**INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024**
https://doi.org/10.1017/pds.2024.212

# Challenges for capturing data within data-driven design processes

Christopher Langner [1,✉], Yevgeni Paliyenko [1], Benedikt Müller [1], Daniel Roth [1], Matthias R. Guertler [2] and Matthias Kreimeyer [1]

[1] University of Stuttgart, Germany, [2] University of Technology Sydney, Australia

✉ christopher.langner@iktd.uni-stuttgart.de

**Abstract**

Cyber-Physical-Systems provide extensive data gathering opportunities along the lifecycle, enabling data-driven design to improve the design process. However, its implementation faces challenges, particularly in the initial data capturing stage. To identify those, a comprehensive approach combining a systematic literature review and an industry survey was applied. Four groups of interrelated challenges were identified as most relevant to practitioners: data selection, data availability in systems, knowledge about data science processes and tools, and guiding users in targeted data capturing.

*Keywords: data-driven design, cyber-physical systems, internet of things (IoT), data science*

## 1. Introduction

The core of a design process is to make many decisions impacting the final product (Hazelrigg, 1998). These decisions are based on information and knowledge (Chaudhari et al., 2020). Data, information, and knowledge build upon each other, which North (2021) visualizes with his knowledge staircase displayed below in Figure 1, which also contains examples of product design for each level.

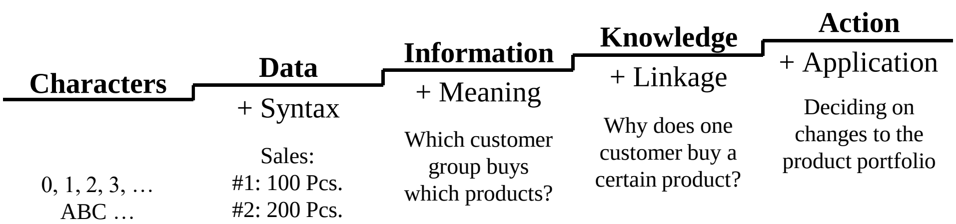



**Figure 1. Knowledge staircase based on North (2021)**

Cyber-Physical-Systems (CPS) provide the possibility to enlarge the sources for data gathering (Bertoni, 2020), since they are enabled to communicate with each other and with the manufacturer by using the Internet of Things (IoT), and therefore connect with (theoretically) every stakeholder or process step during operation (Huber, 2018). This opens the door to unprecedented possibilities to get access to extensive, varied, and context-related data along the whole product life cycle (PLC). With the growing capabilities of Data Science, this data can be exploited to extract valuable information and knowledge from it (Kim et al., 2017). By integrating data collection, processing, and utilization into the design process, this potential can be used to improve the design process (Altavilla et al., 2017). In this context, the term data-driven design (DDD or D$^3$) has become established over the last years

(Kim et al., 2017). It describes decision-making in the design process based on preceding data gathering and analysis (Gerschütz et al., 2021; Holmström Olsson et al., 2019).
Although it holds great potential, the implementation of DDD is hindered by several challenges which have been investigated by previous work. The initial capturing of data has been identified as one of the most urgent issues to be solved (Briard et al., 2021; Mehlstäubl et al., 2022).
However, it remains unclear what the related concrete challenges for data capturing are and which implications result from this for product development in creating a suitable solution. Therefore, the goal of this work is to investigate which concrete issues for the capturing of data within design processes exist and what implications they impose onto product development. Accordingly, this paper focuses on two research questions:

RQ 1: What are the concrete challenges of capturing data within a design process?
RQ 2: Which implications do these challenges impose onto product development?

To answer these questions, this paper is structured as follows: Section 2 presents related work in current research, followed by the description of the used research methods in Section 3, namely a systematic literature review (SLR) and an industry survey. The findings are presented in Section 4 and discussed in Section 5 as well as concluding this paper and providing an outlook for potential future research.

## 2. Related work

In this section, a selection of publications dealing with the topic of challenges within DDD is presented. Mehlstäubl et al. (2022) conducted a survey with industry practitioners, mainly with participants in managing positions in research and development departments. The collection of data is identified as the most important challenge, along with an insufficient variety of data, followed by the pre-processing of data and the interdisciplinary collaboration with data science experts. Furthermore, the correct interpretation of data, as well as the storage of data and data security are found to be urgent issues. In contrast, a too small data volume as well as a missing will for change within both designer and management levels are revealed as non-acute aspects.
Based on the experience of several industry use cases with manufacturers of embedded systems, Holmström Olsson et al. (2019) derived key challenges in adopting DDD. Data collection is named as a major issue here as well. Furthermore, the struggle to define factors describing the value-adding potential of data-driven features is mentioned. This can be related to an insufficient understanding of the target setting, and what exact goal a data-driven process should reach. That is also connected to the necessary change management in traditional hardware-based companies to establish trust in data-driven processes. In addition, the need to shorten development cycles to create the ability to adapt to fast-changing information based on up-to-date data analysis is revealed, as well as the complexity of data governance.
Briard et al. (2023) conducted an industry workshop to investigate challenges with DDD related to the early stages of product development. The required investment in new digital technologies and digitally skilled experts is identified as one of them. Connected to that, missing use cases to assess the potential of data usage make it hard to convince management to make these investments. Furthermore, methods for supporting designers with the new tasks coming from DDD are missing, addressing not only the process of DDD but also the design of products equipped with suitable sensors to collect data itself. Data security and acting compliant with data privacy laws are described as additional challenges.
According to Cantamessa et al. (2020), the challenges related to the introduction of DDD can be divided into three groups: related to the designers, the design processes, and methods for data analytics. Regarding designers, the need to learn how to design smart products being able to gather data in a suitable way, as well as understanding the value-adding potential of data is identified. Regarding design processes, most importantly the challenge of developing a strategy for data collection and processing needs to be faced, again emphasizing the need to understand the concrete issues related to the capturing of data. Methods for data analytics must be empowered to unveil new information and knowledge from big datasets to effectively support design processes.
Bertoni (2018) investigates challenges of DDD in the product innovation process. First, a general issue of decision-making is described: people often avoid the choice of radical new ideas, preferring to stay



