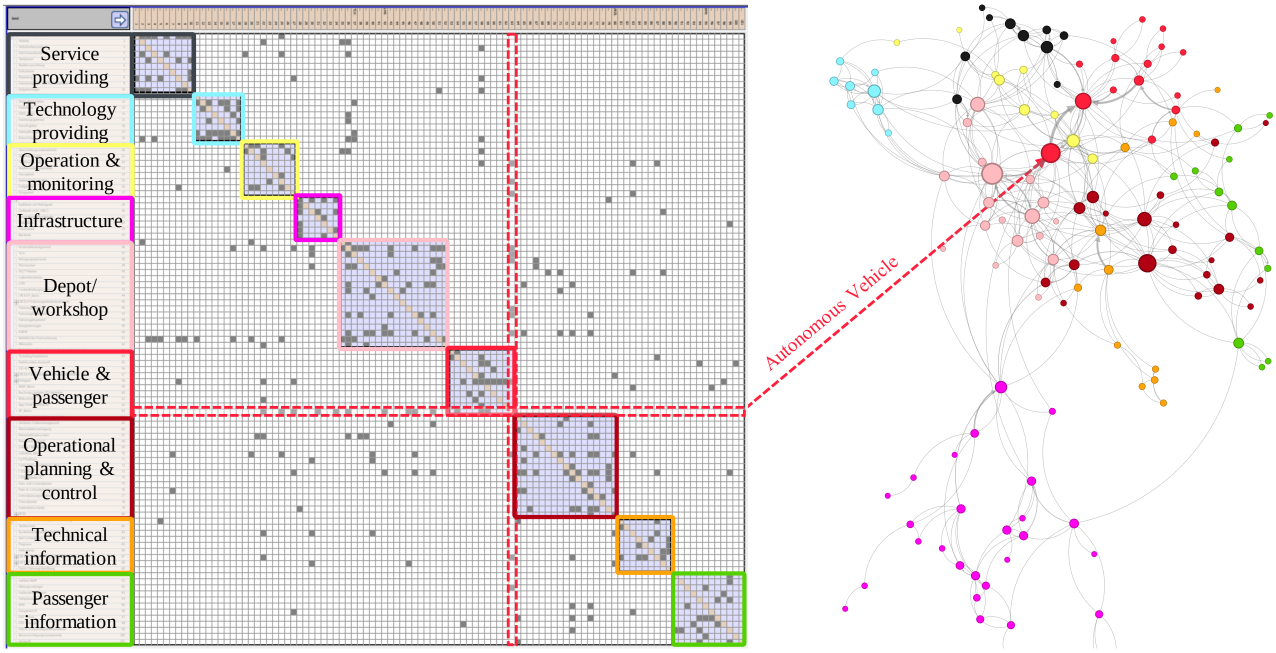

Figure 2. Systems overview with clusters visualized as DSM (left) and graph model (right)

Figure 2 furthermore shows clustering, another important analysis approach when discussing organizational structures or systems designs enabled through DSM and graph models (Lindemann et al., 2009). In this case, it helps to identify entities with closely entangled communication, for systems and organizational design, resulting in the opportunity to enhance communication by installing shared systems or interfaces between those entities. To cluster a directed graph, the *Louvain algorithm* is well-proven and finds numerous successful applications (Shirazi et al., 2019). It was first described by Blondel et al. (2008) and targets to optimize the modularity of a network by separating the network into individual clusters. Transferred to the case study, it means to ideally have groups of entities that only communicate within themselves and have no communication paths to other groups. Of course, this is not a practically reachable target, but the algorithm strives to reach the optimal clusters as closely as possible. The outcomes are presented in Figure 2, with the derived clusters being colored in the DSM and graph model. It is apparent how the algorithm works, by finding a lot of interactions within and fewer outside the clusters in the DSM, or by having the nodes closely together in the graph. At the same time, integrative entities of the overall system can be identified when finding many interactions of it outside its cluster. This applies, as highlighted before, for the autonomous vehicle, but also the workshop management and others. These integrative entities can be derived as the key players for the functioning of the system. Overall, the suggested clusters are reasonable, and a fitting umbrella term can be derived, as displayed in Figure 2. However, a detailed look is required to examine how far they can be translated into realistic recommendations for actions for the operation of autonomous buses in public transport. Compared to the initial structuring of Table 2, the planning and disposition aspects, the authorities, and the databases are dissolved and distributed over the other clusters. For each cluster, the practical applicability of the grouping of entities must be investigated. Figure 3 shows one cluster as an example. It could be labeled as the *depot-related cluster*, with the workshop management and the depot management system (E-BMS) being the core entities in this cluster.

