Challenges in Data Science Project Management: A Case Study in a European OEM

Completed Research Full Paper

Christian Haertel
MRCC VLBA
christian.haertel@ovgu.de

Vincent Donat
MRCC VLBA
vincent.donat@ovgu.de

Daniel Staegemann
MRCC VLBA
daniel.staegemann@ovgu.de

Matthias Pohl
German Aerospace Center (DLR)
matthias.pohl@dlr.de

Klaus Turowski
MRCC VLBA
klaus.turowski@ovgu.de

Abstract

In today's world, the immense amount of data generated led to the emergence of Data Science (DS) as an integral discipline to extract knowledge and value from data. As DS projects suffer from high failure rates, new approaches for DS project management need to be developed and evaluated. In this paper, the Data Science Lifecycle (DSLC) process model is applied in a real-world DS undertaking at a European Original Equipment Manufacturer to support the initiation and planning of a project concerning the product development process. In this course, challenges in the application of the DSLC and the project at hand are identified and discussed. As this study only focuses on the initial Situation Assessment activity, in the future, the extension of the research to the remaining project phases and the inclusion of additional cases are needed to generalize the findings.

Keywords

Data Science, Project Management, Case Study, Data Science Lifecycle.

Introduction

The significance of data has manifested in a manner comparable to the role of oil in earlier industrial eras (Ransbotham et al. 2016). Combined with advanced capabilities for its utilization, data has become a cornerstone in various aspects of our daily lives. In this regard, Data Science (DS) aims to extract valuable insights from data that can be used to assist decision-makers in their strategic and operational decisions. Accordingly, this allows the redesign and optimization of existing processes which reveals previously unknown potentials. Furthermore, future market trends can be forecasted and synergies between different business areas or practices become visible (Medeiros et al. 2020). In conclusion, the use of DS technologies can positively impact operational productivity (Brynjolfsson et al. 2011; Müller et al. 2018), revenue growth (Troilo et al. 2016), and organizational agility (Côrte-Real et al. 2017). However, in DS project execution, stakeholders often face a variety of complex challenges, both technical and managerial. For instance, the identification of relevant data, as well as their quality and integrity, are crucial to enable reliable analyses and predictions. At the same time, legal aspects, such as security concepts for data storage, should not be ignored (Medeiros et al. 2020). Managerial hurdles include establishing new organizational structures, defining new roles and responsibilities related to data management and analysis, and fostering a "data-first mentality" (Roy 2018) to recognize and promote the value of data-based decisions (Medeiros et al. 2020). Additionally, the explorative nature of DS projects due to the data focus (Das et al. 2015) aggravates setting adequate expectations and establishing realistic timelines (Martinez et al. 2021). Given these challenges, DS projects often face a considerable failure rate (VentureBeat 2019). This fact underscores the need for careful planning, execution, and post-analysis of DS projects to realize the desired results. Several process models exist for the implementation of DS projects (e.g., CRISP-DM) to support DS projects in this regard. Nevertheless, research suggests inherent weaknesses in these models (Kutzias et al. 2021; Martinez et al. 2021; Saltz 2021). Based on these shortcomings, a recent study by Haertel et al. (2022) derived a Data Science Lifecycle (DSLC) model from best practices from existing DS methodologies. Yet, there are currently no evaluations and application examples for the DSLC present in the scientific literature. To bridge this gap, this study focuses on the practical utilization of the DSLC in a DS project at a European Original Equipment Manufacturer (OEM). Due to the extensive nature of the project under consideration, in this paper, we specifically focus on the execution and evaluation of the substage *Situation Assessment*, which is part of the *Business Understanding* phase within the DSLC. Consequently, the first research question (RQ) is addressed:

RQ1: How can the DSLC model support the initiation and planning of a project for data-driven product development at a globally operating European OEM?

As an OEM, the company is involved in the development of new products. Accordingly, the OEM aims to optimize this process using DS practices. The adoption of the DSLC in this project not only provides an initial practical application example for this process model but also offers a platform to discuss the specific challenges and issues that may arise during the implementation of DS projects, which is covered by the second RQ:

RQ2: What challenges arise in the initiation and planning of the DS project under review?

This RQ addresses challenges that are related to the application of the DSLC as well as general issues in the context of the DS undertaking. To answer the posed RQs adequately, the Case Study Research methodology (CSRM) of Eisenhardt (1989) will be applied in this paper and is described in the next section in conjunction with related work. After conveying the relevant theoretical fundamentals regarding DS and the DSLC model, the preparation and execution of the case study are outlined according to the adopted CSRM. Subsequently, the discussion of observed challenges follows and is contextualized with the literature. The work closes with a summary and an outlook on future research endeavors.