https://doi.org/10.1017/pds.2024.212 Published online by Cambridge University Press

with at least in parts known concepts from previous experiences. With the described potential of data to unveil new patterns or ideas, this could lead to a mistrust of data-driven processes of designers. Second, the right data interpretation connected with the information completeness is named as a possible challenge. Finally, the cognitive limitation of designers to work with the data processing results is identified.
Numerous other publications describe issues related to DDD, showing the importance of this topic. As many of the described works have shown, the initial step of DDD, the collection of data, is one of the major issues, however not describing the concrete challenges for it, substantiating the need to further investigate in this field.

# 3. Methodology

To answer the posed research questions, a comprehensive approach combining an SLR and an industry survey is applied. The outcomes of the SLR are used as a basis for the questions in the industry survey. Both steps are described in more detail in the following.

## 3.1. Initial collection of issues

Followed by a general literature review, a SLR visualized in Figure 2 is conducted. The search was performed in two databases *Scopus* and *Web of Science* in July 2023 with a first iteration based on a systematic search string and a second iteration of analyzing the cross references of publications found through the search string in the first iteration.

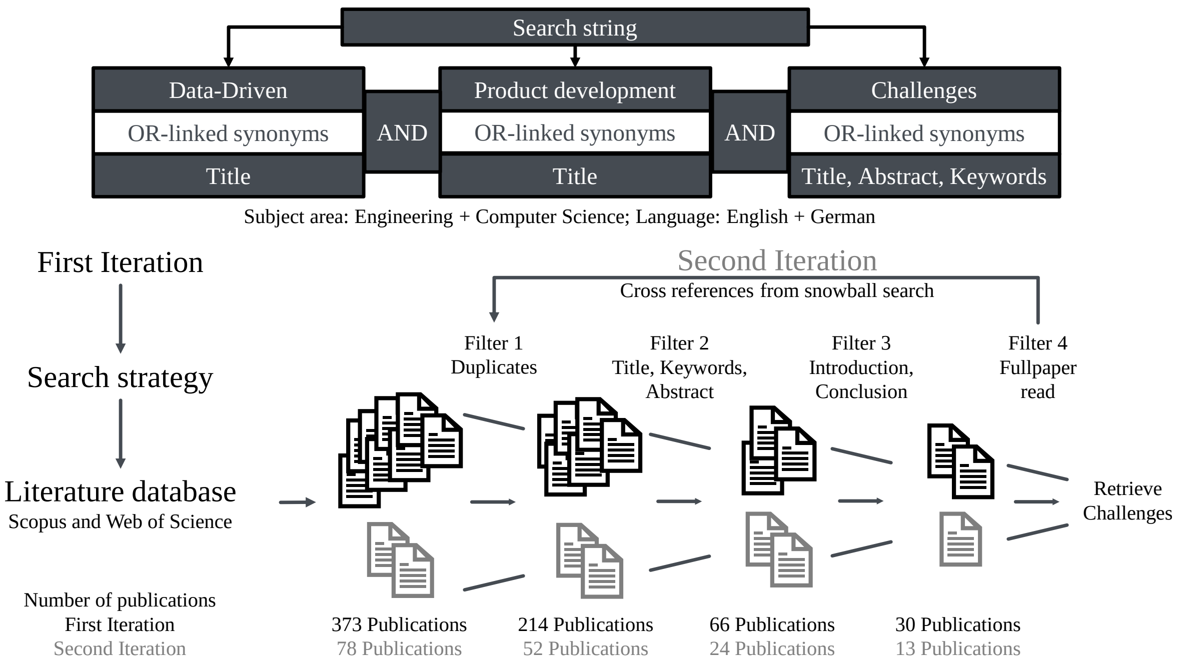



**Figure 2. Systematic literature review process**

Three terms are considered relevant: *data-driven, design / product development*, and *challenge*. For each term, synonyms in both German and English are added to result in the following search string: *("data$driven" OR "data$based*" OR "data$support" OR "daten$getrieben*" OR "daten$basiert*" OR "daten$gestützt*" OR "daten$unterstützt*") AND ("design$process" OR "development$process" OR "product$develop*" OR "product$design*" OR "data$driven$design" OR "data$driven$development" OR "data$driven$engineering" OR "*entwicklung*" OR "*konstruktion*") AND ("challeng*" OR "problem*" OR "issue*" OR "difficult*" OR "dilemma*" OR "trouble*" OR "require*" OR "demand*" OR "need*" OR "obstacle*" OR "complicat*" OR "herausforder*" OR "hindernis*" OR "schwierigkeit*" OR "forderung*" OR "komplikation*" OR "hürde*")*. To limit the outcomes to the core topic of DDD, the terms *data-driven* and *design* are required in the title of the publication. The term



https://doi.org/10.1017/pds.2024.212 Published online by Cambridge University Press

