A Reference Process Model for Usage Data-Driven Product Planning

Maurice Meyer
Paderborn University
Maurice.Meyer@hni.upb.de

Ingrid Wiederkehr
Paderborn University
Ingrid.Wiederkehr@hni.upb.de

Melina Panzner
Paderborn University
Melina.Panzner@hni.upb.de

Dr.-Ing. Christian Koldewey
Paderborn University
Christian.Koldewey@hni.upb.de

Prof. Dr.-Ing. Roman Dumitrescu
Paderborn University
Roman.Dumitrescu@hni.upb.de

Abstract

Cyber-physical systems generate and collect huge amounts of usage data during operation. Analyzing these data may enable manufacturing companies to identify weaknesses and learn about the users of their products. Such insights are valuable in the early phases of product development like product planning, as they facilitate decision-making for product improvement. The analysis and exploitation of usage data in product planning, however, is a new task for manufacturing companies. To reduce mistakes and improve the results, companies should build upon a suitable reference process model. Unfortunately, established models for analyzing data cannot be easily applied for product planning. In this paper, we propose a reference process model for usage data-driven product planning. It builds on three well-established models for analyzing data and addresses the unique characteristics of usage data-driven product planning. Finally, we customize the model for a manufacturing company and demonstrate how it could be implemented in practice.

1. Introduction

The further development of products is the main development focus of engineers. In a survey with 247 engineers, ALBERS *et al.* found that only 7% of developments are true new developments without carry overs or adjustments of existing products. In contrast, 93% of developments represent further developments of existing products [1]. To further develop a product, engineers need to answer questions like: How does the product perform in the field? What are its strengths and weaknesses? How do users utilize the product? What are new requirements for the product? Answering such questions is not trivial, as companies often lack high quality information about the product usage phase [2].

A new solution for this problem emerges from the progressive digitalization of products. In recent years, the digitalization has transformed mechatronic products into so-called cyber-physical systems that integrate hardware, sensors, data storage, microprocessors, software, and connectivity. These systems can collect data about themselves and their environment during their utilization [3].

The analysis of such usage data promises to be especially valuable during the early stages of product development like product planning [4]. As the initial phase or phase zero of product development, product planning aims at finding success potentials of the future [5] and promising products to be developed [6]. In this regard, usage data analytics can help exploit insights about predecessor products and their users that lead to new success potentials and set the agenda for product improvements. In conjunction, usage data analytics and product planning span the new research area *usage data-driven product planning* [7].

At present, the analysis of usage data in product planning is not widely researched. For example, BERTONI found that research on analyzing data to identify customer needs mainly focuses on data from social media and online reviews [8]. In addition, approaches that address usage data often have a broader perspective than product planning. For example, WILBERG *et al.* present a stakeholder-oriented procedure for the development of a use phase data strategy [10]. The approach addresses stakeholders in different business functions, e.g., service or marketing. While such approaches help to identify the potential value of usage data for the different stakeholders of a company, they do not help analyze usage data in product planning.

As there are no existing approaches, the creation of a reference process model for usage data-driven product planning is required. In this paper, the term *reference process model* describes a generic process model that formalizes recommended practices for a certain domain [11]. Our research question is as follows: *How does a reference process model for usage data-driven product planning have to be designed?*

HICSS

Utilizing a design science approach, the result of this paper is a reference process model for usage data-driven product planning. It guides companies with a stepwise procedure and helps structure projects in usage data-driven product planning. It is especially useful for manufacturing companies that want to start exploiting usage data in product planning.

The paper is structured as follows: In section 2, we describe our research design for the construction of the reference process model. Section 3 presents the reference process model for usage data-driven product planning in detail. In section 4, we show the customization of the model for a manufacturing company. Section 5 concludes the paper.

2. Research design

For the development of the reference process model for usage data-driven product planning, we followed the design science research (DSR) guidelines. According to HEVNER *et al.*, DSR aims at developing viable artifacts (e.g., models or methods) for relevant business problems. These artifacts must be novel and provide a clear contribution through solving so far unsolved problems. They are created and evaluated by applying rigorous methods and utilizing the knowledge base [12].