Methodology and Related Work

The CSRM is characterized by an in-depth examination of phenomena and dynamics within specific, real contexts or environments. It allows researchers to gain deep insights into specific situations or problems and use diverse data sources (Eisenhardt 1989). Typical methods that enable data collection in this context include interviews, analysis of archives or documents, and surveys of involved actors. The individual steps of the adopted CSRM by Eisenhardt (1989) are depicted in Figure 1. *Getting Started* refers to the definition of RQs to focus the efforts of the research. These were already introduced in the previous section. *Selecting Cases* is described as a crucial aspect of building theory from case studies (Eisenhardt 1989). One goal of this research is the application of the DSLC in real-world DS projects to assess its usefulness. While other authors, such as Kristoffersen et al. (2019), have already explored the application of DS process models (e.g., CRISP-DM) in OEMs, no comparable investigations exist for the DSLC (RQ 1). Since studies such as (Vicario and Coleman 2020) have demonstrated that DS can be effectively integrated into the operational processes of OEMs and provide benefits, the application of the DSLC in a DS project in an OEM is sensible. Although multiple studies have examined DS projects in OEMs, to the authors' best belief, no studies currently discuss the use of DS in product development processes. Additionally, in comparison to other publications, the paper at hand focuses explicitly on the early project activities concerning the initiation and planning of the undertaking and the encountered challenges (RQ 2). Nevertheless, to strengthen the findings, especially for the comprehensive evaluation of the DSLC, extending the population and considering all DS project phases will be required in the future. Next, *Crafting Instruments and Protocols* refers to how data is collected. Primary (qualitative) data sources are interviews and observations resulting from the authors' participation in the described project. For *Entering the Field*, sufficient induction with the business case and case study subject's processes are required. This is described in the fourth section.

Additionally, to overlap data analysis with data collection (Eisenhardt 1989), field notes are created during the familiarization phase and for the interviews. As only one case is examined in this research, *Analyzing the Data* merely includes within-case analysis based on the field notes from interviews and observations obtained in the DS project. In *Shaping Hypotheses,* insights from the previous step are utilized to identify the challenges related to the DS project and the application of the DSLC. Furthermore, to verify the hypothesized relationships, these are discussed with members of the OEM, which is described in the fifth section. Afterward, a comparison with extant literature (*Enfolding Literature*) on DS case studies and project management challenges is performed as "an essential feature of theory building" (Eisenhardt 1989). The scientific body of knowledge offers some studies that discuss DS project challenges and success factors (Haertel et al. 2023; Martinez et al. 2021). Accordingly, assessing whether the observations from practice align with the theory presents a valuable contribution. The last activity, *Reaching Closure*, includes a stop in adding cases. As the study only focuses on a singular case at this point, this final step is not applied in this article. In the future, an extension of the population quantity and depth (i.e., more DS project stages) is planned to improve the generalizability of the findings.

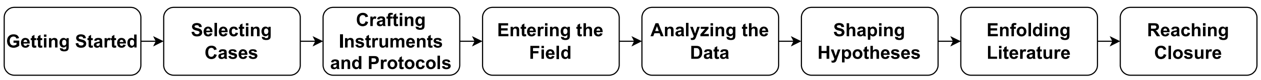

Figure 1. Applied Case Study Research methodology of Eisenhardt (1989)

Theoretical Background

As this work primarily deals with DS, specifically DS project management, a detailed explanation of the relevant terms is required. In addition to DS, related concepts will be defined. Thereafter, an insight into DS project management is provided. The DSLC model, which is essential for this study, will be presented.

Data Science

Historically, the roots of DS can be traced back to the disciplines of statistics and mathematics (Cao 2018). DS "comprises the use of scientific methods and techniques, to extract knowledge and value from large amounts of structured and/or unstructured data" (Martinez et al. 2021). DS refers to conducting data analysis as an empirical science, learning directly from the data itself (Chang and Grady 2019). DS is considered an interdisciplinary field. A typical team within a DS project consists of experts from various fields, including scientific fields such as mathematics, statistics, computer science, and specific domain knowledge relevant to where DS is conducted (Chang and Grady 2019). In addition to these capabilities, communication skills, natural curiosity, and good presentation skills are equally important in a DS project (Chang and Grady 2019). Because of the historical development and interdisciplinarity, several other concepts such as Data Analytics (DA) are related to DS. DA is characterized as "the theories, technologies, tools, and processes that enable an in-depth understanding and discovery of actionable insight into data" (Cao 2018) and is subdivided into descriptive, predictive, and prescriptive analyses. Data Mining (DM) is considered a subfield of DA and DS. Additionally, DS is closely connected to Big Data (BD) (Chang and Grady 2019), since traditional DM algorithms were no longer sufficient to handle the rapid increase in data volumes. BD systems developed for processing and handling large and complex data required new technical and analytical approaches to continue enabling meaningful analyses (Chang and Grady 2019). Other related terms to DS are, for example, information processing, exploratory data analysis, data-driven discovery, descriptive analytics, Knowledge Discovery in Databases (KDD), Business Intelligence (BI), or Machine Learning (Cao 2018). Because of conceptual similarities, various methodologies and approaches find application across these domains (e.g., process models).

