Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 130 (2024) 1454–1461

www.elsevier.com/locate/procedia

57th CIRP Conference on Manufacturing Systems 2024 (CMS 2024)

On the Current State of Industrial Data Science: Challenges, Best Practices, and Future Directions

Sebastian von Enzberg[a,b,*], Julian Weller[a], Jonathan Brock[a], Silke Merkelbach[a], Melina Panzner[a], Jonas Lick[a], Arno Kühn[a], Roman Dumitrescu[a,c]

[a]*Fraunhofer Institute for Mechatronic Systems Design, Zukunftsmeile 1, 33102 Paderborn, Germany*
[b]*Hochschule Magdeburg-Stendal, Breitscheidstr. 2, 39114 Magdeburg, Germany*
[c]*Chair for Advanced Systems Engineering, Heinz Nixdorf Institute, Fürstenallee 11, 33102 Paderborn, Germany*

* Corresponding author. Tel.: +49-391-886-4472; fax: +49-391-886-4126. *E-mail address:* sebastian.von.enzberg@h2.de

Abstract

Data science provides organizations with the capability to draw insights from large amounts of data. Numerous sectors such as banking and marketing leverage this potential already. In comparison, industrial enterprises are slow in creating profit from their data. This originates from various factors, such as availability and quality of data, reliability of solutions, or a lack of organizational embedding. Traditionally, Industrial Data Science (IDS) is referred to as the application of data analytics within industrial settings. In this paper, we expand this definition, positioning IDS as an interdisciplinary field to encompass the holistic integration and organization of data science efforts into industrial enterprises. We provide three contributions. First, we conduct a thorough literature review on current issues and the state of the art of implementing IDS. Second, we conduct eight interviews with practitioners from different industrial enterprises to identify best practices and research gaps. Third, we condense our findings into a single framework which structures multiple layers of IDS in real-world organizations, such as projects, organizational embedding, and roles. Our findings can lead practitioners and academics in their quest to streamline their IDS efforts, guiding them into structured application and further research directions.

Keywords: Manufacturing Analytics; Data Analysis; Smart Factory; Industrial Analytics; CRISP-DM; Best Practices; Data Analytics Management; Portfolio Management

1. Introduction

In recent years, data science has been proven as a powerful tool across various sectors, with applications like data-driven decision making or process optimization and automation. In context of the digital transformation, data science allows enterprises to create value with the help of data [1]. While sectors, such as finance or marketing, have rapidly embraced data-driven strategies, industrial enterprises have been comparatively slow in leveraging this potential [2]. Reasons for this are manifold and range from issues in data infrastructure and data governance, over insufficient skills and acceptance, to lack of support in management and organization [3].

Industrial Data Science (IDS) has been defined as the use of Data Analytics, Machine Learning and similar analytical tools in industrial settings [4–7]. IDS is also seen as a intersectional discipline between computer science, statistics and engineering [8–11,1]. While the latter point of view acknowledges the scope of IDS as a field on its own, it is still limited to the mere application of data analytics in industrial settings, not fully addressing the aforementioned challenges that stem from the implementation of IDS applications. This underscores the need for a more comprehensive, holistic approach to IDS.

10.1016/j.procir.2024.10.266

We define IDS as an interdisciplinary field that encompasses the study, design and implementation of data science methodologies across the facets of industrial operations. Its application context are industrial enterprises, as which we consider organizations that mostly rely on manufacturing and product generation processes for value creation. The result of the implementation process of an IDS use case is a working IDS application, that is successfully embedded into a real-world organization by using tools and methodologies (IDS solutions).

Our definition includes the technological understanding (view on the technical implementation of industrial data science) of the previously cited definitions and also considers broader aspects as people, processes, infrastructure, and culture, as proposed e.g. by socio-technical systems theory [12,13].

Given this definition, this paper aims at reviewing the current state of IDS. Our main contributions are:

1) a structured review on current issues and state of the art of IDS solutions from a literature review,
2) the deduction of best practices and research gaps from an interview series with practitioners from industrial enterprises, and
3) a holistic framework that structures implementation aspects of IDS in organizations

The remainder of the paper is organized as follows: In section 2 we conduct a literature review resulting in a review of current challenges as well as practical tools and methodologies which are comprised in the state of the art. In section 3 we describe the research design for this paper. In section 4, we present the combined results of the conducted literature review and interviews, giving a structured summary of challenges, best practices, and future research directions for IDS. Section 5 presents a framework for the management of IDS and section 6 concludes our paper.

2. Industrial Data Science in the literature

In the following, current challenges (see section 2.1) and methodological and artifact-based approaches (see section 2.2) regarding industrial data science in practice are presented. The provided overview is grounded in an extensive literature search and our years of experience in the field. We primarily show applicable, validated knowledge from the IDS research area. We researched on existing literature focusing on industrial data science use cases, overarching concepts, frameworks, and other literature studies.

2.1. Current challenges

As shown, a plethora of authors use the term industrial data science and its synonyms. Even though there is not a clear definition of what IDS includes, most authors gather best practices, research questions and foremost challenges in the ever-emerging field of IDS. In the following, we propose a structuring approach for these challenges based on the data analytics canvas [9,11] and the commonly known structure of human, technology and organizational aspects [13]. We use the analytics canvas layers (use case, analytics, data pools, data sources) to detail the dimension technology and further enrich it by also analyzing human and organizational driven aspects to emphasize a holistic view on industrial data science (see Table. 1). The layers serve as a horizontal view on interdisciplinary competencies needed to practice efficient industrial data science. The vertical layer of human and organization serve as a setup and foundation for success [4]. In the following, we refer to this approach as the extended analytics layer model (EALM). The dimension of analytics and data science is usually structured by the 4 levels introduced by Gartner as a de facto standard (descriptive, diagnostic, predictive, prescriptive) [14].