While design science research represents the foundation of our research design, our concrete procedure broadly derives from the *Process Model for an Empirically Grounded Reference Model Construction* by AHLEMANN and GASTL. This process model emphasizes the importance of consulting domain experts when constructing and validating reference process models [13]. In addition to that, we analyzed established models following DE LA HIDALGA *et al.* [14] and FRANK *et al.* [15]. To carve out a suitable reference process model from all the input data, we utilized VOM BROCKE's *design principles* [16]. Overall, our procedure comprises four phases: *Domain analysis*, *Model construction*, *Empirical validation*, and *Customization* (see Figure 1). Subsequently, the phases are described in detail.
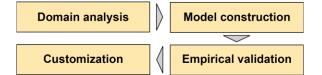

Figure 1: Procedure model for the development of the reference process model

2.1. Domain analysis

In the domain analysis phase, relevant knowledge about the domain shall be captured and prepared for the model construction [13].

For the considered domain, MEYER *et al.* provide a recent comprehensive systematic literature review [7]. In their work, the authors derive the main concepts, advantages, success factors and challenges of usage data-driven product planning [7] und thus present a substantial overview of the topic. Therefore, we used the literature review as our knowledge base.

Next, we investigated our knowledge base in search of necessary process steps for the reference process model. AHLEMANN and GASTL propose this step as a suitable way to structure and aggregate the knowledge available [13]. We analyzed the knowledge base using the question: *Which process steps does the desired reference process model need to include?* Table 1 shows the derived process steps for the reference process model. They represent domain requirements for the model construction and conclude our domain analysis.

Table 1: Necessary process steps for the reference process model derived from the knowledge base

ID	Process step
P-1	Analysis of the product and strategy [10, 17]
P-2	Analysis of the data analytics capabilities [10]
P-3	Definition of use cases [4, 17, 18]
P-4	Definition of data needs [3, 19]
P-5	Collection of usage data [19–21]
P-6	Pre-processing of usage data [19, 21–23]
P-7	Analysis of usage data, i.a. [19, 22, 24, 25]
P-8	Validity-check of data analysis results [19, 26]
P-9	Interpretation of the analysis results [19, 27]
P-10	Creation of new ideas [28]
P-11	Identification of requirements [4, 27, 29, 30]
P-12	Planning the improvement of existing and future products, i.a. [17, 21, 25, 26, 28, 31, 32]

2.2. Model construction

The model construction phase aims at developing a first version of the reference process model [13]. This version is based on the necessary process steps from the knowledge base and on existing process models.

As our desired reference process model shall describe how to analyze usage data of existing products to find improvement potentials, we focused our analysis of existing process models on models for analyzing data. For this, numerous process models exist. MARISCAL *et al.* and PLOTNIKOVA *et al.* provide an overview about existing process models and show that almost all identified approaches are based on two original models:

the *Cross Industry Standard Process for Data Mining* (CRISP-DM) and the *Knowledge Discovery in Databases* model (KDD) [33, 34]. Therefore, these two models are subsequently described in more detail.

CRISP-DM is an industry-, tool- and application-neutral process model for data mining. It was created by the CRISP-DM CONSORTIUM and is based on the experience of numerous data mining practitioners [35, 36]. For several years, it has been considered the de facto standard for data mining projects.

The process model consists of six iterative phases: (1) *Business understanding* aims at understanding the business objectives and requirements and converting them into a concrete data mining problem definition. (2) In *data understanding*, the data are collected and investigated to spot, for example, data quality problems. (3) The *data preparation* includes tasks like the transformation and cleaning of the data to construct the final dataset. (4) During *modeling*, models are built and refined with the data. (5) In the *evaluation*, the model must be evaluated regarding the business objectives. (6) Finally, in *deployment*, the model is put into operation and a report is generated [35].

KDD refers to the overall process of discovering useful knowledge from data [37]. FAYYAD *et al.* define the KDD process as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [38].

The process consists of nine iterative steps: (1) In *learning the application domain*, relevant domain knowledge and the desired goals are captured. (2) *Creating a target dataset* aims at selecting the dataset to be analyzed. (3) In *data cleaning and pre-processing*, basic operations like noise reduction and outlier removal take place. (4) *Data reduction and projection* focuses on tasks like feature engineering and dimensionality reduction. (5) In *choosing the function of data mining*, the purpose of the desired model is decided (e.g., classification, regression). (6) *Choosing the data mining algorithm(s)* addresses the selection of suitable method(s) for searching patterns in the data, e.g., by comparing different models and parameters. (7) *Data mining* describes the search for patterns in the data. (8) In *interpretation of the results*, the patterns discovered are interpreted and translated into the domain language. (9) The final step *usage of the discovered knowledge* covers documenting the new knowledge, reporting it and taking action [38].

