NORDDESIGN CONFERENCE
REYKJAVIK, ICELAND, 12-14 AUGUST 2024

Spearhead Data-Driven Model-Based Systems Engineering: Interview Study on Definition, Preconditions, Challenges, Potentials, and Use Cases

Denis Tissen[1], Ingrid Wiederkehr, Ruslan Bernijazov, Christian Koldewey, Roman Dumitrescu

Heinz Nixdorf Institute, Paderborn University, Fürstenallee 11, 33102 Paderborn, Germany

Abstract: Digital technologies change how companies develop their products and services. Data-driven model-based systems engineering (DDMBSE) addresses this by integrating data analytics (DA) into MBSE. As an underexplored area, DDMBSE needs deeper observation from practical side. We conducted an interview study with 16 experts and operators from MBSE and DA. The results of this paper are a refined definition of DDMBSE including a concept called nexus model. Additionally, preconditions, challenges, potentials and use cases in context of DDMBSE are described, open the path for further research.

Keywords: Complex Systems, Data Driven Design, Traceability, Systems Engineering (SE), Data Analytics

1 Introduction

With technological advancements in the past decade, the global megatrend of digital transformation has changed the way how companies organize themselves and proceed their processes. New technologies and methodologies in the areas of artificial intelligence and data analytics (DA) have shown huge potentials to better understand products, services, and users and automate repetitive tasks (Meyer et al., 2022). This is also shown in statistical facts: By 2026, worldwide investments in digital transformation are anticipated to reach 3.4 trillion U.S. dollars (Statista, 2022). Producing companies are highly interested in using data-driven technologies and methodologies in and on their products to optimize their processes and working artifacts, e.g. simulations and system models (Mütsch et al., 2023). On the one hand this is enabled by their products and services themselves since they evolve from mechatronic to cyber-physical systems (CPS). CPS are characterized by integrating classical mechanical systems and actuators with sensors and associated information processing units connected to other systems (Tränkler and Kanoun, 2007). Predicted to be at 181 zettabytes by 2025, CPS generate huge amounts of data throughout their life cycle which can be analyzed to gain unknown insights about the system and the user (Statista, 2021). On the other hand, digitally based processes and methods, especially in the development phase, are anticipated to keep pace with the rapid digitalization. Since more and more stakeholders from different domains come together through the engineering of CPS, managing the complexity of the development and systems is challenging. Model-based systems engineering (MBSE) has risen as one of the most promising representatives. By setting a system model (simplification of the real system) as the method`s focus, engineers and development teams can understand the system and its environment better. Through the application of MBSE mainly supported by a software-based modeling tool, engineers are able to use different analysis features, e.g. trade-off analysis or impact analysis (Biggs et al., 2018). Even with these features, MBSE cannot handle the huge amount of data generated to its optimum. Research in this area is scarce and only addressed indirect by prototypical demonstrators as Tissen et al. (2023) describe. It needs further knowledge from the practical side, since first innovative use cases are realized. Therefore, this paper aims to answer the research question of *how data-driven approaches in the field of model-based systems engineering are perceived by experts and practitioners*? An exploratory interview study with 16 experts and operators from MBSE and DA was conducted to answer the research question. The results reveal the first insights of DA in MBSE from practitioners. The paper at hand is structured as follows: first, the scientific background is presented in section 2, followed by the research design of the interview study in section 3. Section 4 then illustrates the research results and discusses these in section 5.

2 Scientific background

Following, the view on the topic by the perspectives of MBSE, DA and combined approaches will be briefly described.

2.1 Model-based Systems Engineering (MBSE)

The evolution of mechatronic to cyber-physical systems changes how these are developed. Through the high degree of complexity, these systems need interdisciplinary methods to handle various stakeholders and domains. As a promising approach model-based systems engineering (MBSE) has set itself as one of the best methodologies in this field. MBSE is defined as "*the formalized application of modeling to support system requirements, design, analysis, verification, and validation activities beginning in the [concept stage] and continuing throughout development and later life cycle [stages].*" (INCOSE, 2007; INCOSE, 2023). MBSE is based on combining modeling methods (e.g. CONSENS, SYSMOD), modeling languages (e.g. SysML) and modeling tools (e.g. Cameo Systems Modeler, Enterprise Architect) to build system models as central development artifacts (Delligatti, 2013). These represent the system under development, its elements, characteristics, and internal relationships between the elements. Engineers use these system models to design and analyze the system at an early phase. With the INCOSE Vision 2035, a main competency of MBSE is seen in the analysis of data (INCOSE, 2021). But today´s standards and methods in MBSE do not address this aspect thoroughly (INCOSE, 2021).