Based on this overview, the results from the literature analysis where structured into challenges; a brief explanation of the challenges is gathered in Table 1. In the dimension of use cases, most authors refer to a lack of a holistic view on implementing analytics as well as problems in transferability and scalability of once implemented use cases. Analytics implementation approaches usually lack a focus on prescribing actions [15]. Implementation techniques are still under constant development and are challenged by industrial data [16]. The availability of data from an industrial setting suffers from them being scattered across the companies' data infrastructure [16].

Table 1 Overview of different IDS challenges according to the layers of the EALM.

	Challenge	Explanation
Use Case	Different data distributions for different operating conditions and machines decrease model performance and **transferability**. [17]	• Depending on the analyzed machine, different operating conditions can cause different distributions. This makes it hard to apply models trained for one condition or one machine to other conditions or machines.
	Use cases are usually implemented as singular case studies and are **mostly lacking the overall smart factory perspective**. [18]	• Interconnected value generation through connected use cases is still hard to achieve. No strategic planning or connection to data strategy when implemented. • Silo Installations or just focus on lighthouse implementations did not create the expected business value [19,18] • Use cases often do not scale without a high investment into the right infrastructure [20]
Analytics	Descriptive: Most implementations focus on **dashboarding** and visualization. [20]	• De Facto standard for most production companies nowadays. A focus still lays on the accessibility for end-users and non-analytics-experts. The actual decision that should be made based on a description is often not included in the use case implementation approach [15].
	Diagnostic: Most algorithms are available only for theoretical problems. [21]	• Many diagnosis algorithms exist, most are applied in theory and not suitable for time series as they occur in production.
	Prescriptive: Still immature technology [22] and complex in nature, thus difficult to manage [16].	• Immature algorithms and the overall complexity of automating decisions is making an efficient deployment of prescriptive analytics hard to achieve. • Many different approaches can be used to create a prescription [22]

Data Pool	Not enough **variety in data**. [23]	• In industry, machines usually work in normal mode. Data of faulty conditions is hard to obtain which results in a low variety of the data and makes it hard to train models for fault detection, prediction, or diagnosis.
	Lack of sufficient (labeled) data from the system under observation. [24,25]	• In the context of industry, the Internet of Things (IoT), and cyber-physical systems, literature often discusses Big Data [26,27], which necessitates new and advanced processes. However, in practical settings, companies frequently lament a data deficit. [25] • This issue arises because many field-deployed products lack sufficient sensors, or data access is impeded by inadequate network connectivity. Even when substantial data volumes are available, there's often a shortage of annotated data, or only a limited amount is accessible, as data annotation is costly and time-consuming [28].
	System Descriptions, which are needed for **diagnosis** tasks, are hard to obtain. [21]	• For Diagnosis, system descriptions that model the behavior of the system are needed. Those models are not available as a data basis in most industrial use cases and need to be created manually.
	Industry data is very diverse in structure, dimensionality and quality. [29]	• Industrial data can be classified according to very different characteristics. On the one hand, the sources can be viewed in terms of their origins in the company, i.e., from an organizational and domain-driven perspective. For example, they are generated in different business areas such as marketing and development and along different steps of the product life cycle.
	Data about decisions is subject to a high degree of uncertainty. [30]	• Decision Data is still not utilized to its full potential. Often, data about decisions is either not formalized or not analyzed.
Data Source	**Not enough Data** Points or varying data quality. [16]	• Industry data often lacks the necessary features for the given task or information about data quality is missing. This highly increases the investment cost (for data cleaning and data gathering) of use case related important industry data. Data is still managed in between quantity, quality, and investment cost. A sweet spot is hard to find. [31]
	Data for Prescriptive Analytics is either not documented or part of semi structured expert documents. [32,30]	• Data is often available, one of the main challenges is their accessibility and a lack of an overview about the data. [32,30] • The accessibility of the quality of a decision is an existing prescriptive analytics challenge. [32]
People & Organization	**Collaboration** between data scientists and domain experts is time-consuming. [33]	• Multi Domain Problems require multi domain teams. Different views on the same use case require practitioners to collaborate. This does require extensive time commitments for collaboration
	Multi domain perspective needed for Data Science in Industrial Applications.	• Collaboration between domain experts and data scientists can be challenging due to different prior knowledge and experiences [34]. • A ’citizen data scientist’ was defined by Gartner as ’a person who creates or generates models that use advanced diagnostic analytics or predictive and prescriptive capabilities, but whose primary job function is outside the field of statistics and analytics’[35]. Not enough citizen data scientists are present in industry.
	Craftwork does not equal mechanical work. Often times there a dissonance appears between how data scientists approach a project and how management plans it to be approached. [36]	• Data Scientists work in iterative cycles. Management expects data scientists to follow a typical waterfall like project management approach. This often leads to tension between different stakeholders inside the project.
	Lack of (industrial) data science **experts** [8]	• Many companies are suffering from a shortage of skilled workers. Small and medium-sized companies in particular lack qualified employees with data science skills [37]. In the industrial context, or industrial data science, the shortage of data scientists is even more severe, according to Bauer et al. because this field requires a combination of skills from computer science, statistics, and engineering [8]
	Industry 4.0 implementation restraints [38] and **missing maturity** of digitalization efforts.	• Digitalization efforts are on the rise but most companies still need to further digitalize their processes and systems to enable data analytics. [20]

As a general framework for the development, standard models like CRISP-DM show the relevant process steps of data analytics projects: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Yet, these models require a great deal of expert knowledge and methodological experience. Hence, we report on practical tools and methods in IDS in the subsequent section.