In addition to these industry-neutral models like CRISP-DM, KDD, and their derivatives, we decided to also analyze one manufacturing-focused approach to account for any industry-specific aspects. The **VDI/VDE 3714 guideline** presents a standard for the implementation and operation of big data applications in the manufacturing industry. It aims at aggregating the numerous contributions towards big data analytics in the manufacturing industry and at unifying them into one model [39].

The guideline describes seven iterative phases: (1) In the *definition* phase, the questions and objectives to be answered or achieved need to be specified. (2) Next, *exploring the data situation* aims at describing and structuring the available data and defining additionally required data. (3) In *data management*, data from different sources are merged. (4) *Modeling* deals with creating an evaluable model from the data. (5) Subsequently, an initial *evaluation of the data analysis results* with respect to the project goals is necessary. (6) *Implementation and rollout* aims at transferring the big data application into continuous operation. (7) The final phase *sustainability* addresses the project documentation as well as an assessment of economic, technical and social aspects to ensure a sustainable impact of the big data project [39].

To construct our first version of the reference process model, first, we analyzed the three existing process models CRISP-DM, KDD, and the VDI/VDE 3714 guideline. We followed DE LA HIDALGA *et al.* and FRANK *et al.* and created a detailed process model for each, including all tasks and steps mentioned in their descriptions [14, 15]. We also synchronized their steps, thereby highlighting gaps and similarities in the models.

Second, we aggregated the established reference models to build one exhaustive model from the three original models [16]. We found that we could sort all process steps into **four main processes**: (1) *Planning of the data analysis*, (2) *Analytics and data preparation*, (3) *Analytics workflow design and modeling* and (4) *Exploitation of the data analysis results*. In detail:

(1) **Planning of the data analysis** contains process steps related to the early business perspective on the project: Parts of CRISP-DM's *business understanding* (e.g., the determination of business objectives and requirements); KDD's *learning the application domain*; VDI/VDE 3714 guideline's *definition*.

(2) **Analytics and data preparation** includes process steps concerning the clarification of the analytics task as well as the selection and collection of the data: Parts of CRISP-DM's *business understanding* (e.g., the determination of data mining goals), CRISP-DM's *data understanding*; KDD's *creating a target dataset*; VDI/VDE 3714 guideline's *exploring the data situation* and *data management*.

(3) **Analytics workflow design and modeling** summarizes the process steps addressing data pre-processing and analysis: CRISP-DM's *data preparation* and *modelling*; KDD's *data cleaning and pre-processing*, *data reduction and projection*, *choosing the function of data mining*, *choosing the data mining algorithm(s)*, and *data mining*; VDI/VDE 3714

guideline's *modeling* and parts of the *evaluation of the data analysis results*.

(4) **Exploitation of the data analysis results** consists of process steps for the interpretation and utilization of the data analysis results: CRISP-DM's *evaluation*; KDD's *interpretation of the results* and *usage of the discovered knowledge*; Parts of VDI/VDE 3714 guideline's *evaluation of the data analysis results*.

We omitted the following process steps, as they aim at the continuous operation of the built models and therefore do not contribute to the purpose of our reference process model: CRISP-DM's *deployment*; VDI/VDE 3714 guideline's *implementation and rollout* and *sustainability*.

Third, again following VOM BROCKE's design principles, we specialized the aggregated model for our considered domain [16]. For each necessary process step from Table 1, we analyzed if it was already sufficiently addressed in the model. If not, we added a new process step (e.g., *derivation of use cases*) following DE LA HIDALGA *et al.* and FRANK *et al.* [14, 15]. As a result, we obtained a first version of the reference process model with four main and 16 sub processes. Figure 2 shows the schematic procedure.
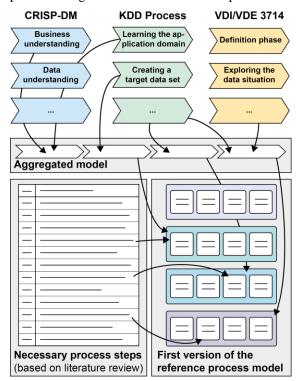

Figure 2: Schematic procedure for the model construction

2.3. Empirical validation

For the validation of the reference process model, AHLEMANN and GASTL recommend conducting interviews with topic experts [13]. We performed three semi-structured interviews of 60-90 mins duration. Our interview partners are characterized in Table 2.