2.2 Data Analytics (DA)

Data analytics (DA) allows one to discover insights and behavior of different objectives, including CPS, by accessing, aggregating, and analyzing data (Tyagi, 2003; Lee et al., 2014). With statistics, artificial intelligence and machine learning, DA is a data-driven approach and relies on context specific aspects to be successful and goal-oriented (Pohl et al., 2022). A typical application structure can be divided into descriptive, diagnostic, predictive, and prescriptive approaches (Ramanathan et al., 2017). The successful integration and use of DA in other domains are numerous: e.g. marketing, finance, manufacturing, distribution (Runkler, 2020). Therefore, the integration in development processes (e.g. MBSE), seems promising.

2.3 Combined approaches

Combined approaches addressing data-driven methods and development methods have risen in the last years but are rarely viewed with MBSE as the main point. Initial literature research here was done by Tissen et al. (2023), where a new concept called data-driven model-based systems engineering (DDMBSE) was proposed. DDMBSE enhances MBSE through the integration of data analytics and is initially defined as follows:

“Data-driven model-based systems engineering (DDMBSE) can be seen as an evolution of model-based systems engineering integrating data analytics aspects and contains six core aspects: (1) DDMBSE fusions utilities (processes, methods, techniques, languages, and tools) between model-based systems engineering and data analytics. (2) By this, data can be integrated into the system model, extending it to a data-driven system model. (3) Targeted data and information sources are here used, (4) iteratively looping and continuously acquiring data. (5) This is made possible by using standardized interfaces for the data-driven system model. (6) The data and system model relies on a strong visualization, enabling easy understanding of the system and data analyzed.” (Tissen et al., 2023)

As the authors state, organizations can gain valuable benefits from DDMBSE, e.g. *better managing of complexity*, *improving development processes*, *enabling data-driven decision making* and *reusing knowledge* (Tissen et al., 2023). But there are also challenges coming with, especially *supporting the application of DDMBSE through methods* (Tissen et al., 2023). With a view to the research question in section 1, this paper addresses these issues by getting knowledge from experts and practitioners in related fields.

3 Research design

To answer the research question above, an exploratory interview study with 16 experts and operators from various operational sectors was conducted. The research design includes initial data collection and subsequent data analysis. For the data collection the guidance of Saunders et al. (2019) and Eisenhardt (1989) is considered, resulting in the initial *preparation of an interview guideline*, the *selection of suitable interviewees*, the *conduction of the interviews* and the *documentation* of these. The interview guideline was designed in a semi-structured form and divided into six sections. The first four sections were used as a database for answering our research questions and the presented results in this article: (1) *Introduction*, (2) *definition of the research field data-driven model-based systems engineering*, (3) *preconditions for the success of data-driven processes in model-based systems engineering* and (4) *potentials and challenges of data-driven model-based systems engineering*. The following questions were addressed:

1. **Introduction**
    a. *What activities do you carry out in the field of MBSE and/or data analytics?*
    b. *How long have you been working in the field of MBSE and/or data analytics?*
2. **Definition of the research field data-driven model-based systems engineering (the DDMBSE definition from above (Tissen et al., 2023) was presented via a slide and read by the interviewees as well as the MBSE triangle (Delligatti, 2013) before asking the next questions)**
    a. *How would you define the research field "data-driven model-based systems engineering"? Do you think the first definition is suitable? Which points would you change/add?*
    b. *Regarding the MBSE triangle: Which elements do you see as necessary components for data-driven MBSE or for creating a data-driven system model?*
3. **Preconditions for the success of data-driven processes in model-based systems engineering**
    a. *What do you think are the preconditions for implementing data-driven model-based systems engineering?*
        i. *Stimuli: companies, people/users, research, processes, methods, technologies...*
    b. *What needs to be changed in the field of MBSE so that data-driven processes are increasingly used?*

4. **Potentials and challenges of data-driven model-based systems engineering**
    a. *What potentials and use cases are enabled by the analysis of data in the context of MBSE (with regard to modeling methods, languages, tools, system models, processes)?*
        i. *Example: Reduction of modeling times, e.g. through model recommendations from existing model elements or AI-generated model elements*
    b. *Which use cases have you or your company already applied?*
    c. *What challenges does the analysis of data in the context of MBSE pose (in terms of modeling methods, languages, tools, system models, processes)?*

For the selection of suitable interviewees, we contacted 24 persons via e-mail. The focus was set on persons working in MBSE and/or DA domains. The resulting number of interviewees with a positive response was 16. A detailed (and anonymized) overview of the interviewees can be seen in Table 1. The interviewees have shown a wide variety of their *working positions*, *years of experience* and *operational sector*, which in reverse, strengthens the validity of the results.