2.2. Practical Tools and Methods for Industrial Data Science

As shown in the previous section, conducting data analytics projects in industrial settings involves various distinct challenges. Consequently, various authors provide different solutions to overcome these challenges. Table 2 provides a brief overview of practical, IDS-oriented solutions, structured according to the CRISP-DM-lifecycle. We chose to structure the tools and methods according to the CRISP-DM-lifecycle, because the procedure model covers all the essential tasks in any data related project. As some of the solutions (e.g., maturity models) address challenges more holistically, we additionally add a ‘General’ column, where we cluster such approaches. The Table 2 is not intended to be complete. We primarily show examples of applicable, validated knowledge from the IDS research area.

A vivid example for such applicable knowledge is the data map proposed by Joppen et al. [39]. It is used within the data understanding phase to systematically map all possible data sources and generated information objects along the business processes. The workshop method thereby allows for data and business experts to systematically communicate and identify existing data, that can be leveraged with IDS.

Regarding the overall conduction of IDS projects, Huber et al. [40] showcase that the conventional CRISP-DM procedure model does not sufficiently cover problems commonly faced in IDS projects. Consequently, they suggest including tasks such as analyzing the system structure, technical test setup, or the selection of adequate software infrastructure.

Zooming out even further, Korsten et al. [41] identify a maturity model for advanced data analytics in organizations. The authors show that successful data analytics not solely depends on technical expertise, but also on the culture, stakeholder support, and governance. Thus, a comprehensive maturity model is provided, that guides organizations to improving their data analytics readiness.

Table 2 Overview of different IDS solutions along the CRISP-DM-lifecycle.

Phases	Solution	Short Explanation
Business Understanding	Business to Analytics Canvas (B2AC) [42].	• The B2AC is a workshop template that systematically matches business questions, data analytics tasks, algorithms, etc.
	Analytics Canvas [9].	• The analytics canvas is a workshop method to match data analytics solutions and data sources. Various sub-forms like [43] exist.
	Cost / Benefit Estimation procedure [44].	• Describe influencing factors for the investment decision into digitalization in production.
Data Understanding	Data Map [39].	• The data map is a workshop method to systematically identify data generated in business processes. Various sub-forms like [45]exist.
	Usage Data Classification Schemes [29].	• Provide three concepts to structure usage data from intelligent products for further analysis.
Data Preparation	Framework for the domain-driven pre-processing of sensor data [46].	• The framework involves domain-experts to create a ready-to-use data from sensor systems.
Modeling	Machine status classification [47].	• Use time series to predict machine conditions.
Evaluation	KPI-set of future production [48].	• The authors provide an overview of 38 key-performance-indicators (KPI).
Deployment	Exemplary architecture for IT-system integration [49].	• Demonstrate how to integrate machine learning models into existing IT-systems.
General	CRISP-DMME [40].	• An extension to the CRISP-DM that includes specific production and maintenance tasks.
	Citizen Data Scientist Concept [50].	• An extension to the CRISP-DM that includes and trains domain experts along the steps.
	Process-reference model for data-driven product planning [51].	• Provides a process model to guide through the data-driven product planning.
	Framework for predictive maintenance [52].	• Architecture and procedure model to realize predictive maintenance in factories.
	Maturity Models [53,54].	• Various maturity models to describe data analytics in organizations have been proposed.

In summary, while multiple approaches on different levels of IDS exist, we find that not all identified challenges in section 2.1 are addressed by the knowledge base. For example, we find that there is no method to adequately identify data necessary for prescriptive use cases. Hence, we further investigate the current challenges and best practices with eight interviews of IDS experts.

3. Research Design

To identify challenges and best practices of IDS in practice, we follow a four-step research design. First, we conducted a thorough literature review, in which we identified challenges and practical tools for IDS (see section 2). Second, we conducted eight semi-structured interviews following the method of Myers and Newman [55]. An overview of the interviewed practitioners, their experience in Data Science, respective industry and job title is given in Table 3 below. All interview partners where from the German industry sector.

Third, we transcribed the interviews and identified challenges and best practices. We dropped statements that are not related to these two aspects (e.g., mentions of other digitalization use cases). Fourth, the research team grouped the statements according to the EALM dimensions. The EALM was chosen because it provides an analytics specific representation of people, technology, and organization [13]. This grouping is presented in the following section.

Table 3 Overview of interviews (Interv.). Industry-Coding: ME = Mechanical Engineering; IA = Industrial Automation; CG = Consumer Goods.