Table 2: Overview about interview partners

ID	Position	Experience
I-1	Innovation and process consultant	6 years
I-2	Digitalization consultant and entrepreneur	15 years
I-3	Head of data science department	11 years

The interviews were aimed at answering the following questions*: Is the reference process model complete or is something missing? Are the sub processes structured in a logical sequence? Are the sub processes clearly separated from each other?*

We discussed each main and each sub-process with all interviewees and protocolled all remarks and questions raised in the interviews.

After each interview, we refined the reference process model according to the remarks of the experts. For example, we added the sub-process *Update product strategy* in the *Exploitation* main process. After all interviews were conducted, we showed the updated reference process model to all experts again. We asked for further remarks, but all experts were satisfied and confident that it fulfills its requirements. Section 3 presents the resulting model.

2.4. Customization

After the validation, AHLEMANN and GASTL suggest to test the reference process model [13]. We performed this practical test with a machinery and plant engineering company, which produces machine tools and production equipment in the field of forming technology.

In the test, we customized the reference process model for the company. We used the RACI method, which is also known as RAM (Responsibility Assignment Matrix) [40]. The method's goal is to assign tasks and responsibilities. For that, it uses a matrix with processes in the rows and roles in the columns [40].

For the assignment of tasks and responsibilities, four options are possible. They are derived from the name of the method: R, A, C, and I. The letter R stands for *responsible* and describes which person is responsible for a task. The letter A stands for *accountable*: It shows who decides whether a task has been performed correctly. For example, this could be a supervisor. The letter C describes that a person should be *consulted* when the task is performed. Lastly, the letter I names all people who need to be *informed* about the events and the results for that specific task [41]. While working with the matrix, several assignment rules

must be respected, e.g., only a single person can be responsible for a given task [41].

At the start of the workshop, we collected the company's relevant roles for usage data-driven product planning. Then, we successively assigned the roles to each sub-process of the reference process model. The results are shown in section 4.

3. Reference process model for usage data-driven product planning

The reference process model for usage data-driven product planning consists of four main processes: (1) *Planning of the usage data analysis*, (2) *Analytics and data preparation*, (3) *Analytics workflow design and modeling* and (4) *Exploitation of the data analysis results*. Each main process comprises four sub-processes. In the following, all sub-processes are described. The sources are given with the corresponding IDs, e.g., [I-1] and [P-1] (see Table 1 and Table 2), and abbreviations, e.g., [CRISP-DM]. Figure 3 shows the reference process model. Please note that the arrows represent sequence flows and not information flows.

(1) Planning of the usage data analysis

(1.1) Identification of investigation needs:
Input: Product strategy
Procedure: This sub-process aims at finding critical needs for investigation concerning the product or its users. For this, the business objectives specified in the product strategy are analyzed [CRISP-DM, I-1, P-1]. This strategy contains details about the product program design, e.g., about the product's strategic focus and its planned evolution [42, 43]. In addition to the strategy, the status quo of the product under investigation is examined [KDD, CRISP-DM, P-1]. From these analyses, critical investigation needs are derived.
Output: Investigation needs

(1.2) Analysis of data analytics capabilities:
Input: -
Procedure: In this sub-process, the capabilities of the organization and the product concerning usage data-driven product planning are assessed [P-2]. For example, the available resources (e.g., data mining experts, usage data) need to be analyzed [CRISP-DM].
Output: Data analytics capabilities

(1.3) Definition of boundary conditions and goals:
Input: Investigation needs (1.1); Data analytics capabilities (1.2)
Procedure: Next, boundary conditions and goals need to be defined [CRISP-DM, KDD, VDI/VDE]. Boundary conditions include requirements, prerequisites, assumptions, and constraints of the project. Just as the goals, they can refer to the organization or the product context. Goals help achieve the business objectives and guide the analytics activities. The boundary conditions and goals set the direction for the subsequent steps.
Output: Boundary conditions and goals