*challenge* is searched in the title, abstract, or keywords. Since the single term of design is too general, more specified terms for product development as shown above are used.

The search was limited to the subject area of *engineering* or *computer science* and the *English* or *German* language. As described in Figure 2, 373 publications were found initially in the first iteration. After filtering out the duplicates, the title, abstract, and keywords of each publication were reviewed to ensure that the term *design* is used in the meaning of *product design / product development*, e.g. excluding game design, health design, investment portfolio design, or teaching design. Furthermore, too specific topics like a specialized water filter controller design were excluded. The introduction and the summary & conclusion sections of the remaining 66 publications were read to filter, if a concrete challenge or at least the term of *challenge* is named, especially in the conclusion. 30 publications were left for a full paper-read. Named challenges in these publications were extracted and with the cross references from them, a second iteration of the literature search was conducted. The included publications underwent the same filtering process as in the first iteration. 13 remaining afterward led to a total of 43 full-read publications.

The described challenges related to the capturing of data within a design process are collected and clustered into thematic groups. They are used as the input for the survey which is described below. Both the challenges and the clusters are presented in Section 4.

### 3.2. Structure of the survey

The survey is divided into five sections. The first section gives a short introduction to the survey. The second section contains general questions about the participants regarding their professional experience, their field of activity, their company, and their position within the organization. The third section provides a brief introduction to DDD by stating definitions and explanations of the terms being used. After that, general questions related to DDD are given to identify the relevancy of DDD within the company and the proximity of the participants to the subject area. Furthermore, it is asked which degree of processing the data the participants encounter in their work activities has. Here, the possibilities to answer are related to the knowledge staircase as in Figure 1.

The fourth section of the survey addresses the concrete issues related to data capturing within DDD, clustered into operational, data-related, technical, organizational, and regulatory challenges. A five-step Likert-style scale from *not relevant at all* to *extremely relevant* is given to participants to assess the relevancy of the named challenges, as well as the possibility to abstain with *can not assess*. The challenges are listed together with an exemplary source from the SLR and the outcomes of the survey in Tables 1 - 5.

The fifth and last section allows participants to state further challenges and comments before finishing the survey. In this section, open-field answers were requested.

The survey was distributed through several industry networks and conducted with German industry partners; therefore, the original questions and answers are translated in this paper.

## 4. Results

This section at first describes the survey participants and then presents the outcomes of the survey and the derived implications they impose onto product development.

### 4.1. Description of survey participants

A total number of 52 participants answered the questions. Their properties can be derived from the answers given to the general questions displayed in Figure 3 (a) - (h).

The majority are very experienced (a) and mainly work in the research or development departments, while the area of IT is related to a fifth of the participants (g). They are distributed over the different levels of responsibility from being employed without management responsibility over project management up to the management of the company (b). This makes the group of participants very valuable since they can give first-hand perspectives from different responsibility levels while having sufficient experience to make profound assessments. Over 50 % are related to vehicle manufacturing and another 25 % to general mechanical engineering (d). The number of employees of the companies is



https://doi.org/10.1017/pds.2024.212 Published online by Cambridge University Press

mostly high, meaning more than 10.000 (e), giving the outcomes a strong perspective of participants with a background in large engineering companies. Half of the participants deal daily with data acquisition, data evaluation, or data-driven decision-making processes in their job, another quarter weekly and a tenth at least monthly (c), underlining the suitability of the participants for this survey. Interestingly, most of them work with pre-processed or completely processed data (= information), followed by a smaller share of interpreted data (= knowledge) (h). This could be based on the smaller proportion of managers within the group over employees and project managers. Only a fourth work with raw data, which can be explained by raw data mostly being processed by data scientists or IT experts rather than engineers. Emphasizing the need to work on the topic of DDD is given by the high relevance of this topic within all companies (f).
Note: Figure 3 h) is not answered by those who stated "never" as their proximity to the topic in c).