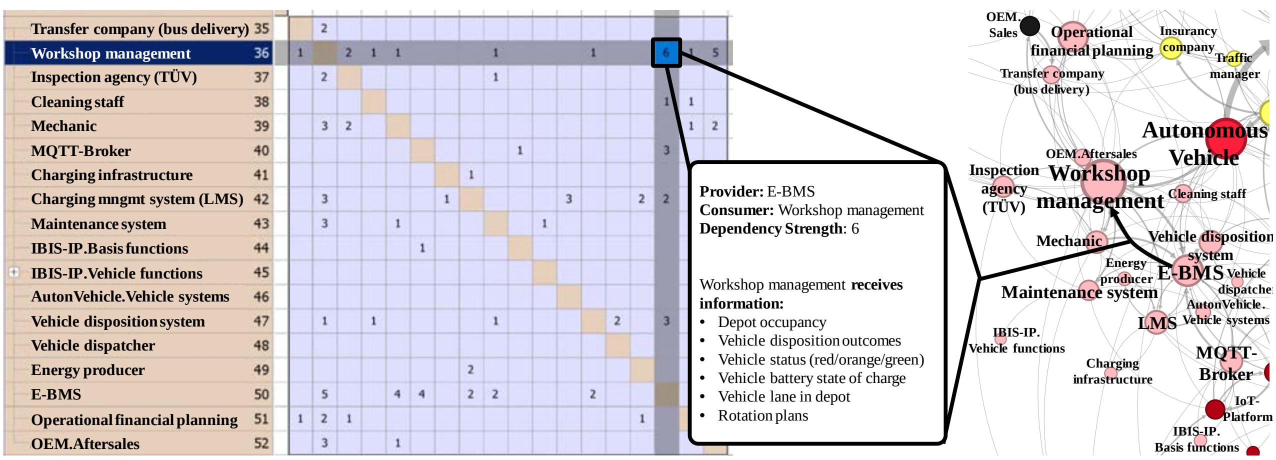

Figure 3. Detailed DSM/Graph view for one exemplary cluster with depot-related entities

Since the data-based clustering is conducted across previous system affiliations, some currently grouped system components are assigned to different clusters. Here, the vehicle-related functions of the integrated on-board information system (IBIS) are assigned to the depot/workshop cluster of Figure 3. This poses an interesting jumping-off point for a systems improvement analysis deep-dive, for each system which components are distributed over several clusters, the system's structure could be re-organized. In addition, Figure 3 shows the possibility of providing more detailed information on a selected interface between two entities. The *information* column from Table 1 is read by the software and included in the *dependency properties*. Figure 3 shows the example of *Lattix*, a similar view showing the same data can be derived from *Gephi* as well by clicking on a selected vertex.

Overall, the DSM/Graph model allows both a systems (over-)view and detailed views on selected areas of the system in one model view by just zooming in. It can fully represent the comprehensive dataset and allows for several analysis approaches. Further steps should include a detailed review of the suggested clusters and derivation of concrete, realistic recommended actions for organizational or systems design.

3.3 Data communication model using SysML

In the subsequent section, the same aggregated data set is utilized to construct a system model using SysML, specifically through an Internal Block Diagram that acts as a logical architecture model. This diagram was created using the *Capella* tool, developed within *Arcadia*, which employs a modeling logic that integrates established principles from UML, SysML, and the NATO Architecture Framework (NAF). *Capella* thus ensures a high degree of semantic interoperability with other Model-Based Systems Engineering (MBSE) environments and has proven its utility across domains such as aerospace, automotive, and defense (Elekramine et al., 2022). The resulting SysML model represents a function- and data-centric abstraction of the multi-stakeholder autonomous public transportation network, with a specific emphasis on data flows, system functions, and organizational interfaces.