Table 1. List of interviewees (anonymized)

Interviewee	Current position	Years of experience	Interviewee domain	Operational sector
(A)	Head of department	ca. 6	MBSE & Data Analytics	Research
(B)	Senior systems engineer	ca. 10	MBSE	Tool Vendor
(C)	Team leader / consultant	ca. 11	MBSE	Tool Vendor
(D)	Chair member of association	ca. 4	MBSE	Research
(E)	Head of department	ca. 16	Data Analytics	Industry
(F)	Senior systems engineer	ca. 11	MBSE	Industry
(G)	Head of department	ca. 18	MBSE	Tool Vendor
(H)	Senior systems engineer	ca. 11	MBSE & Data Analytics	Industry
(I)	Researcher	ca. 2	MBSE & Data Analytics	Research
(J)	Professor	ca. 16	MBSE	Research
(K)	Senior systems engineer	ca. 7	MBSE	Consulting
(L)	Senior systems engineer	ca. 32	MBSE	Research
(M)	Innovation manager	ca. 10	MBSE	Consulting
(N)	Head of Company	ca. 6	MBSE & Data Analytics	Industry
(O)	Professor	ca. 13	Data Analytics	Research
(P)	Researcher	ca. 5	Data Analytics	Research

All interviews were conducted and recorded online as Microsoft Teams video meetings and lasted about 60 min each. Led by one interviewer, every video meeting was conducted individually. The records were then transcribed word-wise and anonymized to prevent tracing or interference. For the data evaluation the thematic analysis and clustering after Braun and Clarke (2006, 2012) was followed. The procedure allows to identify relevant patterns within the interview answers. This started by coding the interviews inductively without prescribing a code system but deriving the codes directly from the transcripts. Two researchers independently coded the interviews, resulting into two separate code systems. After the first interview both codes were compared to one another and discussed regarding similarities and differences. By this, both code systems were adapted to one another and led to one general code system. The interview was then re-read and adapted to the general code system. After the second interview, the general code system was updated with the new codes also discussed here by the researchers. The general code system was then applied to all following transcripts by both researchers. With the help of the codes, themes were searched within them oriented along with the research question. Clusters were formed with the themes, setting a core aspect into focus, and building patterns. Each quote was linked to one code. The patterns were then positioned to the respective research question. The results of this clustering will be shown in the following section.

4 Results

Analyzing the interviews yielded four main results. They contain a *refined definition for DDMBSE*, including a new model, called the *nexus model* (Section 4.1). Furthermore, *preconditions for DDMBSE* are shown (Section 4.2), enhanced with emerging *challenges* (Section 4.3), *potentials and use cases* (Section 4.4).

4.1 Refined definition of data-driven model-based systems engineering

Based on the initial definition of DDMBSE after Tissen et al. (2023), a refined formulation is derived from the database including a new description model for DDMBSE. The **refined definition for DDMBSE** is as follows:

"Data-driven model-based systems engineering (DDMBSE) can be seen as an evolution of classical model-based systems engineering enhanced by integrating data analytics aspects. DDMBSE fusions utilities (processes, methods, techniques, languages, and tools) between classical model-based systems engineering and data analytics, including the data engineering and algorithmic (e.g. by artificial intelligence) behind these.

It supports handling and understanding complex systems during their whole system life cycle (from the beginning of life: conceptualization, development, production; mid of life: distribution, utilization, support; to the end of life: system life span extension, material reuse, disposal, incineration). The purpose of DDMBSE is to integrate and link data into the (static) system model, extending it to a (dynamic) data-driven system model. The data-driven system model relies on a strong visualization, enabling easy understanding of the system and data analyzed within its context and environment.

Targeted data and information sources from system model internal (e.g. diagrams, partial models), system model external (e.g. system behaviour), company internal and/or company external sources are used. Current and/or previous data is iteratively looped and continuously acquired from defined points, phases, or the whole system´s life cycle and/or other system life cycles. This is made possible by using standardized ontological and technological interfaces and data formats for the data-driven system model, the data sources and used tools."

The new definition of DDMBSE includes a specification of its purpose, which lies in the handling and understanding of complex systems, e.g. CPS (noted by interviewee B, E, F, G, M, O). The purpose of DDMBSE is the modeling of a dynamic data-driven system model, which enables a continuous optimization within its partial models. The data sources for this can be identified along the life cycle of the system of interest within the system model (system model internal), e.g. its requirements, outside of the system model (system model external), e.g. test data from physical prototypes, and company internal and external data sources (noted by Interviewee B, C, E, F, J, K, L, N, O, P). A sharper definition of what DA aspects need to be considered is mentioned by interviewees A, B, F, and H, including the data engineering and algorithmics (e.g., through artificial intelligence). Many of interviewees recognized the initial definition as sufficient (Interviewee C, D, G, I, K, M, P).

As an extension of the refined definition of DDMBSE, a new description model is derived which connects (Latin: "Nexus" stands for English: "Connection") DA and MBSE. It is called the **"Nexus model for data-driven model-based systems engineering"**. The nexus model is presented in Figure 1.