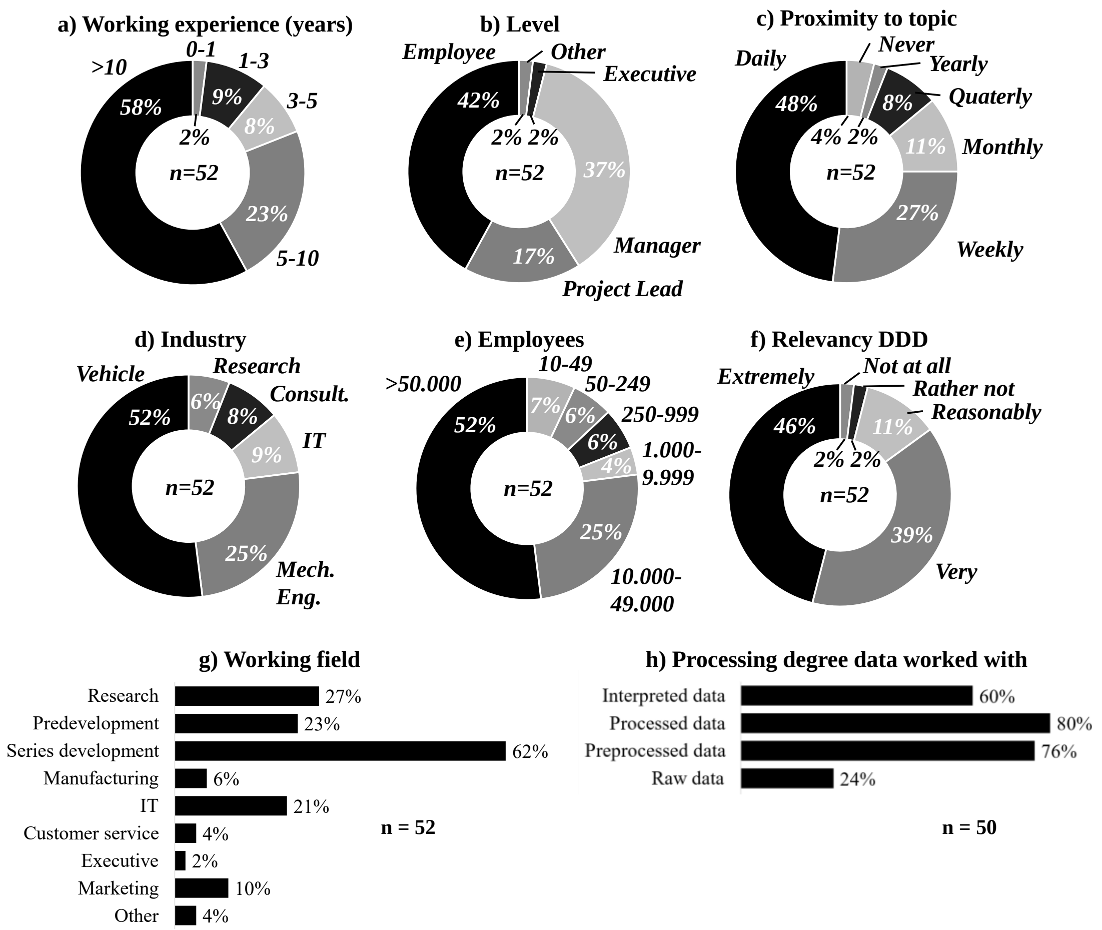



Figure 3. Information about participants

## 4.2. Challenges for data capturing

As described, a five-step scale is given to participants to assess the relevancy of the named challenges. Translated into the numbers given in the outcomes, this can be divided into *not relevant at all (1)*, *rather not relevant (2)*, *reasonably relevant (3)*, *very relevant (4)*, and *extremely relevant (5)*. Further, the possibility of answering with *can not assess* to abstain is given, explaining the varying numbers of answers ($n$) over the challenges.
In the following, the displayed tables include the challenges extracted from the SLR sorted in groups, each group being represented by a separate table and their rating of relevancy from the survey. An exemplary source from the SLR is given for each challenge. The challenges are sorted by the rating given in the survey from highest to lowest and supplemented by a graphical representation of the distribution of answers by showing the *average* $\bar{x}$ and *standard deviation* $s$. In general, most of the issues are rated as at least reasonably relevant, with only *ethical hurdles* being rated as rather not relevant. This



https://doi.org/10.1017/pds.2024.212 Published online by Cambridge University Press

makes sense since the challenges were retrieved from the SLR, it rather confirms the outcomes of the SLR as correct and relevant.