Data Science Project Management

Achieving the goals in DS projects requires addressing multiple challenges that demand specialized tools, processes, and methods (Hesenius et al. 2019). On this note, project management describes "the application of knowledge, skills, tools, and techniques to project activities to meet the project requirements" (PMI 2017). As approaches from IT project management cannot be entirely transferred to DS projects due to their exploratory (Das et al. 2015) and data-intensive (Saltz et al. 2017) nature, DS process models are applied to help perform the typical activities of DS suitably (Haertel et al. 2022; Kutzias et al. 2021). A notable example is CRISP-DM but there are various other process models with different origins. In a comprehensive study, Martinez et al. (2021) examined a total of 19 process models for their strengths and weaknesses. Next to various challenges identified in DS project execution, the authors conclude the lack of integral methodologies for DS (Martinez et al. 2021). Consequently, based on the consolidation of best practices from 35 process models, the DSLC model was developed by Haertel et al. (2022). The DSLC consists of six broad stages (see Figure 2) and several subactivities that are detailed in an accompanying process map, including corresponding artifacts and functional roles.

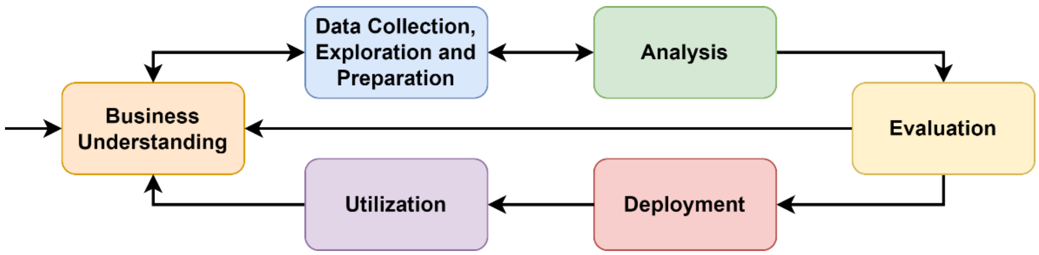

Figure 2. DSLC main phases, adapted from Haertel et al. (2022)

The DSLC model phases are worked on iteratively, allowing for potential backward steps depending on the phase (Haertel et al. 2022). In *Business Understanding*, the *Situation Assessment* entails identifying existing and required resources, developing requirements, and documenting risks and terminology. This can also include suitability tests or research on previous undertakings to obtain a more detailed picture of possible solutions, appropriate tools, and data. For the management and a required project proposal, these findings are used to better estimate the expected value of the DS project and derive success criteria as well as DS objectives. Furthermore, the project plan is defined, along with associated milestones and tasks (Haertel et al. 2022). *Data Collection, Exploration and Preparation* primarily involves data acquisition and subsequent steps like initial investigations and the resulting data preparation based on the preliminary insights. In *Analysis,* suitable analytical (ML) models are developed and technically evaluated to fulfill the initially documented DS targets. Afterward, *Evaluation* assesses the models concerning business goals and requirements. Also considering factors such as budget and time capacities, the next steps may involve refined test designs or models, project discontinuation, or the *Deployment* phase (Haertel et al. 2022). In this penultimate phase, various preparatory organizational activities and tests are initially carried out to adequately set up the actual implementation of the analytical artifacts in the productive environment. The *Utilization* phase refers to the actual use of implemented models, including supporting, monitoring, and maintenance activities. If the results no longer meet the requirements, a new DS project can be initiated (Haertel et al. 2022). In this paper, the application of the DSLC model in practice is described, focusing on the *Situation Assessment* task.

Case Study: Globally Operating European OEM

For anonymity protection and due to the complexity of the company in which the study was conducted, detailed insights into the organizational aspects of the company are not provided. The study focuses on a selected organizational unit (OU) within the company. The analyzed European OEM operates globally. The company has branches and subsidiary organizations on various continents, which act in different configurations and work collaboratively or autonomously. The OU plays a central, company-wide role with the mission of comprehensively supporting the entire OEM and its subsidiaries. The benefits of DS have already been recognized within the OU. Although the OEM already has a dedicated DS department, the OU independently initiates and implements DS projects. Based on diverse data analyses based on the datasets at its disposal, the OU significantly contributes to supporting an enterprise-wide Controlling and Reporting Tool (CRT). The CRT archives various historical datasets essential for the conducted data analyses. BI tools, actively maintained and supported by the OU, are employed for optimal visualization and presentation of the analysis results. The OU continually strives to explore new collaboration opportunities with other departments within the OEM, seeking and investigating shared data sources.