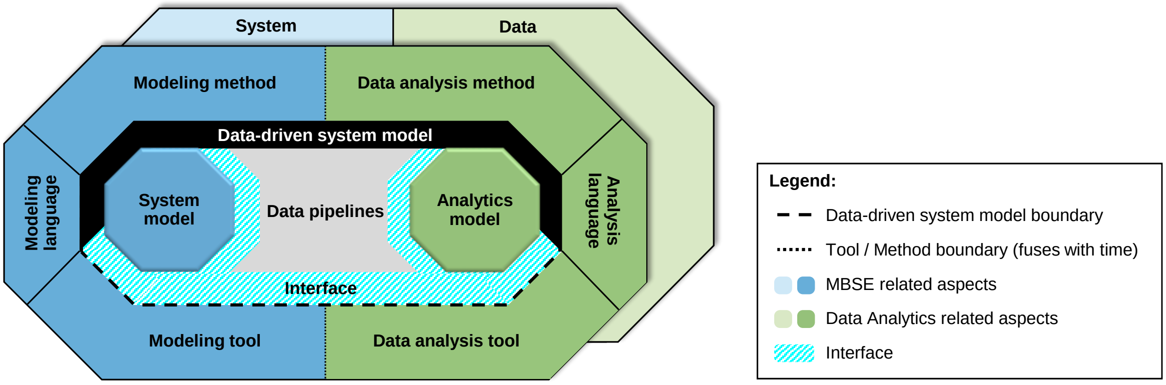

Figure 1. The new nexus model for data-driven model-based systems engineering

The DDMBSE nexus model extends the classical MBSE triangle and describes the necessary elements to model and develop a data-driven system model (see Figure 1). As before, a modeling method, modeling language and modeling tool remain necessary elements to model a system model based on the underlying system of interest. Enhanced is this by the aspects: data analysis method (e.g. CRISP DM, artificial intelligence), analysis language (e.g. Python, R), and data analysis tool (e.g. KNIME). With data as a base, these three aspects are necessary to form an analytics model corresponding to the system model. The analytics model builds and combines all necessary data together and sends these via standardized interfaces and data pipelines into the system model. The results and changes that happen by this are manageable through an interface to the modeling and data analysis tools. These are currently separated but tend to be combined in future tools. The same testimony can be seen in the methods, which are today mostly domain-specific but will be combined with DDMBSE methods in future research. The data-driven system model is the sum of system model, analytics model, interfaces, and data pipelines in between. Future research will form new content in all areas of the nexus model and build out the description of DDMBSE.

4.2 Preconditions for data-driven model-based systems engineering

With the refined definition of the DDMBSE, the question of how to explore and apply this new field of research now rises. Several preconditions must be fulfilled for this. The analysis of the interviews revealed five areas in which preconditions

for DDMBSE must be achieved: *Research*, *Company & User*, *Tools & Tool provider*, *Applicability*, and *Data* (see Table 2).

In terms of **research**,

"*[...] for many (people) from the application sector, this (DDMBSE) is a topic of spearhead research." (Interviewee N)*

Use cases and best practices are called for above all, closely followed by the development of the actual DDMBSE methods. In addition, the interviewees see the need to link the two research areas and to develop basic concepts for the application and description of DDMBSE. This must be achieved by bringing together research and the user community. In addition, overlaps and differences to MBSE must be identified. The question must be asked whether existing basic concepts of MBSE are not called into question, as new technologies (e.g. AI) must now be included. This primarily concerns the data analysis perspective, whereby new perspectives must be incorporated and supplemented. Accordingly, modeling languages need to be more formalized to make greater use of data-driven processes.

Table 2. Preconditions for data-driven model-based systems engineering

Research *What preconditions does the research have to fulfill?*	Description of use cases and best practices (D, F, J, K, N, P)	Development of DDMBSE methodologies (A, B, C, L, P)
	Research fields must be combined (A, F, P)	Basic concepts for the application and description of the DDMBSE (B, J, P)
	Bringing the research and user community together (N)	Identify differences/synergies with MBSE (D)
	Question the basic concepts of MBSE, as new technologies (J)	New views (data analytics) must be added (B)
	Formalization of the modeling language (B)	
Company & User *What preconditions must companies and users fulfill?*	Clarification of added value and potentials of DDMBSE in the company (A, D, F, H, J, K, N, P)	MBSE must be established and widely used (B, F, H, K, M, N)
	Organizational integration of (MB)SE (B, D, K)	Create clear processes at the organizational level (E, K, O)
	Consider scaling (industry / SMEs) (D, E)	Create technical infrastructure (G, K)
	Clarification of data protection when collaborating with partners (J, M)	Target definition/analysis of the current situation (J, P)
	Data analytics must be established (N)	Companies must think data-oriented (J)
Tools & Tool providers *What preconditions must tools and tool providers fulfill?*	Interfaces of tools (D, E, K, L, O)	Standardization of (modeling) languages (B, L, M)
	Standardization of interfaces (B, C, M)	Creating traceability of change processes (A, C)
	The maturity of data analytics tools must increase (N, O)	Simplified use of data-supported processes (I)
	DDMBSE providers and services must be available (H)	Tools must have the very latest technological maturity (O)
Applicability *What preconditions must be fulfilled for application?*	Training and user education (A, B, G, H, K, L, M, N)	Competencies and mindset for MBSE (A, B, D, F, L, M, N, O)
	Employee acceptance and motivation (E, F, H, K, M, O, P)	Competencies and mindset for data analytics (A, L, M, O)
	Clear description of objectives during implementation (P)	Re-definition of roles (A)
Data *What preconditions must be fulfilled by the data?*	Data (critical amount of necessary data, labeled) (A, B, E, G, H, I, K, L, P)	Data management (quality, processing, backup) (A, C, D, E, G, H)
	Providing data pipelines (H, P)	Allow data access and manipulation (H, K)
	Identify data sources (G, L)	Standardization of data models (B, G)
	Allow/enable data and network continuity (F, G)	(Increased) inclusion of the system life cycle (G)
	Data networking/reduction of data silos (J)	