Building on the data-centric modeling methodology for smart PSS proposed by Paliyenko et al. (2024), the autonomous transportation network is systematically modeled. As an initial modeling step, stakeholders are manually organized into informal clusters based on their primary functional roles and responsibilities. This clustering serves as a foundational abstraction layer to reduce representational complexity and to highlight entities with closely coupled communication patterns. These clusters form a basis for systems and organizational design, revealing latent opportunities for shared interface development, co-location of services, or middleware integration to enhance inter-organizational communication efficiency. Each cluster is encapsulated as a logical subsystem in the block diagram, containing functions that are either executed by or allocated to the underlying stakeholder entities. The overarching blocks in the diagram represent major socio-technical domains (e.g., public operators, mobility platforms), while sub-blocks denote actor roles or system components with tightly integrated responsibilities. This allows for both horizontal (cross-organizational) and vertical (within-organization or subsystem) traceability. Figure 4 illustrates the top-level network clusters and the inter-cluster data channels that represent the functional interrelations and coordination mechanisms.

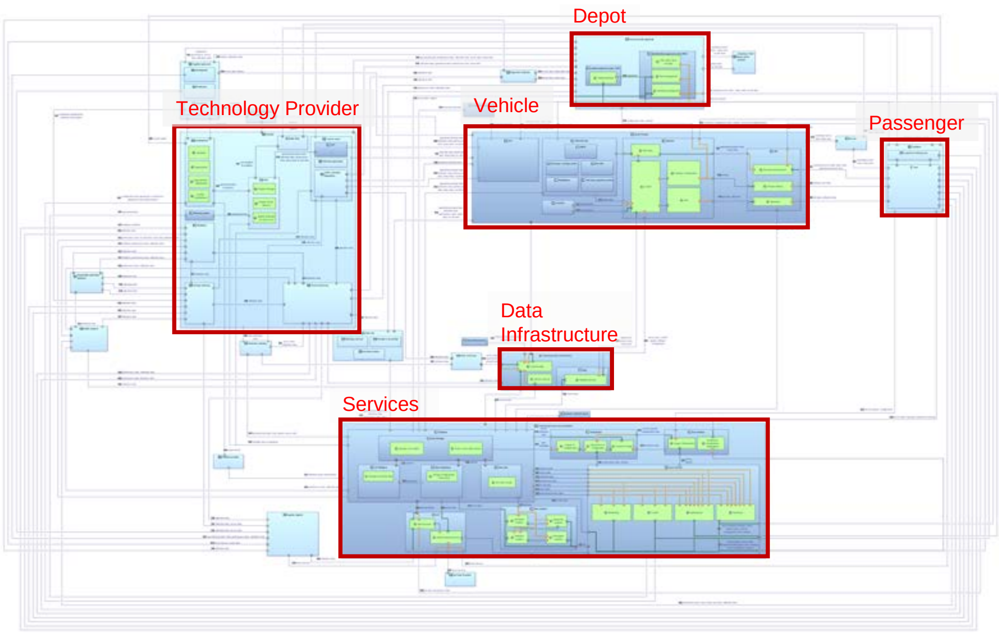

Figure 4. Systems overview with clusters visualized as SysML

The complete SysML model represents multiple domains - technical, organizational, and service-related - that are decomposed into subsystems (highlighted in dark blue), sub-actors (light blue), and logical functions (green). These visual encodings enhance model readability and enable cross-domain mapping. Data transfers between these elements are semantically typed and linked to system-level goals, thereby supporting traceability from stakeholder requirements to data infrastructure and service behavior.

The data and information flows between entities and functions are categorized using a predefined taxonomy of data types, grounded in best practices from information systems engineering (Paliyenko et al., 2024). This classification significantly reduces visual noise and supports the semantic compression of communication patterns. While the model presents a lean visual form, individual data channels can be interactively "unpacked" to explore the underlying data streams, enabling a dynamic granularity of analysis. These data channels are bidirectional or unidirectional, depending on the operational logic, and represent either utilization data, operational product data, performance data, service data, status data, or on-site data. The SysML model further allows the isolation and in-depth analysis of communication links between two functions or entities. This facilitates the identification of affected roles, subsystems, and data interactions for scenario-based analysis (e.g., failure cases, system upgrades, or integration scenarios).

An excerpt of the model centers around the autonomous bus as a smart product and central system node is displayed in Figure 5. The bus integrates core technological components, like sensors, actuators, edge computing units, and human-machine interfaces, to enable autonomous operation and real-time user interaction. Sensor data gathered onboard includes contextual (e.g., traffic and route conditions) and operational variables (e.g., passenger occupancy and vehicle status). This data is processed locally to generate control signals and operational decisions. Additionally, pre-filtered and structured data is transmitted via a GSM communication infrastructure to centralized systems, where it is further processed, stored, and fused with contextual data in data lakes or data warehouses, depending on the structure and analysis objectives.