Business Case: Product Development Process

As an OEM, the corporation is involved in the development of new products. Manufacturing companies aim to conceptualize, produce, and distribute products that can thrive in the market, aiming to maximize corporate profits (Henning and Moeller 2011). Activities of the product development process (PDP) are complex and multifaceted. These include the initial development idea, technical design, or the final validation of the product to ensure its market viability and quality (Henning and Moeller 2011). These phases are embedded in an overarching process, often organized on a project basis. A characteristic feature of this process is the inherent uncertainty regarding technical specifications and market potential (Verworn 2005). The PDP is impacted by time and cost pressure, as no revenue can be generated until the new product is introduced to the market, while development activities already incur significant costs. Hence, minimizing the duration of the PDP and reducing the costs incurred during this phase is crucial (Henning and Moeller 2011). The early phases of the PDP are characterized as highly influential since decisions regarding product design, material selection, and process efficiency can have significant impacts on overall expenses (Ehrlenspiel et al. 2014). Additionally, the costs of changes increase exponentially the further the product advances in its lifecycle (Ehrlenspiel et al. 2014). Optimizing and properly designing the PDP, especially its early phases, is therefore essential for market success and competitiveness. The early phases involve activities that begin with the idea or conception of a new product and end with the actual initiation of product development (Verworn 2005). As the OEM subsidiaries sometimes operate autonomously and independently, the early phases of the PDP are conducted according to individual standards, processes, and procedures. This diversity leads to a range of challenges, including a lack of transparency, inefficient resource utilization, and difficulties in project management, which hinder the utilization of synergies. To address these issues, the OEM has taken the initiative to harmonize and standardize the PDP, specifically concerning its early phases, across the entire corporation. In this process, the OU plays a central role by both providing expertise to help integrate and standardize different processes effectively and utilizing the CRT to support and monitor the newly structured process. In addition, the OU aims to leverage the CRT for more data-driven and advanced analytical approaches to gain deeper insights into the data generated during the early phases of the PDP. By using these methods, the OU strives to recognize patterns, make forecasts, and thus promote more informed, data-driven decision-making throughout the PDP. Accordingly, the integration of DS practices into this process is a crucial step to make the PDP more efficient, forward-looking, and adaptable. Within this initiative, authors of this work are involved with developing a suitable concept for a data-driven PDP according to the stakeholder requirements. Hence, this includes not only the technical design and implementation of data analysis and visualization but also close collaboration with the domain users to capture expectations.

Preparation of the Case Study

The case study focuses on the initiation and planning of the DS project to harmonize the PDP. This corresponds to the *Situation Assessment* activity in the *Business Understanding* phase of the DSLC. In preparation for the investigations, extensive measures were taken to establish a solid foundation for the analysis. An in-depth understanding of internal processes, as well as the functionality and technical capacities of the CRT, which was intended to serve as the basis for subsequent data analyses, was worked out. This included the comprehension of the domain-specific language used within the corporation, as well as an overview of the complex organizational structures. Additionally, internal documentation within the corporation was examined for further insights into the early phases of the PDP. Another crucial step involved multiple consultations with the management of the OU. These discussions primarily focused on the anticipated benefits for the corporation, the methodological approach for the investigations, and the planned milestones in the conceptualization and implementation of data analyses. Moreover, meetings with the Business Analyst of the OU and the Requirements Manager of the CRT were employed to design interview questions to capture relevant information accurately. Based on the induction activities, a list of interview targets for requirements elicitation was created. This encompassed a wide range of project stakeholders, including domain experts and executives, to ensure a comprehensive and representative understanding of the project context and requirements.

Application of the DSLC Process Model

The *Business Understanding* phase of the DSLC marks the beginning of a DS undertaking. This phase primarily aims to develop a profound understanding of the business requirements and goals that should be achieved with the deployment of DS solutions (Haertel et al. 2022). As the initial activity, *Situation Assessment* involves steps to identify required resources and requirements, as well as to capture risks. Possible solution approaches are also explored in this subprocess. Therefore, comprehensive interviews were conducted with a variety of stakeholders and users to obtain a holistic picture of the needs, expectations, and challenges of different user groups. These user groups can be categorized into three main categories, each utilizing the CRT in different ways and therefore having specific requirements. *Value Engineers* extensively use the CRT for documentation and analysis of their results (e.g., calculations for procurement), considering it a central instrument for their daily tasks. Based on open tasks, *Lower Management* uses the CRT to optimize team performance and resource allocation. *Middle Management* primarily needs the CRT for capturing and analyzing key performance indicators (KPIs) for the precise comprehensive assessment and control of their business areas. **Fehler! Verweisquelle konnte nicht gefunden werden.** displays the sequence of the conducted tasks within the *Situation Assessment* substage in this project (RQ 1). In the first step, a total of eight interviews were performed to gain a profound understanding of the domain matter. Some interviews took place in multiple, spaced-out sessions because of the high information density which required more time for thorough interview execution. Furthermore, the opportunity to revisit complex topics in later sessions provided significant benefits. The interview questions were categorized into three main groups, namely *General*, *Workflow*, and *Project*. *General* focuses on questions aimed at developing an understanding of the early phases of the PDP from the interviewee's perspective. The *Workflow* category concentrates on the specific work steps in the PDP that the interviewee performs. Special emphasis is put on relevant data, employed tools, and necessary coordination processes. Finally, *Project* is dedicated to the assessment of the overall project. These questions are designed to capture the perception of the project and provide insight into the objectives and challenges.