In terms of **companies and users**, many respondents felt that the most important precondition for DDMBSE is that the added value or business impact for the company must be clear. MBSE must first be established and, more importantly, integrated into the organization along with the fundamentals of systems engineering:

"[…] MBSE is a subject area that is easily over 20 years old in research, perhaps we could even say almost 30, but it has only partially arrived in the industrial context." (Interviewee B)

This requires clear processes at the organizational level and broad industrial dissemination of MBSE in general. In addition, the technical infrastructure must be in place and the cooperation with partners must be specified with regard to data security. Above all, companies must know their current situation and be able to define their target situation. They must also be able to scale for large and small organizations. Companies must enable and continue to think data-driven and have data analytics embedded in the organization.

With respect to **tools and tool providers**, respondents see a demand for interfaces to the tools themselves:

"[…] tool providers have to create these interfaces at all so that there is no obstacle to this." (Interviewee D)

This needs to be complemented by the standardization of modeling languages and the interfaces themselves. Traceability of change processes is also requested to make the multitude of data integrations understandable for the user. There is a lack of tools, vendors, and services for DDMBSE. Therefore, existing MBSE and data analysis tools need to be enhanced in terms of functionality and technological maturity. In addition, the handling of data-driven processes needs to be simplified.

With view on the **applicability,** most users will need training for DDMBSE.

"[…] it is super important, that you have the right competence carriers, that you also have a certain excellence in there [...]. In other words, you simply need the people who deal with these topics and know them." (Interviewee A)

First, the skills and mindset for MBSE must be available and accepted by employees. Then, the skills and mindset for data analysis must be motivated. From the interviewees' point of view, this requires a clear description of the objectives for the applicability of DDMBSE. This is inherently linked to redefining roles and tasks within teams, departments, and organizations.

Looking at the preconditions in terms of **data**, data must first be available:

"In order to be able to work with data at all, I first need to have data." (Interviewee L)

The interview partners see data management in terms of quality, processing, backup and the corresponding data as a decisive precondition. Appropriate data pipelines are also required. In addition, it must be ensured that the data can be accessed and manipulated, which is also a possible risk when not available. The correct identification of data sources is seen by some respondents as a necessary step. Regarding data models, standardization, and consistent networking of data across the entire system life cycle is required. In addition, data silos need to be reduced and made accessible.

4.3 Challenges for data-driven model-based systems engineering

Following the presentation of the preconditions in section 4.2, the question now arises as to what challenges DDMBSE faces. Six areas of challenges have been identified: *Tools*, *Data handling*, *Users & Application*, *Organizational aspects*, *Scope definition*, and *General aspects* (see Table 3).

When it comes to **tools**, the question arises as to how the DA functionalities should be implemented. Due to the large number of thematic isolated solutions, system exchange and formalization is a major challenge. This is supplemented by **data handling**. The interview partners see the challenge here in identifying suitable data and data sources that ensure sufficient data quality.

"If we now want to (operate) data-driven MBSE, […] then I first have to identify the right information or data sets." (Interviewee G)

This is supplemented by data management and the selection and use of suitable databases. The integration of the data into the system model itself and suitable data pipelines and associated algorithms are the next priority here, as well as the high level of data diversity and data security.

For the **users & application** of the DDMBSE, the greatest factor is user acceptance as well as the corresponding training and knowledge development. The question here is where the boundary lies between user tasks and software tasks. A challenge for **organizations** is the initial effort assessment of the DDMBSE as well as the implementation of decision-making processes and the associated change management:

"[…] to promote this acceptance not only at management level, but also at user level. I think it will be quite a challenge to show that this (DDMBSE) is better now or that you really can save resources." (Interviewee O)

Large companies have the problem of sluggish organizational structures or inadequate IT infrastructures. Due to the heterogeneity of information and data silos in companies as well as the diversity of the domains themselves, tailoring must take place. Regarding the **scope definition**, the main challenge is to define and specify the right level of abstraction in terms of modeling, but also the purpose. This can result in product-specific semantics.

In **general,** the combination of DA and MBSE represents a challenge.

"So, if there is no support, in terms of tools and methods, then it (DDMBSE) really becomes a burden for people." (Interviewee C)

It is unclear how the technological implementation should take place and how the results of the data-based procedures should be validated. There is a lack of standards for terminology. It is unclear what impact MBSE will have on system development and the associated modeling languages, methods, and tools.