End-users and service stakeholders interact with the system via the bus HMI, mobile applications, and web portals that connect to a central operational platform. These interfaces facilitate dynamic service configuration (e.g., routing preferences, accessibility settings), real-time information delivery (e.g., delay notifications), and integration with third-party systems. Thus, the model not only reflects technical communication flows but also incorporates service-oriented design principles.

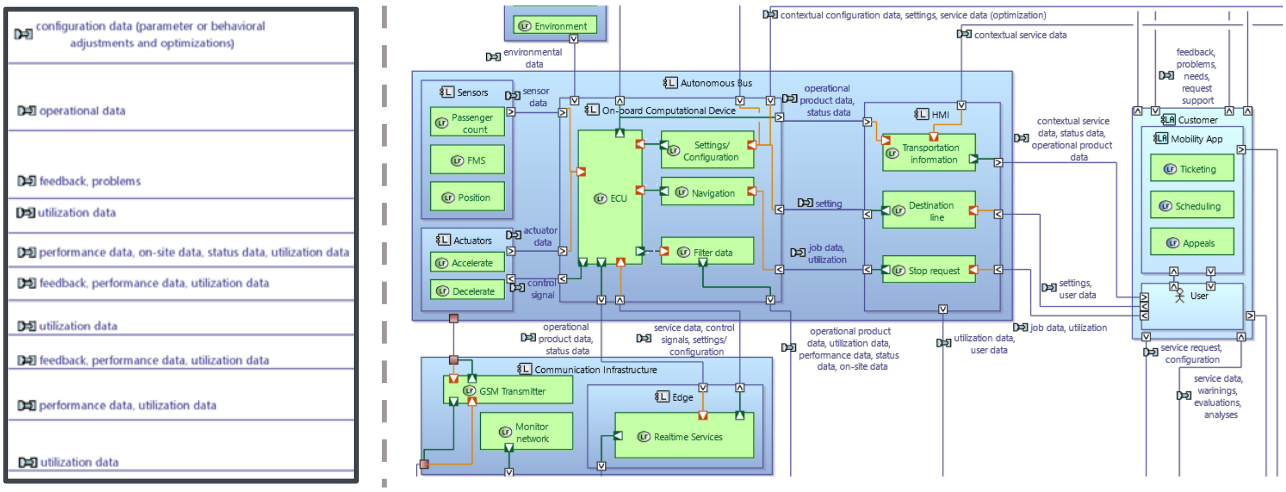

Figure 5. Excerpts illustrating vehicle-associated blocks (right) and categorized data channels (left).

The SysML-based modeling approach demonstrates a high level of expressiveness in capturing the structural, functional, and informational architecture of the complex, multi-stakeholder autonomous transportation network. By integrating clustering, logical function allocation, data channel classification, and multi-layered visualization, the model supports both top-down system design and bottom-up analysis, accommodating different stakeholder perspectives.

3.4 Evaluation of the DSM/Graph and SysML models by industry experts

As described in Chapter 2, a total of 16 interviews with industry experts are conducted to evaluate the presented models. The professional backgrounds of the participants are given in Figure 1. The participants span all important stakeholders in the ecosystem of the case study and are therefore a suitably diverse group to give valuable insights for this work.

The interviews are conducted as follows: at first, the target setting of the model build is presented, and the methodology for building the models is introduced. Following, the acquired dataset with a focus on the included entities is discussed regarding its completeness. Consequently, the basic concepts of DSM, Graph, and SysML are explained before showing the model for the case study of autonomous public transportation in detail. To ensure a comparable evaluation, eight guiding questions, as shown in Figure 6, are discussed with the participants, leading to a numeric evaluation by the participants on a scale from one (not at all) to five (completely). Additionally, the participants are asked to give any other feedback or thoughts in an open discussion at the end.