Interv.	Experience	Industry	Job-Title
I1	9 years	ME	Industrie 4.0 Manager
I2	4 years	IA	Software Product Owner
I3	12 years	ME	Industrie 4.0 Manager
I4	6 years	Logistics	Data Product Owner
I5	5 years	Electronics	Supply Chain Analyst
I6	3 years	CG	Industrie 4.0 Manager
I7	7 years	Consulting	Technology Consultant
I8	16 years	CG	Data Scientist

4. Industrial Data Science in practice

In the following, the findings from the interviews will be structured into two main sections. The first section focusses on the existing challenges in industrial data science. They can be seen as a call for research and emphasize possible pitfalls for researchers and practitioners alike. The second section reflects on given best practices from the practitioners. The best practices serve as a baseline for future implementation endeavors. They summarize the current relevant knowledge base in industrial data science for practitioners. The impact of trends and the current development of the field is discussed in chapter five. As indicated in the research Design, the EALM approach is used as a structuring method. We expanded it to be able to also include management and people related aspects. Some reflections on the findings are provided in chapter five.

4.1. Challenges

This subchapter provides a brief overview of the challenges in industrial data science from a practitioner's point of view presented in Fig. 1. According to interviewee I4, the **Domain Understanding and Use Case Definition** is hindered by an existing threshold built up by bureaucracy and company specific regulations. Analytics capabilities, software tools or knowledge are often not available to everyday users. Thus, creativity and potentials for optimization from a grass root kind of perspective are suppressed. Data Analytics Algorithms were not developed for the domain "industry". Thus, problems like sparse data and lack of data quality exist when talking about the efficient use, **integration,** and deployment of algorithms (according to interviewee I7). The **Data Infrastructure** as an enabler for analytics is bound to high initial investments to connect machines and create a baseline infrastructure for analytics. Same applies for possible tools to enable domain experts to become citizen data scientists (interviewee I3). When building use cases on the existing **infrastructure**, the Data to use case match is still a challenge (interviewee I5).

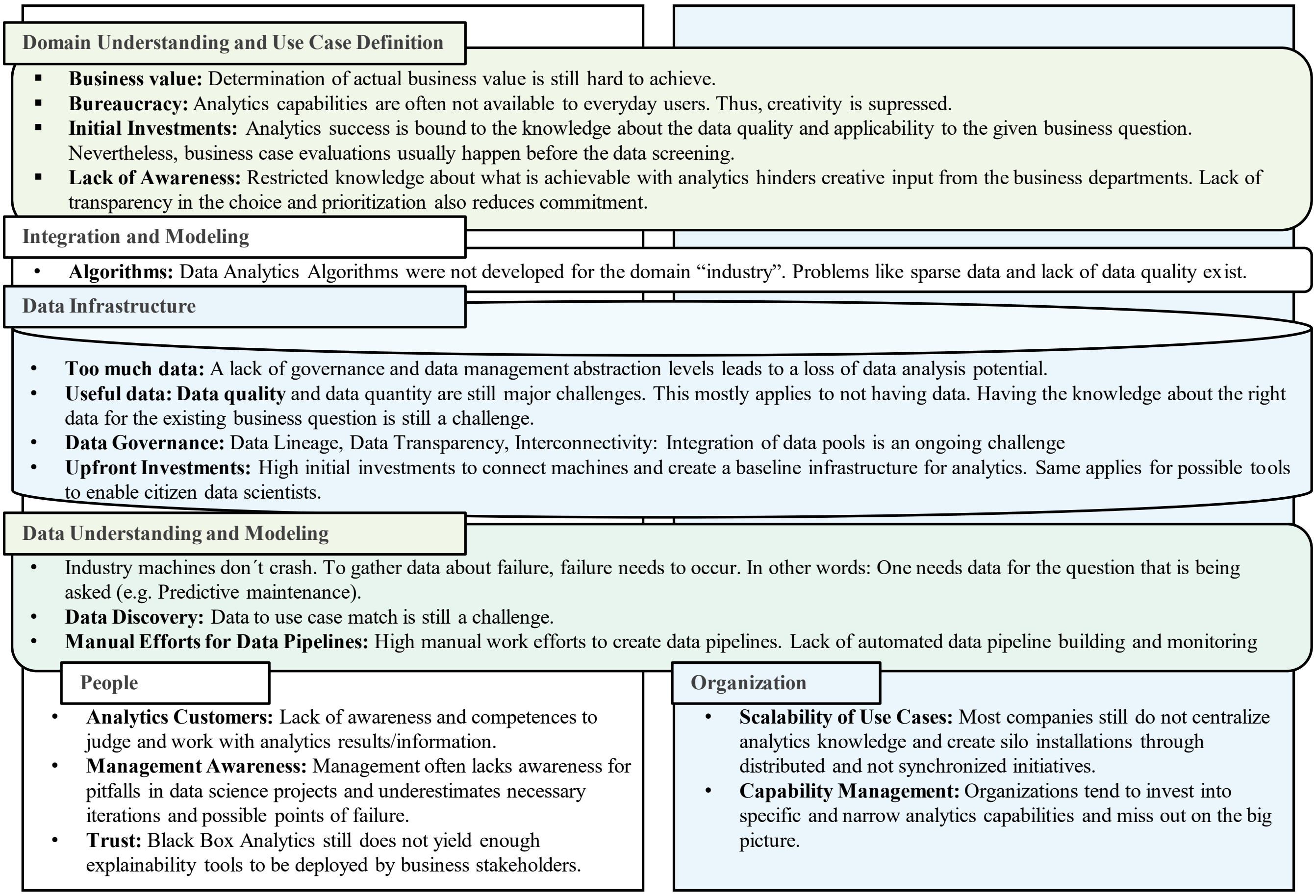

Fig. 1. Current Challenges in Industrial Data Science according to the interviewees, structured in the EALM framework.