Table 3. Challenges of data-driven model-based systems engineering

Tools *What challenges must be faced for the tools?*	Implementation of data analytics features in the modeling tool (C, H, O)	Large number of topic-related stand-alone software solutions (O, B)
	System model exchange (I, B)	Model formalization (I)
Data handling *What challenges must be faced for handling the data?*	Identification of suitable data and data sources (B, D, G, P)	Data quality (E, O, P, L)
	Data maintenance and management (G, N, O)	Database selection and connection (I, P)
	Data acquisition (E, L)	Integration of data into system model (E, M)
	Data pre-processing/pipeline (O, P)	Selection of algorithms (O, P)
	High heterogeneity of data (E, J)	Basic quantity of data (L)
	Data security and release (A)	
User & Application *What challenges must be faced by the user and the application?*	User acceptance (A, B, D, M, O)	Employee training and knowledge development (D, M, K)
	Task distribution between user and software (A, F)	Little to no best practices (P)
	Generational conflict of users (K)	
Organization aspects *What challenges must be faced by organizations?*	Effort estimation (D, K, O)	Execution of decision-making process and change management (A, C)
	Slow organizational structures in large companies (A, F)	Required IT infrastructure and digitalization (O)
	Introduction and application of DDMBSE (K)	Securing respective authorizations and accesses (M)
	Data and information silos in companies (D)	Tailoring of DDMBSE processes and methods (F)
Scope definition *What challenges must be faced for defining a scope of application?*	Finding the right modeling level across the domains (E)	Finding the right abstraction levels for purpose (E)
	Product-specific semantics (E)	Finding the right analysis method for purpose (B)
General aspects *What general challenges must be faced?*	Combination of data analytics and MBSE (A, F, H, P)	technological implementation (D)
	Lack of method for realizing the DDMBSE at system level (H)	Validation of data-driven results (L)
	Anticipating the modeling process for later DDMBSE application (B)	Standardization of terms (J)
	Unclear impact on system development (H)	Unclear impact on modeling language (H)

4.4 Potentials and use cases for data-driven model-based systems engineering

Along the preconditions and challenges shown in section 4.2 and 4.3, DDMBSE offers various potentials and use cases. The potentials and use cases are clustered into: *Company & Supplier*, *Advanced traceability*, *Development process*, *Complexity management*, *System model properties*, and *Tool usability & Assistance* (see Table 4).

DDMBSE offers many advantages for **companies and suppliers**. Among other things,

"[…] supporting people by providing the right information at the right time in a context-sensitive manner […]" (Interviewee A)

empowered by stakeholder-specific views and documents can also be extracted. This enables customers, but also partners, to analyze and integrate more closely. Furthermore, company-specific information and knowledge can be reused and displayed, and documentation activities can be automated. Overall, this reduces the entry threshold for (DD)MBSE.

Table 4. Potentials and use cases of data-driven model-based systems engineering

Company & Supplier *What potentials and use cases emerge for companies and suppliers?*	Stakeholder-specific views and document derivation (A, D, F, N)	Improved analysis and integration of customers and partners (C, F, H, N)
	Provide and reuse company-specific knowledge and data (A, D, N)	Automating activities and documentation (A, M)
	Lowering the entry barrier for MBSE (G)	
Advanced traceability *What potentials and use cases emerge for traceability?*	Establishment and improvement of cross-model traceability (C, F, G, E)	Improved general traceability (A, E, K)
	Automated generation of trace links (M, K)	Traceability of models across generations (C)
	Automatic consistency checks of data records for a product (M)	
Development process *What potentials and use cases emerge for the development process?*	Reduction of modeling time through reuse of data and models (B, C, H, I, J, L)	Optimization and support of transparency of (development) processes (L, P, E)
	Automation and scaling of development processes (I, K)	Retrospective and prospective evaluation and creation of non-MBSE models (B, O)
	Reduction of information search times (H)	Support in the application of methods (D)
Complexity management *What potentials and use cases emerge for managing complexity?*	Gaining knowledge through analysis of field data (E, F, L)	Improved analysis of the (system) context (C, G)
	Decision-making support (G, H)	Reduction and optimization of variant diversity (I, N)
	Automation of information provision (F, H)	Derive expert knowledge from models (D)
	View-specific extraction of data (D)	Standardized data formats (A)
System model properties *What potentials and use cases emerge for the system model properties?*	(interdisciplinary) automatic generation of system models (B, C, E, H, I, N)	Increased model quality through continuous updates and optimization (A, G, L, O, P)
	Plausibility and consistency check of system models (B, H, M)	Simulation and evaluation of systems based on real data from the field (digital twin) (H, O, J)
	Generation of system architecture alternatives (N, G)	Reduction of error sources through automatic notifications (M, L)
	(Automatic) derivation of system model (J, O)	store expert knowledge in system models (demographic change) (B)
	Checking requirements (K)	
Tool usability & Assistance *What potentials and use cases emerge for tool usability and user assistance?*	Context-specific and user-oriented support for system modeling (through data) (A, B, D, O)	AI-based interface assistance between tools (A, M)
	Mapping of source data via various development artifacts to the system model (E)	Change/variant management assistant (A)
	Digitalization of workshops (O)	Deriving knowledge from models for domain beginners (B)