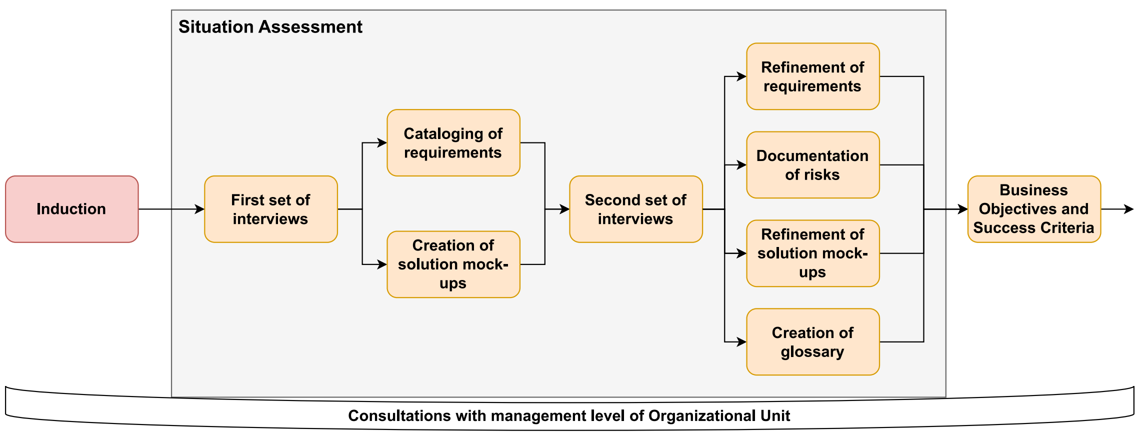

Figure 3. Conducted activities of Situation Assessment (RQ 1)

After conducting interviews, initial drafts for data analyses, visualizations, and reports for the PDP were initiated. These mock-ups were directly based on conversations with stakeholders and the information and insights they provided. Thereafter, a second round of discussions with selected stakeholders was initiated. In these conversations, the developed solution concepts were presented, and feedback was collected. The intention behind this approach was to gather early impressions and reactions to the work while ensuring that key points and stakeholder requirements were accurately captured and understood. This iterative approach allowed for continuous refinement of the work, ensuring that the results met the expectations and needs of the stakeholders.

Findings

Based on the interviews, this subsection summarizes the results of the analysis of the current situation across the three categories of questions. For *General*, it became evident that each new product is considered and treated as an independent project in the PDP. The roles and responsibilities of the interviewees naturally varied. Typical activities include evaluating products regarding their costs and technical features, as well as developing new ideas and approaches to create added value. In the context of collaboration, the finance departments and the units for technical development were particularly mentioned. These departments are often actively involved in the feedback process, especially in the evaluation and development of ideas and approaches. Furthermore, the importance of the early phases of the PDP was equally recognized by all respondents. For *Workflow*, the responses of the participants varied due to the different strategies across the different corporate entities. A common activity was the creation of cost indications, which involved estimating and evaluating the value of a component. Additionally, systematic comparisons with other products (benchmarks) on the market were executed to set performance standards and identify potential improvement opportunities or find alternative approaches, such as for material selection. Notably, one corporate subsidiary already established a comprehensive and standardized process to structure and streamline the early phases of the PDP, providing a significant competitive advantage. This specific company also stands out with a database where historical calculations are stored to enable future projects by providing reference values and forming a solid data foundation for planning and analysis. In contrast, discussions with stakeholders of other subsidiaries revealed that comparable systems did not exist for them and office administration tools are primarily utilized for controlling, documenting, and storing project-related data as well as for the creation of reports and result sheets. Due to the diverse structures and different activities within the OEM, the data required for these described processes exhibit considerable variety. In particular, financial data regarding components play a critical role in determining cost ceilings and include cost estimates from technical development, forming essential input for decision-making and planning. Specific data about individual components, such as required material(s), are also of great importance for categorization. Repeatedly, two KPIs were described as benchmarks for performance evaluation. The first concerns the discrepancy between the financial target and reports from technical development teams. A specific percentage threshold is set within which deviations between planned and actual costs are allowed. One goal of the early phase of the PDP is to minimize this gap. The second KPI relates to the penetration rate, which indicates the percentage of components of a product being subject to a calculation by value engineers. A specific percentage target is also defined for this KPI. During the discussions, critical views regarding the effectiveness of these KPIs were expressed, as they appeared to rather rely on the intuition of the management than on solid, data-driven analyses. In this context, the desire for suitable KPIs or requirements based on insights from the data was voiced. Furthermore, interviewees expressed the lack of a unified data repository and specialized tools leading to increased use of general office software products, which are not optimally aligned with the specific requirements of process control. Dependence on such tools requires significant manual effort, which is not only time-consuming but also susceptible to errors. Many participants wished for more transparency in workflows within other departments or corporate entities to uncover potentially untapped synergies. Another important aspect raised is the need for increased collaboration between different departments. To achieve this and improve the efficiency and quality of work processes, the introduction of a unified process in the PDP is considered crucial to facilitate coordination and cooperation between departments and help overcome current challenges in data management and process control. In this context, some participants expressed the expectation that insights gained from analyses and evaluations of collected data could contribute to improving the PDP.