When managing industrial data science**:** Management often lacks awareness for pitfalls in data science projects and underestimates necessary iterations and possible points of failure (interviewee I1). **Organizations** tend to invest into specific and narrow analytics capabilities and miss out on the big picture. (I3). This creates the potential for inefficiency due to the possibility of creating the same solutions multiple times because of missing efforts for synchronisation of distributed data scientists and data science teams (interviewee I1). All presented challenges (aggregated in Fig. 1) are industrial data science focused. Because of the resulting sum of interviews, no numerical ranking was possible. The focus of all interviewees was laying on use case discovery, data and organizational (management) aspects. The challenges provided in the text were the most prominent ones.

4.2. Best Practices

This subchapter provides a brief overview of the challenges in industrial data science from a practitioner's point of view presented in Fig. 2. With regard to the **Use Case Definition**, Testbeds are an essential tool to earn business owners trust on analytics and the underlying use case specific potential (interviewee I7). The degree of standardization per use case is company specific but needs to be agreed on by the relevant stakeholders. This can ensure a smoother rollout and enable the scalability of solutions for **integration and modeling**. This is especially important in terms of how to integrate the use cases into the existing **Data Infrastructure** (interviewee I3). When purchasing new machines for the shopfloor, the requirements list for new resources/machinery needs to include data analytics related aspects to ensure data availability, data quality and data lineage (interviewee I7).

Teaching rudimentary analytics knowledge to business stakeholders (EALM dimension: **people**) can greatly increase use case discovery capabilities. (Interviewee I2, I4 and, I1). Implementing use case catalogues or standardized services can help **organizations** to scale use cases and make their area specific efforts visible (Interviewee 3).

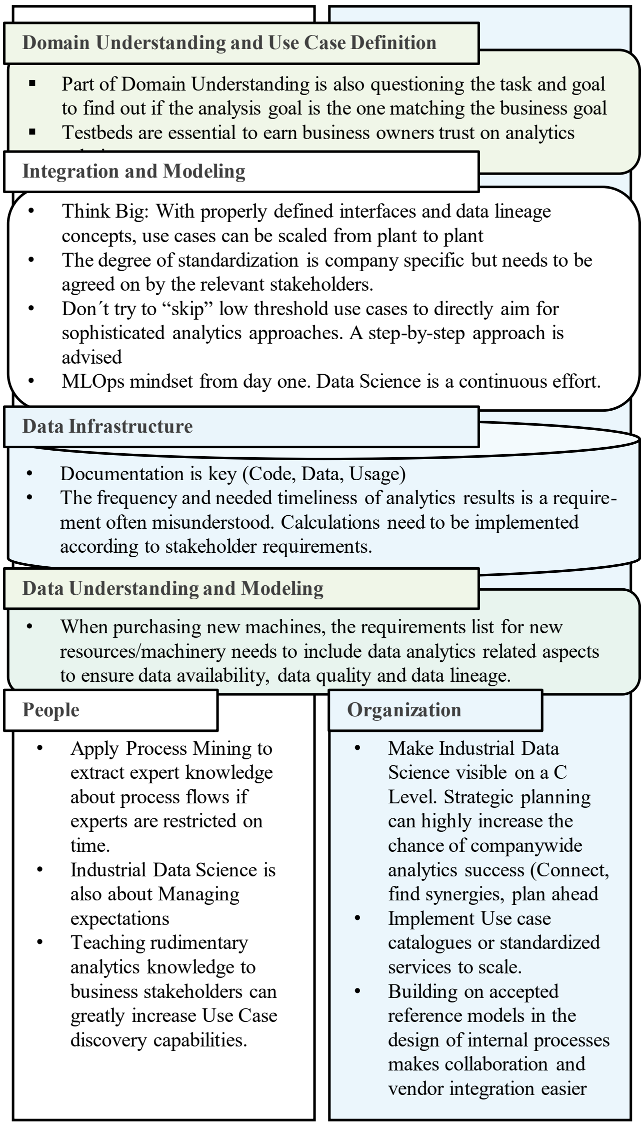

Fig. 2. Best Practices in Industrial Data Science

4.3. Future Research Directions

As per the last part of the interviews, we challenged the interviewees to name prominent trends and developments they expect in the ever-expanding field of industrial data science.

I3 raised the question of how much IDS will be outsourced or standardized into products and services. For this, sticking to reference processes will become more important than ever to foster cooperation and continue the trend of developing scalable solutions (**Procedure Models and Project Implementation)**. I4 expects the continuous development of more advanced use cases (e.g. prescriptive). For these, a connection to data strategy and more management related aspect to steer IDS initiatives and move on from opportunity driven use case deployments is expected by I3 (**Management and multiple Projects**). The same interviewee emphasizes the role that AI (e.g. Coding assistance for citizen data scientists) will play in the future to develop more sophisticated industrial data science Solutions. **Team Composition** is a vital factor that enables successful projects. The composition might change in the future as more and more data science tools may become a commodity. Common enabler technologies like Generative AI, Process Mining and Computer Vision will still play an important role. **Enterprise Architecture Management** could serve as a baseline to further align Business and IT regarding the usage and deployment of industrial data science applications.