Question	Ø DSM	Ø Graph	Ø SysML
How understandable was the modeling methodology overall?	*4.75*	*4.75*	*4.88*
To what extent does the model represent all relevant aspects of the system?	*5.00*	*5.00*	*5.00*
To what extent did the modeling approach contribute to effectively integrating knowledge and requirements from different domains (e.g., mechanics, electronics, software, service)?	*4.44*	*4.31*	*4.50*
How useful was the modeling approach overall for modeling the system?	*4.50*	*4.19*	*4.36*
To what extent did the support help to reduce the complexity of the system to be developed?	*3.75*	*4.19*	*4.00*
How well does the model support the understanding of dependencies between system elements?	*4.34*	*4.09*	*4.04*
To what extent was the model flexible enough to represent different variants of the system?	*4.06*	*4.25*	*4.32*
Would you use the methodological support in your daily practice?	*4.13*	*4.31*	*4.43*

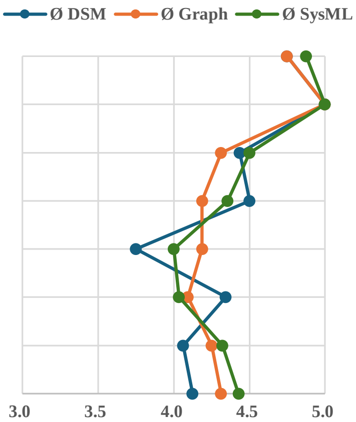

Figure 6. Evaluation of the DSM/Graph and SysML models by industry experts

Figure 6 shows the average evaluation separately for each question and each modeling approach. To visualize the comparison of all three models, a corresponding graphic is given on the right-hand side of the figure. Overall, all three modeling approaches are assessed between four and five, and the differences between the models within one question are small, emphasizing their general suitability for a data-centric modeling of an interdisciplinary, complex stakeholder network. Nevertheless, specific strengths and weaknesses of all approaches are visible and can be derived. In the following, the evaluation is discussed with enrichment of the expert's feedback given beyond the posed guiding questions.

Insights concerning DSM

The application of the DSM is predominantly perceived as intuitive and easy to understand by the interview participants. It is particularly appreciated for providing a quick and high-level overview of the system. The matrix structure allows for a clear representation of incoming and outgoing relationships between entities, especially concerning data communication or interaction paths. However, a key limitation noted is the loss of clarity as the number of entries increases, since users tend to "get lost" in the matrix when the complexity grows. While this reduces its usefulness for detailed or highly networked systems, DSM remains a valuable tool for early-stage system analysis and coarse structural understanding. Additionally, DSM is pointed out as a structured framework for the required data capturing, since it ensures all possible interfaces between two entities are investigated when discussing the network.

Insights concerning Graph Theory

Graph-based modeling is also described as intuitive and well-suited for providing a comprehensive system-level overview. The visual representation of nodes and vertices enables users to quickly grasp complex interrelations, offering rapid insight into the structure of a system. Nonetheless, participants point out that the method has limited capacity to represent detailed structures, especially when the graph becomes cluttered and difficult to interpret due to a high density of vertices. Additionally, the need to use filters or other navigational steps to dive deeper into the model is regarded as a weakness, especially in comparison to the DSM, where, in one view, just by adjusting the zoom, a system's overview and a detailed view on selected entities are possible. Overall, graph-based models are seen as particularly useful during exploration phases or when communicating with non-technical stakeholders.

Insights concerning SysML

The interviews reveal that SysML is a highly capable tool for modeling complex systems, particularly when it comes to the explicit and formal representation of service functions, logical relationships, and involved entities. Its engineering-style visual syntax, resembling electrical schematics, is seen as a benefit, as it supports understanding among technically trained users. On the other hand, this is also pointed out as a disadvantage by users without a deep engineering background. Interviewees also emphasize SysML's ability to support user-centered modeling and to enable targeted analyses via different viewpoints. However, two main challenges are identified: the manual effort required to meaningfully model logical structures and functions, and the steep learning curve due to the necessity of understanding the underlying modeling logic. Despite these challenges, SysML is considered particularly suitable for the development and analysis of multi-layered systems due to its depth, scalability, and formal rigor.