**Table 1. Operative challenges extracted from SLR and their rating in the survey**

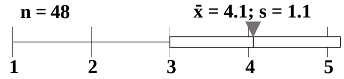
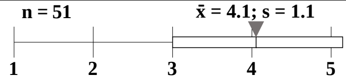
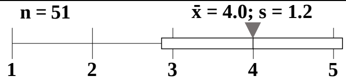
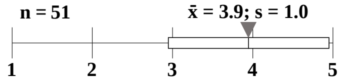
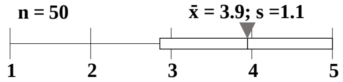
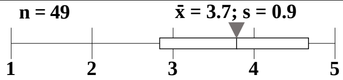
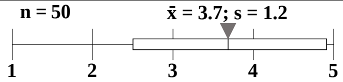
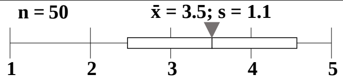
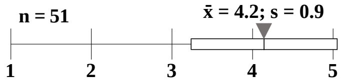
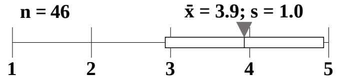
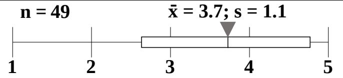
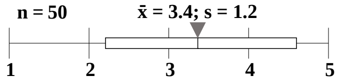
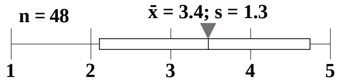
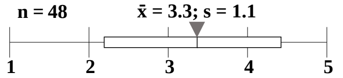
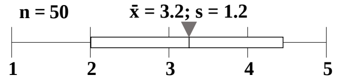
| # | Challenge | Rating of relevancy |
|---|---|---|
| 1 | Data silos / unknown data existence (*Which data is possibly available in my system?*) (Ebel et al., 2021) | n = 48; $\bar{x}$ = 4.1; s = 1.1  |
| 2 | Aggregation of data gathered from different sources (*How do I merge data, e.g. from different software systems?*) (Altavilla et al., 2017) | n = 51; $\bar{x}$ = 4.1; s = 1.1  |
| 3 | Missing understanding of the target setting (*What is the desired data used for?*) (Holmström Olsson et al., 2019) | n = 51; $\bar{x}$ = 4.0; s = 1.2  |
| 4 | Missing methodical support for data capturing (*What is the best way to proceed? What must be paid attention to?*) (Cantamessa et al., 2020) | n = 51; $\bar{x}$ = 3.9; s = 1.0  |
| 5 | Missing data assessment criteria (*Which data is suitable and which is not?*) (Jiang et al., 2021) | n = 50; $\bar{x}$ = 3.9; s =1.1  |
| 6 | Identifying suitable data to collect (*Which data is helpful for the present task?*) (Lachmayer and Mozgova, 2021) | n = 49; $\bar{x}$ = 3.7; s = 0.9  |
| 7 | Aggregating different data formats and structures (*How do I, e.g., combine qualitative and quantitative data or different file types?*) (Bertoni, 2018) | n = 50; $\bar{x}$ = 3.7; s = 1.2  |
| 8 | Uncertainty about when to collect which data (*At which stage of the lifecycle should what data be collected?*) (Machchhar and Bertoni, 2021) | n = 50; $\bar{x}$ = 3.5; s = 1.1  |

**Table 2. Data-related challenges extracted from SLR and their rating in the survey**

| # | Challenge | Rating of relevancy |
|---|---|---|
| 1 | Bad data quality (*E.g. completeness, correctness, credibility of the data source, timeliness*) (Sadiq et al., 2007) | n = 51; $\bar{x}$ = 4.2; s = 0.9  |
| 2 | Difficulties with the combination of data of different quality (*E.g. detailed data vs. rough data or a mix of high and low quality*) (Altavilla et al., 2017) | n = 46; $\bar{x}$ = 3.9; s = 1.0  |
| 3 | Missing transparency/traceability of data (*Where does data come from, and who changed it*) (Franch et al., 2020) | n = 49; $\bar{x}$ = 3.7; s = 1.1  |
| 4 | Too small amount of data (*E.g. to ensure statistical significance*) (Bach et al., 2017) | n = 50; $\bar{x}$ = 3.4; s = 1.2  |
| 5 | No available data (*E.g. for new development without predecessor model*) (Höhn et al., 2017) | n = 48; $\bar{x}$ = 3.4; s = 1.3  |
| 6 | Missing or limited access to data (*Cf. open-access data vs. confidential corporate data*) (Sadiq et al., 2007) | n = 48; $\bar{x}$ = 3.3; s = 1.1  |
| 7 | Too large amount of data (*resulting in overload to select the right data or lack of resources for storage and processing*) (Mehlstäubl et al., 2022) | n = 50; $\bar{x}$ = 3.2; s = 1.2  |



https://doi.org/10.1017/pds.2024.212 Published online by Cambridge University Press

Table 3. Technical challenges extracted from SLR and their rating in the survey

| # | Challenge | Rating of relevancy |
|---|---|---|
| 1 | Ensuring data security (*Protection against unauthorized access to data, also e.g. concerning intentional manipulation of data*) (Mehlstäubl et al., 2022) | n = 49; $\bar{x}$ = 3.7; s = 1.1 |
| 2 | Technical limitations to collect data (*E.g. products not equipped with required sensors*) (Gorkovenko et al., 2020) | n = 48; $\bar{x}$ = 3.5; s = 1.3 |
| 3 | Missing resources (*E.g. supporting software, computing power, capacities for storage of large amounts of data*) (Ebel et al., 2021) | n = 50; $\bar{x}$ = 3.0; s = 1.2 |

Table 4. Organizational challenges extracted from SLR and their rating in the survey

| # | Challenge | Rating of relevancy |
|---|---|---|
| 1 | Missing knowledge about data science processes (*How is data processed, and what needs to be paid attention to?*) (Relich, 2023) | n = 49; $\bar{x}$ = 3.7; s = 1.2 |
| 2 | Unwillingness of customers to share data (*E.g. through data privacy settings on devices*) (Franch et al., 2020) | n = 49; $\bar{x}$ = 3.7; s = 1.2 |
| 3 | Challenging interdisciplinary collaboration with data scientists (*E.g. different approaches, "speaking the same language"*) (Cantamessa et al., 2020) | n = 48; $\bar{x}$ = 3.5; s = 1.2 |
| 4 | Missing management support for data-driven processes or projects (*E.g. not approved but needed capacities*) (Briard et al., 2023) | n = 48; $\bar{x}$ = 3.5; s = 1.2 |
| 5 | Overload with data management and data governance (*Storage, maintenance, management of access rights & user roles*) (Altavilla et al., 2017) | n = 47; $\bar{x}$ = 3.2; s = 1.1 |