5. Management of Industrial Data Science

Based on the **EALM** and the findings from the interviews, we propose Fig. 3 as an overarching Framework to summarize our findings for the four focus topics in industrial data science. We include the presented Layer Model, CRISP-DM, Roles in industrial data science as well as the overarching Management perspective. We emphasize that different interdisciplinary roles are needed but refer to respective studies on which to apply in which circumstance [33,56]. The elements creating the framework are grounded in the interview findings from chapter 4.3. All mentioned aspects structure the field of industrial data science to enable the discovery of possible **Synergies and Synchronization** points in between projects and initiatives. We advocate practitioners to take an **organizational perspective** on industrial data science and try to synchronize existing use case implementation efforts. Our goal is to emphasize the importance of taking management related measures into account and including efforts for synchronization and IT infrastructure. We emphasize that IDS is a team effort.

6. Conclusion and Outlook

We contributed a condense summary of the current state of research on industrial data science. The state of research was structured into both challenges and current artifact-like approaches to tackle these. Based on an interview study, we derived current challenges and best practices based on practitioner's feedback. For the study, 8 different interviews with experienced practitioners were conducted. Current shortcomings of our approach include a narrow domain focus and a general focus on German (but internationally operating) companies from the manufacturing sector.

The best practices can be used as a checklist for success factors in the planning phase for future industrial data science endeavors. This way, an integration into the existing knowledge base regarding Industrial Data Science can be ensured. In the future, we plan on expanding the provided best practices into solution patterns for the deployment of industrial data science.

Future work includes the following: Further integration of Analytics Solutions into existing Decision-Making Processes. Management of industrial data science Regarding interconnections to overarching aspects (data strategy) and underlaying aspects (infrastructure). Especially the management support and integration still poses challenges to researchers and practitioners alike.

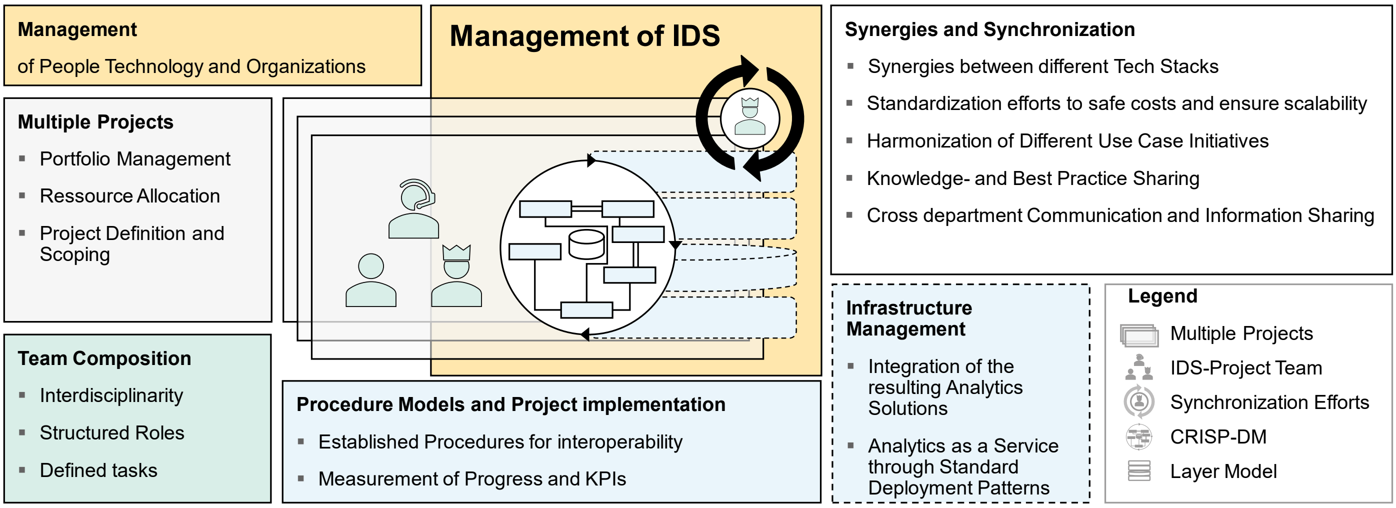

Fig. 3. Management of Industrial Data Science

References

[1] Schulte, L., Schmitt, J., Stankiewicz, L., Deuse, J., 2020. Industrial Data Science - Interdisciplinary Competence for Machine Learning in Industrial Production, in: Schüppstuhl, T., Tracht, K., Henrich, D. (Eds.), Annals of scientific society for assembly, handling and industrial robotics. Springer Vieweg, Berlin, Heidelberg, pp. 161–171.

[2] Raut, R., Narwane, V., Kumar Mangla, S., Yadav, V.S., Narkhede, B.E., Luthra, S., 2021. Unlocking causal relations of barriers to big data analytics in manufacturing firms, 30 pp.

[3] Wang, J., Zhang, W., Shi, Y., Duan, S., Liu, J. Industrial Big Data Analytics: Challenges, Methodologies, and Applications.

[4] Deuse, J., West, N., Syberg, M., 2022. Rediscovering scientific management. The evolution from industrial engineering to industrial data science. Int. J. Prod. Manag. Eng. 10 (1), 1–12.

[5] Mazarov, J., Wolf, P., Schallow, J., Nöhring, F., Deuse, J., Richter, R., 2019. Industrial Data Science in Wertschöpfungsnetzwerken. ZWF 114 (12), 874–877.

[6] Nolte, V., Sindram, T., Mazarov, J., Deuse, J., 2020. Industrial Data Science erfolgreich implementieren. ZWF 115 (10), 734–737.