For the *Project*-related questions, particularly emphasized challenges were the process complexity and diversity between various corporate entities. There was agreement that the project would be considered successful if it led to the establishment of a utilized and refined PDP. Such a process should be characterized by the unified management of data and enable efficient control of early PDP phases within the OEM and its entities. Furthermore, creating synergies and the need for more transparency were emphasized as essential goals. One participant specifically mentioned creating a corporation-wide database that includes calculations for all possible projects to allow precise analyses to support future undertakings. The realization of these goals is considered crucial to improve collaboration between OEM entities and ensure efficient resource utilization. Overall, these discussion points indicate a growing awareness of the importance of effective process design and data management within the corporation, seen as keys to future success in the early phase of the PDP. After the second set of interviews, sufficient information was collected to continue with the remaining tasks of *Situation Assessment*. Initial drafts of requirements and solution mock-ups were revisited and refined based on the feedback. Additionally, the extracted obstacles and challenges were documented as risks. Finally, domain-specific knowledge was captured in a glossary to establish a common understanding for project members. With these activities being completed, the findings can be utilized to perform the definition of *Business Objectives and Success Criteria*, which corresponds to the consequent substage of *Business Understanding* in the DSLC.

Discussion

The conducted activities during *Situation Assessment* have revealed a series of issues that can attributed to challenges in the context of the DS project and problems in the application of the DSLC. These were further discussed with a Business Analyst (BA) and the Data Science Lead (DSL) of the OEM. The results are summarized in this section, corresponding to the focus of RQ 2. A comparison with the typical challenges in data-intensive projects according to the literature is also conducted, which refers to the *Enfolding Literature* step in the adopted CSRM.

DSLC Implementation Challenges

Due to the explorative nature of DS projects, the execution of the various phases and subprocesses of the DSLC relies on an iterative approach. This did not necessarily align with the working procedures of the OU's management. A definitive schedule with start and end points was expected. Deviations due to revisiting an activity or returning to previous steps were not envisioned. This discrepancy led to challenges in project implementation. Literature suggests that establishing an adequate corporate and work culture is one of the central challenges in DS projects (Li et al. 2022; Nakayama et al. 2017). This specific issue appears to be a typical example of such adaptation hurdles. Additionally, this observation hints at the difficulty of establishing realistic expectations for DS in terms of timelines (Martinez et al. 2021). Here, the BA emphasized the importance of communication skills to raise awareness for this issue due to the explorative character of DS in regular consultations with stakeholders. This aspect is also mentioned in the literature (Chang and Grady 2019). The application of the DSLC was further complicated by a partial lack of willingness to follow DS-specific approaches within the OEM. Certain activities of the process, such as decisions about data acquisition and the design of data infrastructure, were already pre-decided by other groups before the learnings of *Situation Assessment*, indicating a low level of process maturity (Martinez et al. 2021). The DSL confirmed that DS process models were not utilized for such undertakings in the OEM. Instead, ad-hoc approaches or methods from traditional IT project management were relied on. For the latter, the limited applicability for DS projects was admitted because of extremely complex and detailed guidelines. Nevertheless, the DSL stressed that some elements of *Situation Assessment* were already executed to some capacity in previous projects. This includes iterative discussions with stakeholders to understand user requirements and grasp all nuances of a process next to the technical perspective, indicating the importance of the interdisciplinarity aspect. While the application of DSLC generally proved to help plan the DS project, the outlined issues need to be accounted for to maximize its effectiveness for supporting DS undertakings.