Table 5. Regulatory challenges extracted from SLR and their rating in the survey

| # | Challenge | Rating of relevancy |
|---|---|---|
| 1 | Restrictions resulting from data privacy laws (*What data are allowed to be gathered, processed, shared, and used?*) (Briard et al., 2023) | n = 48; $\bar{x}$ = 3.6; s = 1.2 |
| 2 | Missing standards (*E.g. on data formats, data exchange, etc.*) (Lachmayer and Mozgova, 2021) | n = 51; $\bar{x}$ = 3.6; s = 1.2 |
| 3 | Unclear data ownership (*Who (legally) owns certain data? May it be used?*) (Machchhar and Bertoni, 2021) | n = 47; $\bar{x}$ = 3.3; s = 1.3 |
| 4 | Missing/unprecise legal regulations regarding the digital law (*What is allowed, what is not?*) (Bauer et al., 2018) | n = 41; $\bar{x}$ = 3.2; s = 1.0 |
| 5 | Ethical hurdles (esp.) with regard to the utilization of customer data (*Does it have a negative impact on the company's image?*) (Gorkovenko et al., 2020) | n = 43; $\bar{x}$ = 2.5; s = 1.3 |

*Bad data quality* is rated as the most relevant challenge with an average score of 4.2. At the same time, *missing data assessment criteria* are rated high as well with 3.9 respectively. These two aspects are highly related to each other: if suitable criteria to assess data are missing, then bad data quality is a logical consequence. Further, the rating of *difficulties with the combination of data of different quality* with 3.9 can be related to these aspects: when capturing data without a suitable assessment of its quality,



https://doi.org/10.1017/pds.2024.212 Published online by Cambridge University Press

not only bad quality results but also a mixture of bad and good quality. One participant described a temporal dimension as well: when working with partners, a certain degree of data quality needs to be defined for fixed dates along the project to ensure every stakeholder can proceed with a usable dataset.
With a score of 4.1, both *data silos / unknown data existence* and *aggregation of data gathered from different sources* are rated as second most relevant. The former describes the missing knowledge about the present systems, their available data distributed over different silos, and subsequently not knowing where data is collected, stored, and processed within the system furthermore explains the rating of *missing transparency/traceability of data*. If the named aspects are incomprehensible, it is a logical consequence that it cannot be traceable where data is coming from and which processing steps it has already undergone. The latter can be viewed as a specific *missing knowledge about data science processes* along with *aggregating different data formats and structures*. Several participants stated that they are introducing in-service training for their employees to meet with this challenge, nevertheless will still always be dependent on working in an *interdisciplinary collaboration with data scientists* for data processing and the (graphic) preparation of the outcomes. Another connection can be drawn here: if designers do not know about the processes of data science, they do not have a good basic understanding of how data scientists think and work, making cooperation with them challenging. An interesting aspect was added here by several free-text answers: the challenge to identify suitable applications for including Artificial Intelligence in data capturing and how to implement it.
The *missing understanding of the target setting* follows with a rating of 4.0. While first having the technical perspective of selecting data unsuitable for the present goal in mind, the statement of two participants brings a second perspective to this aspect, which is the organizational and financial perspective. They stated in the comment section, that, because of the missing understanding of the target setting, simply "all" data is captured to prevent missing out on any relevant. This leads to a huge amount of data having to be managed and therefore related to higher costs. Further, they build a bridge to *identifying suitable data to collect* when stating that because of the large, un-targeted data amount, they are overburdened with this aspect.
Surprisingly, regulatory challenges are generally rated lower, as are the technical challenges except for ensuring data security. Interestingly, several of those participants rating the *restrictions resulting from data privacy laws* as extremely relevant stated that they give this assessment based on the culture in their company described as cautious towards data capturing resulting from a general fear of violating the regulations. This connects to challenges related to DDD in general and not only specific to data capturing addressing the required change management to introduce data-driven processes as it has been unveiled by literature such as Mehlstäubl et al. (2022). However, others giving a low rating at the same time shows how different companies are dealing with this. Another aspect is brought up by the statement of several participants: the issue of an undefined responsibility within the company for the collection of data. An example is given describing the service department having access to the products when they return for repair and maintenance work, however, it is not defined as their responsibility to collect the data from the product and therefore not being done, while the design department does not get in contact with the products once they left manufacturing and therefore is unable to capture data.

| Data selection | Systems analysis | Data Science knowledge | Targeted capturing |
|---|---|---|---|
| Assessment of data quality | Identification of data availability & traceability | Basis for interdisciplinary collaboration | Identification of suitable data |