Overall insights

The interviews show that the data collection has led to a complete depiction of all relevant players and system components. The modeling methods are also able to capture the entirety of the system landscape and depict it in such a way that complexity is reduced and understanding is promoted. Concerning the analysis or the perspective of observation, it is predominantly stated that a view focusing on oneself is the most relevant. The participants think that the greatest interest lies in examining their own relationships with other entities, the incoming and outgoing data channels, and the data streams they contain. Fewer participants find that the holistic representation of all entities, connections, and data communications is most important to them. Nevertheless, most interviewees recognize the relevance of this perspective, especially for cross-stakeholder communication and interaction during the development and operation of such a system. At the same time, it is noted that the strict clustering of actors according to organizational units plays a subordinate role because the organizational units are often predetermined. Rather, a clustering according to related and associated functional units is desired, which are characterized by a proximity of functions, tasks, responsibilities, data requirements, or availability. Unsurprisingly, two entities are named as the two most important objects of consideration, as well as their connections and data requirements: the autonomous bus and the passenger as the customer. Numerous interviewees wish for an analysis showing a comparison of the current status with bus drivers and manually driven buses and autonomous, driverless vehicles, which is not part of the showcased models yet.

4 Discussion and outlook

In this paper, the suitability of DSM, Graph, and SysML models for an interdisciplinary data-centric design of complex stakeholder networks is investigated. The outcomes emphasize the general applicability of those modeling approaches for the task of modeling data-related links of entities in a complex, multi-domain stakeholder network. Overall, the graph model is highlighted for its capability of allowing a fast, well-visualized systems overview without the need for specific previous knowledge. The DSM's structured approach supports a clear representation of all dependencies in the system at one viewing level and further helps as a guiding framework for data gathering efforts in collaborative formats. SysML's integrative representation of functional and communication-related dependencies in the system is pointed out, as well as the intuitive understandability for users with engineering backgrounds.

The research is conducted along the case study of autonomous public transportation, which gives real-world application and insights on the one hand, but limits the outcomes on the other hand. All participants of the conducted industry workshops and expert interviews are working in related fields to public transportation. To give the gained insights general validity across multiple industry domains, further case studies must be conducted. In addition, the number of 16 consulted industry experts for the interviews is reasonable to give valuable insights. At the same time, however, more interview partners should be presented with the model and asked for their feedback. Especially, an in-depth evaluation by participants familiar with the modeling approaches in comparison with persons not familiar with them would pose an interesting further research path.

A general limitation of the evaluation outcomes is the models being built by the authors. While the data gathering workshops are conducted along the modeling approach's methodology, and explained to the interviewees in detail, the model building is still done by the authors, who are experts in this field. If new users are faced with the modeling task in its wholeness, outcomes may vary. Nevertheless, the gained insights are of high value, since it is frequently observed that experts in the modeling approach build the models while the data gathering and outcomes discussion take place with a broad group of domain experts, just as it is done in this work.

The interviewees stated that implementing the design models in sophisticated software, which allows for extensive interactions, such as increasing or reducing information depth, zooming in and out in the models, search and analysis functions/fields, is crucial for ease of use. In addition, such software would reduce the required skills from the user, thus increasing the potential application in real development and operations scenarios. While for each model build, a separate software tool was used, it would be an interesting further research direction to look into the interoperability of the models. By using the same dataset and generating all three models from it, this work has shown that it is generally possible to build DSM, Graph, and SysML models with the same data basis. However, this dataset does not make use of the full range of functionality of the models, resulting in questions arising, such as: *to what extent can DSM, Graph, and SysML models be translated into one another? Which limitations occur from the basic, mathematical description? Is there a shared data model framework that supports the simultaneous translation of the models? For which use cases is which model more suitable?* Especially the last question could be answered using this work, enriched with numerous other case studies in a similar investigation style.

In conclusion, further needed actions are: further detailing of the models and implementation of deep-dive analysis of the system, as well as validating the analysis outcomes with industry experts; conducting more interviews within the case study, but especially extending the research to other industry branches, and investigation of the interoperability and translation of the three models into one another.

Acknowledgments

The authors would like to thank Frank Waldman and Code Clinic LLC for kindly providing a license of the *Lattix* software at a reduced price for academia.

This research was funded by the Federal Ministry of Transport of Germany (BMV) on the basis of a decision of the German Bundestag with ~ 12.7 million euros as part of the project *MINGA* in the framework of the funding guideline “autonomous and connected driving in public transports”.

References

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment P10008. http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

Brovar, Y., Menshenin, Y., Knoll, D., Fortin, C., 2022. Modelling of Engineering and Manufacturing Data Structures Through a DSM-Based Approach, in: Canciglieri Junior, O., Noël, F., Rivest, L., Bouras, A. (Eds.), Product Lifecycle management. Springer, Cham, 96–109. https://doi.org/10.1007/978-3-030-94399-8_8.