[7] West, N., Gries, J., Brockmeier, C., Gobel, J.C., Deuse, J., 2021. Towards integrated Data Analysis Quality: Criteria for the application of Industrial Data Science, in: 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI). 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), Las Vegas, NV, USA. 10.08.2021 - 12.08.2021. IEEE, pp. 131–138.

[8] Bauer, N., Stankiewicz, L., Jastrow, M., Horn, D., Teubner, J., Kersting, K., Deuse, J., Weihs, C., 2018. Industrial Data Science: Developing a Qualification Concept for Machine Learning in Industrial Production.

[9] Kühn, A., Joppen, R., Reinhart, F., Röltgen, D., Enzberg, S. von, Dumitrescu, R., 2018. Analytics Canvas – A Framework for the Design and Specification of Data Analytics Projects. Procedia CIRP 70, 162–167.

[10] Reinhart, F. Industrial Data Science - Data Science in der industriellen Anwendung. Industrie 4.0 Management.

[11] Reinhart, F., Kühn, A., Dumitrescu, R., 2017. Schichtenmodell für die Entwicklung von Data Science Anwendungen im Maschinen- und Anlagenbau.

[12] Münch, C., Marx, E., Benz, L., Hartmann, E., Matzner, M., 2022. Capabilities of digital servitization: Evidence from the socio-technical systems theory. Technological Forecasting and Social Change 176, 121361.

[13] Ulich, E., 2013. Arbeitssysteme als soziotechnische Systeme–eine Erinnerung.

[14] Steenstrup, K., Sallam, R.L., Eriksen, L., Jacobson, S.F., 2014. Industrial Analytics Revolutionizes Big Data in the Digital Business.

[15] Hertog, D. den, Postek, K., 2016. Bridging the gap between predictive and prescriptive analytics - new optimization methodology needed.

[16] Karim, R., Galar, D., Kumar, U., 2023. AI factory: Theories, applications and case studies, First edition ed. CRC Press Taylor & Francis Group, Boca Raton.

[17] Zhang, B., Li, W., Li, X.-L., Ng, S.-K., 2018. Intelligent Fault Diagnosis Under Varying Working Conditions Based on Domain Adaptive Convolutional Neural Networks. IEEE Access 6, 66367–66384.

[18] Padovano, A., Longo, F., Nicoletti, L., Gazzaneo, L., Chiurco, A., Talarico, S., 2021. A prescriptive maintenance system for intelligent production planning and control in a smart cyber-physical production line. Procedia CIRP 104, 1819–1824.

[19] Meister, M., Beßle, J., Cviko, A., Böing, T., Metternich, J., 2019. Manufacturing Analytics for problem-solving processes in production. Procedia CIRP 81, 1–6.

[20] Geissbauer, R., Bruns, M., Wunderlin, J., 2022. PwC Digital Factory Transformation Survey: Digital backbone, use cases and technologies, organizational setup, strategy and roadmap, investment focus.

[21] Diedrich, A., Niggemann, O., 2022. On Residual-based Diagnosis of Physical Systems. Engineering Applications of Artificial Intelligence 109, 104636.

[22] Lepenioti, K., Bousdekis, A., Apostolou, D., Mentzas, G., 2020. Prescriptive analytics: Literature review and research challenges. International Journal of Information Management 50, 57–70.

[23] Chen, H., Liu, R., Xie, Z., Hu, Q., Dai, J., Zhai, J., 2022. Majorities help minorities: Hierarchical structure guided transfer learning for few-shot fault recognition. Pattern Recognition 123, 108383.

[24] Fredriksson, T., Mattos, D.I., Bosch, J., Olsson, H.H., 2021. Data Labeling: An Empirical Investigation into Industrial Challenges and

Mitigation Strategies, in: Morisio, M., Torchiano, M., Jedlitschka, A. (Eds.), PRODUCT-FOCUSED SOFTWARE PROCESS IMPROVEMENT. 21st international, vol. 12562. Springer, [S.l.], pp. 202–216.

[25] Meyer, M., Fichtler, T., Koldewey, C., Dumitrescu, R., 2022. Potentials and challenges of analyzing use phase data in product planning of manufacturing companies. AIEDAM 36.

[26] Li, J., Tao, F., Cheng, Y., Zhao, L., 2015. Big Data in product lifecycle management. Int J Adv Manuf Technol 81 (1-4), 667–684.

[27] Yin, S., Kaynak, O., 2015. Big Data for Modern Industry: Challenges and Trends [Point of View]. Proc. IEEE 103 (2), 143–146.

[28] Fan, C., Chen, M., Wang, X., Wang, J., Huang, B., 2021. A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. Front. Energy Res. 9.

[29] Panzner, M., Enzberg, S. von, Meyer, M., Dumitrescu, R., 2022. Characterization of Usage Data with the Help of Data Classifications. J Knowl Econ.

[30] Meister, F., Khanal, P., Daub, R., 2023. Digital-supported problem solving for shopfloor steering using case-based reasoning and Bayesian networks. Procedia CIRP 119, 140–145.

[31] Klees, M., Evirgen, S., 2022. Building a smart database for predictive maintenance in already implemented manufacturing systems. Procedia Computer Science 204, 14–21.

[32] Becker, W., Ulrich, P., Botzkowski, T. Data Analytics im Mittelstand: aus: Management und Controlling im Mittelstand. Springer 2016.