Traceability is experiencing significant progress, particularly through cross-model traceability. This makes it possible to link models from different domains but also to connect model generations across generations. Further potential is the automation of training and consistency checks of the data for a specific product. Regarding the **development process**, the majority of those interviewed see a significant reduction in modeling time through the reuse of data and models. In addition, to optimize and support transparency throughout the entire development process and beyond by

“[…] increasing the efficiency of modeling, efficiency in that I can access information more quickly, for example.“ (Interviewee H)

This is made possible by automating and scaling development processes and supporting the application of methods. The construction and completion of holistic development databases can be made possible through the retrospective and prospective evaluation and creation of non-MBSE models. **Complexity management** is supported by analyzing field data by providing new insights and knowledge about the system. Contextual analysis of the system is also promoted, which allows decisions to be made more informed. An automated and view-specific extraction of data, information and expert knowledge from the models allows improved knowledge management. For **system model properties**, the automated interdisciplinary generation of system models represents the greatest potential, followed by quality improvement through continuous updates.

"We have already seen how quickly generative AI can produce 80% solutions, which may not be perfect, but not everyone has to aspire to that. So, if you really manage to get an 80% solution, an MBSE solution, for your new project at the click of a mouse, [...] that's a dream come true." (Interviewee I)

The system model is optimized through plausibility and consistency checks of the system model as well as the evaluation of the system using field data. When modeling, the generation of system architecture alternatives opens new possibilities for system development. Through automated notifications of errors, these can be dealt with as early as possible and can also be checked against existing requirements. For **tool usability and assistance**, the most important thing is context-specific and user-oriented support in the system modeling process. AI-based interfaces can serve as assistants to improve exchange between tools. This allows a lot of data to be built up beyond various development artifacts to form the system model. Further potential arises from the digitalization of workshop results as well as through variant management assistants.

When asked which of the participants had already developed or implemented their first use cases, 9 out of 16 people stated that they had not yet developed or implemented any use cases. However, 10 different use cases have already been implemented at a prototype level (see Table 5). These partly dealt with the application of AI algorithms to optimize parameters in subsystems, but also with the creation of system architectures based on natural language. Further use cases dealt with the use of real data from the field, which was fed back into the system model. In teaching, among other things, some project work was carried out with students. Nevertheless, the implementation of the use cases of DDMBSE is minimal. Accordingly, there is still a lot of untapped potential that needs to be further explored.

Table 5. Potentials and use cases already implemented or realized by the interviewees

None	C, E, F, G, H, L, M, P, N
Prototype level *Implementation of first use cases at a very low level of technological maturity*	Integration of a language model in recommendation loops (A)
	Stakeholder-specific documentation of a system model (A)
	AI algorithm for optimizing parameters of a subsystem (B)
	Deriving offer documents from a system model (D)
	Mapping of requirements and passing criteria (I)
	Creating a system architecture based on natural language (I)
	Recognition of components within an MBSE system architecture (I)
	Comparing a system model against real data (O)
	Back-transformation of error data from the field into the system model (K)
	Project work with students (J)

5 Discussion and conclusion

Contribution: The paper at hand provides a comprehensive review of the preconditions, challenges, potentials and use cases in the context of integrating data analytics into MBSE, called data-driven model-based systems engineering (DDMBSE). **Initially** a **refined definition for DDMBSE** is formulated based on the initial definition after Tissen et al. (2023) derived from a systematic literature review. The refined definition aims to clarify the research field and manifests it. Additionally, a new description model is presented which is called the **nexus model for DDMBSE**, connecting MBSE and data analytics aspects. The nexus model contains necessary elements to successfully build a data-driven system model. It enhances the classical MBSE triangle, which is widely used and accepted. **Second**, the analysis reveals **preconditions** which need to be fulfilled for establishing DDMBSE. The preconditions address different areas, from the *Research*, *Company & User*, *Tools & Tool provider*, *Applicability*, and *Data*. All preconditions in the areas are described in detail. **Third**, the paper contains a comprehensive overview of the identified **challenges faced by DDMBSE**. The challenges are clustered into the six groups *Tools*, *Data handling*, *Users & Application*, *Organizational aspects*, *Scope definition*, and *General aspects*. Main challenges are the need for further buildup of DDMBSE and its concepts, the identification of suitable data sources and the training and acceptance of DDMBSE by users. **Fourth**, an overview of **potentials and use cases** offered by applying DDMBSE are described. The potentials reach from *Company & Supplier*, *Advanced traceability*, *Development process*, *Complexity management*, *System model properties*, to *Tool usability & Assistance*. Main potentials lay in the interdisciplinary generation of system models, the reduction of modeling time by reusing data and models, stakeholder-specific views and documents, improved integration of customers and partners and advanced traceability of system models and properties. It also reveals that currently, most realized use cases are prototypes and not yet used actively in development processes of companies and in research.