General DS Project Challenges

A fundamental challenge pertains to the complexity of the subject of investigation. Despite acquiring in-depth insights into the OEM's work processes, relevant topics, and terminology in the preparation, capturing all relevant content during *Situation Assessment* proved to be demanding. The workflows of the globally operating OEM involved a multitude of stakeholders. Furthermore, both BA and DSL outlined sluggish organizational structures as a reason for unnecessarily prolonged processes. DS projects are interdisciplinary. Consequently, understanding domain-specific knowledge is crucial in DS (Nakayama et al. 2017; Para et al. 2018), and multidisciplinary teams are required (Martinez et al. 2021). In conjunction with the large number of involved departments in the project, team management becomes a significant challenge. Moreover, various stakeholders responsible for providing crucial data were not integrated into the data analysis process, which especially constituted a challenge when each department employed different strategies for data collection and storage. Consequently, this increased the difficulty of adequately assessing and utilizing the provided data. From a technical perspective, the CRT proved to be inadequate for storing and, especially, analyzing large data volumes. While sufficient data for effective analyses was not available in the first place, the lack of corresponding interfaces to other systems of the OEM exacerbated the integration of further data sources. According to the literature, data access and the absence of suitable tools and IT infrastructure for conducting analyses are highly relevant challenges in DS projects (Leone

2023; Martinez et al. 2021), especially in complex corporate environments. Some but not all stakeholders showed little interest in providing additional data due to concerns about slowing down processes by capturing more data. This collaboration issue might be a sign of a missing "data-first mentality" (Roy 2018) due to a lack of understanding of the benefits DS offers for the PDP. This increases the danger of the project results not being used (Martinez et al. 2021), which equals project failure. It became clear that closer collaboration between stakeholders and analytics teams is necessary to improve the quality and availability of data to enable more effective data analyses. The discussed challenges are summarized in Table 1 and grouped according to the categorization of Martinez et al. (2021).

Project Management	**Team Management**	**Data & Information Management**
Complexity of project and organization	Lack of collaboration between departments	Insufficient infrastructure for data storage and analysis
Low level of process maturity	Lack of DS-supportive work culture	lacking data volume, quality, and understanding due to non-participation of stakeholders
Inadequate project timeline expectations	Project interdisciplinarity	Absence of interfaces for data collection and integration

Table 1. Encountered challenges during the initiation and planning of the DS project (RQ 2)

Conclusion and Future Research Directions

DS has gained significant popularity due to the potential for improvements to business performance (Müller et al. 2018). However, DS projects are often not successful, which indicates the need for new and revised approaches (e.g., process models) (Saltz and Krasteva 2022). In this paper, the recently developed DSLC model was demonstrated in a case study in a European OEM, following the methodology of Eisenhardt (1989) to allow for a systematic and reproducible procedure. The research specifically addressed how the DSLC's *Situation Assessment* activity can be conducted to support the initiation and planning of a DS project (RQ 1). Furthermore, encountered challenges in the project and the application of the DSLC were discussed and contextualized (RQ 2). The observed obstacles in project, team, and data and information management need to be addressed in the further course of the undertaking to enable a successful outcome. Accordingly, this study is subject to limitations. Since only a specific project phase was investigated in this article, further research is needed to review the application of the remaining DSLC guidelines. Moreover, extending the case population to organizations with different characteristics (e.g., size, industry sector, specific conditions) is required to allow for cross-case analysis (Eisenhardt 1989), enhance the generalizability of statements concerning the usefulness of the DSLC model, and assess the transferability and impact of the findings to DS projects in other firms. In the future, these lessons learned will enable focused revisions of the DSLC model. Finally, while the identified DS project challenges aligned with the literature, adding further cases is necessary to also achieve confirmation in this regard and derive suggestions and success factors to overcome the commonly observed obstacles in DS undertakings.

REFERENCES

Brynjolfsson, E., Hitt, L. M., and Kim, H. H. 2011. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" *SSRN Electronic Journal* (doi: 10.2139/ssrn.1819486).

Cao, L. 2018. "Data Science," *ACM Computing Surveys* (50:3), pp. 1-42 (doi: 10.1145/3076253).

Chang, W., and Grady, N. 2019. "NIST Big Data Interoperability Framework," Gaithersburg, MD: NIST Big Data Public Working Group, Definitions and Taxonomies Subgroup.

Côrte-Real, N., Oliveira, T., and Ruivo, P. 2017. "Assessing business value of Big Data Analytics in European firms," *Journal of Business Research* (70), pp. 379-390 (doi: 10.1016/j.jbusres.2016.08.011).

Das, M., Cui, R., Campbell, D. R., Agrawal, G., and Ramnath, R. 2015. "Towards methods for systematic research on big data," in *IEEE Big Data 2015*, pp. 2072-2081 (doi: 10.1109/BigData.2015.7363989).

Ehrlenspiel, K., Kiewert, A., Lindemann, U., and Mörtl, M. 2014. *Kostengünstig Entwickeln und Konstruieren*, Berlin, Heidelberg: Springer Berlin Heidelberg.