[33] Bokrantz, J., Subramaniyan, M., Skoogh, A., 2023. Realising the promises of artificial intelligence in manufacturing by enhancing CRISP-DM. Production Planning & Control, 1–21.

[34] Piorkowski, D., Park, S., Wang, A.Y., Wang, D., Muller, M., Portnoy, F., 2021. How AI Developers Overcome Communication Challenges in a Multidisciplinary Team. Proc. ACM Hum.-Comput. Interact. 5 (CSCW1), 1–25.

[35] S. Moore, 2017. Gartner says more than 40 percent of data science tasks will be automated by 2020. https://www.gartner.com/en/newsroom/press-releases/2017-. Accessed 6 December 2023.

[36] Konstantin Hopf, Oliver Müller, Arisa Shollo, and Tiemo Thiess. Organizational Implementation of AI: Craft and Mechanical Work.

[37] Shirley Coleman, Rainer Göb, Giuseppe Manco, Antonio Pievatolo, Xavier Tort-Martorell, Marco Seabra Reis. How Can SMEs Benefit from Big Data? Challenges and a Path Forward.

[38] Kletti, J., Rieger, J., 2022. Die perfekte Produktion: Manufacturing Excellence in der Smart Factory.

[39] Joppen, R., Enzberg, S., Kühn, A., Dumitrescu, R., 2019. Data map – method for the specification of data flows within production. Procedia CIRP 79, 461–465.

[40] Huber, S., Wiemer, H., Schneider, D., Ihlenfeldt, S., 2019. DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. Procedia CIRP 79, 403–408.

[41] Korsten, G., Aysolmaz, B., Turetken, O., Edel, D., Ozkan, B., 2022. ADA-CMM: A Capability Maturity Model for Advanced Data Analytics.

[42] Panzner, M., Meyer, M., Enzberg, S. von, Dumitrescu, R., 2022. Business-to-Analytics Canvas - Translation of Product Planning-Related Business Use Cases into Concrete Data Analytics Tasks. Procedia CIRP 109, 580–585.

[43] Brock, J., Enzberg, D.-I.S. von, Kühn, D.-I.A., Dumitrescu, D.-I.R., 2023. Process Mining Data Canvas: A method to identify data and process knowledge for data collection and preparation in process mining projects. Procedia CIRP 119, 602–607.

[44] Joppen, R., Lipsmeier, A., Tewes, C., Kühn, A., Dumitrescu, R., 2019. Evaluation of investments in the digitalization of a production. Procedia CIRP 81, 411–416.

[45] Massmann, M., Meyer, M., Frank, M., Enzberg, S. von, Kühn, A., Dumitrescu, R., 2020. Method for data inventory and classification. Procedia CIRP 93, 234–239.

[46] Brock, J., Rempe, N., Enzberg, S. von, Kühn, A., Dumitrescu, R., 2023. A Framework For The Domain-Driven Utilization Of Manufacturing Sensor Data In Process Mining: An Action Design Approach. Hannover : publish-Ing, 771 pp.

[47] Merkelbach, S., Afroze, L., Janssen, N., Enzberg, S. von, Kühn, A., Dumitrescu, R., 2022. Using vibration data to classify conditions in disk stack separators. Vib. proced. 46, 21–26.

[48] Joppen, R., Enzberg, S. von, Gundlach, J., Kühn, A., Dumitrescu, R., 2019. Key performance indicators in the production of the future. Procedia CIRP 81, 759–764.

[49] Weller, J., Roesmann, D., Eggert, S., Enzberg, S. von, Gräßler, I., Dumitrescu, R., 2023. Identification and prediction of standard times in machining for precision steel tubes through the usage of data analytics. Procedia CIRP 119, 514–520.

[50] Merkelbach, S., Enzberg, S. von, Kuhn, A., Dumitrescu, R., 2022. Towards a Process Model to Enable Domain Experts to Become Citizen Data Scientists for Industrial Applications, in: 2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS). 2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS), Coventry, United Kingdom. 24.05.2022 - 26.05.2022. IEEE, pp. 1–6.

[51] Meyer, M., Wiederkehr, I., Panzner, M., Koldewey, C., Dumitrescu, R., 2022. A Reference Process Model for Usage Data-Driven Product Planning.

[52] Enzberg, S. von. Implementation and Transfer of Predictive Analytics for Smart Maintenance: A Case Study.

[53] Brock, J., Löhr, B., Brennig, K., Seger, T., Bartelheimer, C., Enzberg, S. von, Kühn, A., Dumitrescu, R., 2023. A process mining maturity model: Enabling organizations to assess and improve their process mining activities. ECIS 2023 Research Papers.

[54] Ginger Korsten, Banu Aysolmaz, O. Türetken, Diederick Edel, Barış Özkan, 2022. ADA-CMM: A Capability Maturity Model for Advanced Data Analytics. Hawaii International Conference on System Sciences.

[55] Myers, M.D., Newman, M., 2007. The qualitative interview in IS research: Examining the craft. Information and Organization 17 (1), 2–26.

[56] Kutz, J., Neuhüttler, J., Schaefer, K., Spilski, J., Lachmann, T., 2023. Generic Role Model for the Systematic Development of Internal AI-based Services in Manufacturing, in: Bui, T.X. (Ed.), Proceedings of the 56th Annual Hawaii International Conference on System Sciences. January 3-6, 2023. Department of IT Management Shidler College of Business University of Hawaii, Honolulu, HI.