Diestel, R., 2024. Graph Theory (6th ed.). Springer.

Delsing, J., Kulcsár, G., Haugen, O., 2024. SysML modeling of service-oriented system-of-systems. Innovations in Systems and Software Engineering 20, 269–285. https://doi.org/10.1007/s11334-022-00455-5.

Elakramine, F., Jaradat, R., Ullah Ibne Hossain, N., Banghart, M., Kerr, C., El Amrani, S., 2022. Applying Systems Modeling Language in an Aviation Maintenance System. IEEE Transactions on Engineering Management 69 (9), 4006-4018. https://doi.org/10.1109/TEM.2021.3089438.

Eppinger, S.D., Browning, T.R., 2012. Design Structure Matrix Methods and Applications. The MIT Press. http://dx.doi.org/10.7551/mitpress/8896.001.0001.

Lindemann, U., Maurer, M., Braun, T., 2009. Structural Complexity Management. Springer-Verlag, Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87889-6.

Langner, C., Rehberg, M., Tüzün, G.-J., Roth, D., Kreimeyer, M., 2024a. Challenges of the real operation of autonomous buses in public transport along the lifecycle. DS 133: Proceedings of the 35th Symposium Design for X (DFX2024), 5-14. https://doi.org/10.35199/dfx2024.01.

Langner, C., Rehberg, M., Roth, D., Kreimeyer, M., 2024b. A Lifecycle Model for Autonomous Buses in Public Transport. DS 134: Proceedings of the 26th International DSM Conference (DSM 2024), 127-136. https://doi.org/10.35199/dsm2024.14.

Langner, C., Paliyenko, Y., Yigiter, O., Roth, D., Kreimeyer, M., 2025a. Identification and classification of the stakeholders in the ecosystem of autonomous buses in public transport. Stuttgarter Symposium für Produktentwicklung 2025 (in press).

Langner, C., Paliyenko, Y., Roth, D., Kreimeyer, M., 2025b. Modeling data flows in a complex stakeholder network - a case study on autonomous public transportation. Proceedings of the Design Society, Volume 5: ICED25 (in press). https://doi.org/10.1017/pds... .

Machchhar, R.J., Toller, C.N.K., Bertoni, A., Bertoni, M., 2022. Data-driven value creation in Smart Product-Service System design: State-of-the-art and research directions. Computers in Industry, Vol. 137, 1-21. https://doi.org/10.1016/j.compind.2022.103606.

Paliyenko, Y., Mueller, S.R., Roth, D., Kreimeyer, M., 2024. Data-Centric Architecture Model for the Development of Smart PSS. DS 134: Proceedings of the 26th International DSM Conference (DSM 2024), 1-10. https://doi.org/10.35199/dsm2024.01.

Pannunzio, V., Friday, L., Kipouros, T., Clarkson, J.P., Brayne, C., 2023. Dependency and Structure Modelling for Stakeholder Management: An Example in Landscape Regeneration. DS 126: Proceedings of the 25th International DSM Conference (DSM 2023), 133-140. https://doi.org/10.35199/dsm2023.15.

Parra-Moyano, J., Schmedders, K., Pentland, A.S., 2020. What Managers Need to Know About Data Exchanges. MIT Sloan Management Review, Vol. 61, No. 4. http://dx.doi.org/10.2139/ssrn.3670680.

Rak, J., Hutchison, D., Tapolcai, J., Bruzgiene, R., Tornatore, M., Mas-Machuca, C., Furdek, M., Smith, P., 2020. Fundamentals of Communication Networks Resilience to Disasters and Massive Disruptions, in: Rak, J., Hutchison, D. (Eds.), Guide to Disaster-Resilient Communication Networks, Computer Communications and Networks. Springer. https://doi.org/10.1007/978-3-030-44685-7_1.

Shirazi, S., Baziyad, H., Ahmadi, N., Albadvi, A., 2019. A New Application of Louvain Algorithm for Identifying Disease Fields Using Big Data Techniques. Journal of Biostatistics and Epidemiology 2019; 5(3), 183-193. https://doi.org/10.18502/jbe.v5i3.3613.

Contact: C. Langner, University of Stuttgart, Institute for Engineering Design and Industrial Design, Pfaffenwaldring 9, 70569 Stuttgart, Germany, +49 711 68568078, christopher.langner@iktd.uni-stuttgart.de