Eisenhardt, K. M. 1989. “Building Theories from Case Study Research,” *The Academy of Management Review* (14:4), p. 532 (doi: 10.2307/258557).
Haertel, C., Daase, C., Staegemann, D., Nahhas, A., Pohl, M., and Turowski, K. 2023. “Toward Standardization and Automation of Data Science Projects: MLOps and Cloud Computing as Facilitators,” *IC3K2023*, pp. 294-302.
Haertel, C., Pohl, M., Nahhas, A., Staegemann, D., and Turowski, K. 2022. “Toward A Lifecycle for Data Science: A Literature Review of Data Science Process Models,” *PACIS 2022 Proceedings*.
Henning, F., and Moeller, E. (eds.). 2011. *Handbuch Leichtbau*: *Methoden, Werkstoffe, Fertigung*, München: Hanser.
Hesenius, M., Schwenzfeier, N., Meyer, O., Koop, W., and Gruhn, V. 2019. “Towards a Software Engineering Process for Developing Data-Driven Applications,” in *2019 IEEE/ACM 7th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering*.
Kristoffersen, E., Aremu, O. O., Blomsma, F., Mikalef, P., and Li, J. 2019. “Exploring the Relationship Between Data Science and Circular Economy: An Enhanced CRISP-DM Process Model,” *18th Conference on e-Business, e-Services and e-Society (I3E)* (11701), pp. 177-189.
Kutzias, D., Dukino, C., and Kett, H. 2021. “Towards a Continuous Process Model for Data Science Projects,” in *Advances in the Human Side of Service Engineering*, pp. 204-210.
Leone, M. 2023. “Decoding the Data Universe: The State of Data Science and Machine Learning,” Enterprise Strategy Group (ed.).
Li, Y., Zhai, L., Yang, C., and Liu, K. 2022. “Research on the Countermeasures of Digital Intelligence Transformation and Upgrading of Innovative Manufacturing Enterprises,” in *Business Intelligence and Information Technology*, pp. 106-115.
Martinez, I., Viles, E., and G. Olaizola, I. 2021. “Data Science Methodologies: Current Challenges and Future Approaches,” *Big Data Research* (24), p. 100183 (doi: 10.1016/j.bdr.2020.100183).
Medeiros, M. M. de, Hoppen, N., and Maçada, A. C. G. 2020. “Data science for business: benefits, challenges and opportunities,” *The Bottom Line* (33:2), pp. 149-163.
Müller, O., Fay, M., and vom Brocke, J. 2018. “The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics,” *Journal of Management Information Systems* (35:2), pp. 488-509 (doi: 10.1080/07421222.2018.1451955).
Nakayama, M., Olbrich, S., and Sutcliffe, N. 2017. “Changing with Grassroots Business Intelligence at a Large Global Manufacturing Firm,” in *Perspectives in Business Informatics Research*: *BIR 2017*, pp. 148-156 (doi: 10.1007/978-3-319-64930-6_11).
Para, J., Del Ser, J., Aguirre, A., and Nebro, A. J. 2018. “Decision Making in Industry 4.0 Scenarios Supported by Imbalanced Data Classification,” in *Intelligent Distributed Computing XII*, pp. 121-134.
PMI. 2017. *A guide to the project management body of knowledge (PMBOK guide)*.
Ransbotham, S., Kiron, D., and Prentice, P. 2016. “Beyond the hype: The hard work behind analytics success,” *MIT Sloan Management Review* (57), p. 6.
Roy, R. 2018. *How to build a data-first culture for a digital transformation*, accessed on Feb 16 2024.
Saltz, J., Shamshurin, I., and Crowston, K. 2017. “Comparing Data Science Project Management Methodologies via a Controlled Experiment,” in *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)*, Hawaii International Conference on System Sciences.
Saltz, J. S. 2021. “CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps,” in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 2337-2344.
Saltz, J. S., and Krasteva, I. 2022. “Current approaches for executing big data science projects - a systematic literature review,” *PeerJ Computer Science* (8:e862).
Troilo, M., Bouchet, A., Urban, T. L., and Sutton, W. A. 2016. “Perception, reality, and the adoption of business analytics: Evidence from North American professional sport organizations,” *Omega* (59), pp. 72-83 (doi: 10.1016/j.omega.2015.05.011).
VentureBeat. 2019. “Why do 87% of data science projects never make it into production?” available at https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/, accessed on Feb 29 2024.
Verworn, B. 2005. *Die frühen Phasen der Produktentwicklung*: *Eine empirische Analyse in der Mess-, Steuer- und Regelungstechnik*, Wiesbaden: Dt. Univ.-Verl.
Vicario, G., and Coleman, S. 2020. “A review of data science in business and industry and a future view,” *Applied Stochastic Models in Business and Industry* (36:1), pp. 6-18 (doi: 10.1002/asmb.2488).