**Figure 4. Identified areas of required support**

The *missing of a methodical support for data capturing* is rated as very relevant with an average of 3.9. The need for it is emphasized by many statements of participants stating that they need to either buy in external expertise or even fail projects because of the inability to overcome the issues. When reviewing the most relevant challenges described above, it can be concluded that within the research field of product development, methodical support is needed which must include several contents as displayed in Figure 4: at first, supporting data selection with some sort of assessment method to meet with the factors explained in the first paragraph regarding bad data quality. Secondly, providing approaches to analyze the present systems regarding their availability of data, where data is generated, captured, processed, and used to cope with the aspects related to data silos / unknown data existence as well as giving a foundation to assess the data quality. Thirdly, giving some sort of basic knowledge about data science processes, tools, and working methods to address the challenges resulting from missing knowledge about it. Fourthly, the method must guide its users to a targeted capturing as described above. Although the group of smaller companies (SC) is underrepresented among the participants, some observations in comparing their answers with those from bigger companies (BC; >10.000) at least indicate differences. E.g., the challenges of *data silos*, *limited access*, and *data security* are rated ~ 0.7 lower in SC. Naturally, these problems increase with increasing company size. In contrast, *missing standards* are rated 0.6 higher by SC, possibly explained by a BC often having its own internal standards.

## 5. Discussion and conclusion

The term DDD describes the utilization of knowledge extracted from the analysis of data to support and improve the design process. The continuously growing digitalization in all fields of engineering holds great potential to exploit more and new fields of data throughout the product lifecycle. Despite its great potential, DDD is facing several problems that need to be overcome. Previous work has shown that the initial step, the capturing of data, is one of the biggest present issues related to DDD. Therefore, the goal of this contribution was to investigate what concrete challenges of the capturing of data within a design process are (RQ 1) and what implications this poses onto product development (RQ 2). To answer these questions, a survey with industry practitioners based on an SLR was conducted.

The conducted survey delivered the view of the industry on challenges related to data capturing within DDD. The vast majority of participants dealing with this topic intensively in their work activities makes it a highly valuable contribution to the research field of DDD since it represents the view directly from the relevant practitioners. The conducted SLR can be evaluated as correct, since, except for only one, all challenges extracted from literature were at least evaluated as reasonably relevant.

However, there are some limitations to be discussed. With the participants mostly coming from large companies, the validity of the outcomes has to be limited to larger organizations. While most of the ratings were given similarly, some aspects were identified to be rated slightly differently. Further, the findings have to be restricted to the industry sections of vehicle manufacturing and mechanical engineering, since other industries are not sufficiently represented in the group of participants.

RQ 1 is answered by the given survey answers in Section 4. Four groups of challenges are concluded to be related to each other and are found to be most relevant to practitioners. These topics must be addressed in a methodical support which is currently missing, therefore answering RQ 2 by implicating that such a methodical support consisting of the aspects shown in Figure 4 must be developed:

1. Support data selection with an assessment method
2. Provide approaches to analyze the present systems regarding the availability of data, where it is generated, captured, processed, and used
3. Giving basic knowledge about data science processes, tools, and working methods
4. Guide users to a targeted capturing

This leads to the next steps that have to be taken. In the following, existing literature needs to be searched and reviewed to investigate which challenges are already addressed by supporting methods, which ones are not, and for which only limited support is available. Subsequently, a suitable method needs to be developed to provide practical support concerning all relevant challenges for data capturing within DDD. In addition, interviewing focus groups from SC to further investigate differences between them and BC as well as other industries to enlarge the validity range of the research are possible next steps.

## References

Altavilla, S., Montagna, F. and Newnes, L. B. (2017), "Interdisciplinary life cycle data analysis within a knowledge-based system for product cost estimation", *Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 5: Design for X, Design to X, Vancouver, Canada, 21.-25.08.2017*.

Bach, J., Langner, J., Otten, S., Holzapfel, M. and Sax, E. (2017), "Data-driven development, a complementing approach for automotive systems engineering", *2017 IEEE International Symposium on Systems Engineering, ISSE 2017 - Proceedings*. https://dx.doi.org/10.1109/SysEng.2017.8088295.

Bauer, D., Maurer, T. and Bauernhansl, T. (2018), "Datenbasierte Produktionsoptimierung Entwicklungsfelder für produzierende Unternehmen", *WT Werkstattstechnik*, vol. 108, no. 3, pp. 108–112.