Limitations: The results shown above are based on 16 interviews with experts and operators in MBSE and data analytics. However, further interviews may lead to a wider overview of the topic. The interviewees were all from Germany, which limits an international generalization of the findings. However, the results are derived from various areas, which supports a solid database. Also, the interviewees present their subjective view on the topic, which is restricted by their academic and industry background. Further insights may have appeared by other executions than via video meetings and the evaluation methods (e.g. using a grounded theory approach).

Future research: The interview study reveals that the integration of DA into MBSE offers various potentials and use cases, which are currently only on the horizon visible. With the advancements in the areas of DA and AI, a rapid change in the execution of development methods will appear and support the further rise of DDMBSE. The preconditions for this take off are further, also linked to risks and restrictions, and more detailed research in the areas of DDMBSE is needed. New methods and tools must be developed, integrating advanced data analysis techniques into MBSE. Additionally standardized interfaces are needed in the future supporting heterogenic domains and data. The research also needs to train future engineers in academic and industry, to support the change.

References

Biggs, G., Juknevicius, T., Armonas, A., Post, K. 2018. Integrating Safety and Reliability Analysis into MBSE: overview of the new proposed OMG standard. INCOSE International Symposium, 28, 1322-1336.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, pp. 77–101.

Braun, V., Clarke, V., 2012. Thematic analysis. In: Cooper, H., Camic, P.M., Long, D.L., Panter, A.T., Rindskopf, D., Sher, K.J. (Eds), APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. American Psychological Association, Washington, pp. 57–71.

Delligatti, L. 2013. SysML Distilled: A Brief Guide to the Systems Modeling Language, 1st ed, Addison-Wesley.

Eisenhardt, K.M., 1989. Building theories from case study research. The Academy of Management Review 14, 532–550.

INCOSE, 2007. Systems Engineering Vision 2020. The International Council on Systems Engineering (INCOSE).

INCOSE, 2021. Systems Engineering Vision 2035 – Engineering solutions for a better world. The International Council on Systems Engineering (INCOSE).

INCOSE, 2023. Systems Engineering Handbook – A guide for system life cycle processes and activities – Fifth edition. John Wiley & Sons Ltd., Hoboken, New Jersey.

Lee, J., Bagheri, B., Hung-An, K., 2014. Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics. Proceeding of Int. Conference on Industrial Informatics (INDIN), 1-6.

Meyer, M., Fichtler, T., Koldewey, C., Dumitrescu, R., 2022. Potentials and challenges of analyzing use phase data in product planning of manufacturing companies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing 36, 1-13.

Mütsch, F., Gremmelmaier, H., Becker, N., Bogdoll, D., Zofka, M.R., Zöllner, J.M., 2023. From Model-Based to Data-Driven Simulation: Challenges and Trends in Autonomous Driving. arXiv.

Pohl, M., Staegemann, D.G., Turowski, K. 2022. The Performance Benefit of Data Analytics Applications. The 12th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS), Elsevier, Porto, Portugal, pp. 679-683.

Ramanathan, R., Mathirajan, M., Ravindran, A.R. (Eds.), 2017. Big Data Analytics Using Multiple Criteria Decision-Making Models, CRC Press, Boca Raton, Florida.

Runkler, T. A., 2020. Data Analytics – Models and Algorithms for Intelligent Data Analysis 3. Springer Vieweg Wiesbaden.

Saunders, M., Lewis, P., Thornhill, A., 2019. Research Methods for Business Students. 8th Edn. London, New York, Ann Arbor: Pearson; ProQuest Ebook Central, Harlow, England.

Statista, 2021. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes). URL: https://www.statista.com/statistics/871513/worldwide-data-created/.

Statista, 2022. Spending on digital transformation technologies and services worldwide from 2017 to 2026 (in trillion U.S. dollars). URL: https://www.statista.com/statistics/870924/worldwide-digital-transformation-market-size/.

Tissen, D., Wiederkehr, I., Koldewey, C., Dumitrescu, R., 2023. Exploring data-driven model-based systems engineering: a systematic literature review, IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), Rabat, Marocco.

Tränkler, H.R., O. Kanoun, O., (2007). Smart Systems and Devices: Innovative Key Modules for Engineering Application. Conference Smart Systems and Devices, Hammamet, Tunisia, pp. 3-12.

Tyagi, S., 2003. Using data analytics for greater profits. Journal of Business Strategy 24, 12–14.

Contact: D. Tissen, Paderborn University, Heinz Nixdorf Institute, Fürstenallee 11, 33102 Paderborn, Germany, Phone: +49 5251 60-6494, E-Mail: denis.tissen@hni.uni-paderborn.de