Bertoni, A. (2018), “Role and Challenges of Data-Driven Design in the Product Innovation Process”, *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1107–1112. https://doi.org/10.1016/j.ifacol.2018.08.455.
Bertoni, A. (2020), “Data-driven Design in concept development: systematic review and missed opportunities’, *Proceedings of the Design Society: DESIGN Conference*, vol. 1. https://doi.org/10.1017/dsd.2020.4.
Briard, T., Jean, C., Aoussat, A. and Véron, P. (2023), “Challenges for data-driven design in early physical product design: A scientific and industrial perspective”, *Computers in Industry*, vol. 145, p. 103814. https://doi.org/10.1016/j.compind.2022.103814.
Briard, T., Jean, C., Aoussat, A., Véron, P., Le Cardinal, J. and Wartzack, S. (2021), “Data-driven design challenges in the early stages of the product development process”, *Proceedings of the Design Society*, vol. 1, pp. 851–860. https://doi.org/10.1017/pds.2021.85.
Cantamessa, M., Montagna, F., Altavilla, S. and Casagrande-Seretti, A. (2020), “Data-driven design: the new challenges of digitalization on product design and development”, *Design Science*, vol. 6. https://doi.org/10.1017/dsj.2020.25.
Chaudhari, A. M., Bilionis, I. and Panchal, J. H. (2020), “Descriptive Models of Sequential Decisions in Engineering Design: An Experimental Study”, *Journal of Mechanical Design*, vol. 142, no. 8. https://doi.org/10.1115/1.4045605.
Ebel, P., Orlovska, J., Hünemeyer, S., Wickman, C., Vogelsang, A. and Söderberg, R. (2021), “Automotive UX design and data-driven development: Narrowing the gap to support practitioners”, *Transportation Research Interdisciplinary Perspectives*, vol. 11. https://doi.org/10.1016/j.trip.2021.100455.
Franch, X., Seyff, N., Oriol, M., Fricker, S., Groher, I., Vierhauser, M., Wimmer, M., Madhavji N. and Pasquale, L. (2020), “Towards Integrating Data-Driven Requirements Engineering into the Software Development Process: A Vision Paper”, *Lecture Notes in Computer Science*, 12045 LNCS, pp. 135–142. https://doi.org/10.1007/978-3-030-44429-7_10.
Gerschütz, B., Schleich, B., Wartzack, S., Krause D., Paetzold K. and Wartzack S. (2021), “A semantic web approach for structuring data-driven methods in the product development process”, *Proceedings of the 32nd Symposium Design for X, DFX 2021.* https://doi.org/10.35199/dfx2021.15.
Gorkovenko, K., Burnett, D. J., Thorp, J. K., Richards, D. and Murray-Rust, D. (2020), “Exploring the Future of Data-Driven Product Design”, *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376560.
Hazelrigg, G. A. (1998), “A Framework for Decision-Based Engineering Design”, *Journal of Mechanical Design*, vol. 120, no. 4, pp. 653–658. https://doi.org/10.1115/1.2829328.
Höhn, M., Hollauer, C., Wilberg, J., Kammerl, D., Mörtl, M., Omer, M., Fadel G., Salustri F., Kim H., Skec S., Van der Loos M., Maier A.M., Kokkolaras M. and Oehmen J. (2017), “Investigating usage data support in development processes - A case study”, *Proceedings of the International Conference on Engineering Design, ICED*, vol. 7, DS87-7.
Holmström Olsson, H., Bosch, J., Franch X., Männistö T. and Martínez-Fernández S. (2019), “Data Driven Development: Challenges in Online, Embedded and On-Premise Software”, *Lecture Notes in Computer Science*, 11915 LNCS, pp. 515–527. https://doi.org/10.1007/978-3-030-35333-9_36.
Huber, W. (2018), *Industrie 4.0 kompakt – Wie Technologien unsere Wirtschaft und unsere Unternehmen verändern*, Wiesbaden, Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-20799-1.
Jiang, S., Hu, J. and Luo, J. (2021), “Data-driven design-by-analogy: State of the art”, *Proceedings of the ASME Design Engineering Technical Conference*, vol. 2. https://doi.org/10.1115/DETC2021-68669.
Kim, H. H. M., Liu, Y., Wang, C. C. L. and Wang, Y. (2017), “Special Issue: Data-Driven Design (D-3)”, *Journal of Mechanical Design, Transactions of the ASME*, vol. 139, 11, SI. https://doi.org/10.1115/1.4035002.
Lachmayer, R. and Mozgova, I. (2021), “Technical Inheritance as an Approach to Data-Driven Product Development”, *Design Methodology for Future Products: Data Driven, Agile and Flexible*. https://doi.org/10.1007/978-3-030-78368-6_3.
Machchhar, R. J. and Bertoni, A. (2021), “Data-driven design automation for product-service systems design: Framework and lessons learned from empirical studies”, *Proceedings of the Design Society*, vol. 1 [Online]. https://doi.org/10.1017/pds.2021.84.
Mehlstäubl, J., Gadzo, E., Atzberger, A. and Paetzold, K. (2022), “Herausforderungen datengetriebener Methoden in der Produktentwicklung”, *Konstruktion*, no. 06, pp. 60–66. https://doi.org/10.37544/0720-5953-2022-06.
North, K. (2021), *Wissensorientierte Unternehmensführung*, Wiesbaden, Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-32771-2.
Relich, M. (2023), “A Data-Driven Approach for Improving Sustainable Product Development”, *Sustainability (Switzerland)*, vol. 15, no. 8. https://doi.org/10.3390/su15086736.
Sadiq, S., Zhou, X. and Orlowska, M. (2007), “Data quality - The key success factor for data driven engineering”, *Proceedings - 2007 IFIP International Conference on Network and Parallel Computing Workshops, NPC 2007*. https://doi.org/10.1109/ICNPCW.2007.4351459.