(1.4) Derivation of use cases:
Input: Boundary conditions and goals (1.3)
Procedure: Use cases address the previously defined content goals. They link possible influential variables to a goal or ask questions about the relationships of goals and influential variables [VDI/VDE, I-2, P-3]. For each use case, costs and benefits must be estimated [CRISP-DM].
Output: Use cases

(2) Analytics and data preparation

(2.1) Specification of analytics objectives:
Input: Data analytics capabilities (1.2); Use case (1.4)
Procedure: This sub-process transforms a use case into concrete data analytics objectives [CRISP-DM]. If domain knowledge is missing, it needs to be captured to derive relevant variables and adequate data analytics approaches (e.g., clustering or association rule mining) [KDD]. For this, the data scientists stay in close contact with the domain experts [I-3].
Output: Data analytics objectives

(2.2) Definition of data needs:
Input: Data analytics objectives (2.1)
Procedure: This sub-process specifies the data and their sources [P-4]. Also, it defines specific measurements or data objects [VDI/VDE]. This includes the transformation of the defined physical variables into concrete data attributes in the target dataset. For example, from the domain knowledge and demand (e.g., noise behavior), raw data and data sources (e.g., sound vibrations via a sound level meter) need to be derived.
Output: Data needs

(2.3) Collection of data:
Input: Data analytics capabilities (1.2); Data needs (2.2)
Procedure: In this sub-process, the data needs are compared with the actual existing data in the company. If the required data are not available, an acquisition concept must be developed and implemented [P-5].
Output: Raw data set

(2.4) Description of data:
Input: Raw data set (2.3)
Procedure: This sub-process is about understanding the collected data and analyzing them regarding their processing options [CRISP-DM]. Here, the goal is data literacy. Steps are a first exploration, a holistic
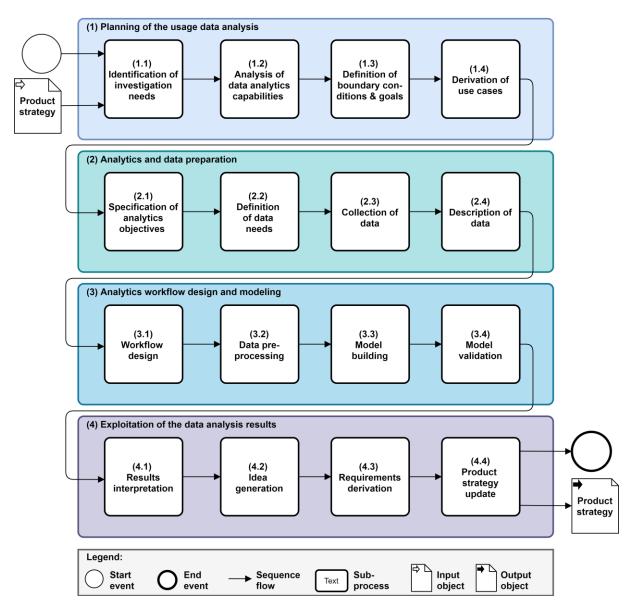
Figure 3: Reference process model for usage data-driven product planning

description of the data with relevant meta data, and a classification of the data [VDI/VDE]. If data quality is not sufficient, it is necessary to go back to the previous step and acquire the data with the required quality [I-3].
Output: Data characteristics

(3) Analytics workflow design and modeling

(3.1) Workflow design:
Input: Data analytics objectives (2.1); Data characteristics (2.4)
Procedure: In this sub-process, pre-processing and modeling techniques are selected and composed into analytics workflows [I-3]. Typically, there are several techniques for the same problem type. Some modeling

techniques need specific data formats or have certain model assumptions. This requires a close link between data preparation and modeling as well as an alignment of these steps as a coherent workflow [I-3].
Output: Conceptual workflows

(3.2) Data pre-processing:
Input: Raw data set (2.3); Conceptual workflows (3.1)
Procedure: In this sub-process, analysts need to refine the data to prepare it for modeling [CRISP-DM, KDD, P-6]. Here, the pre-processing methods defined in the workflows must be implemented (e.g., record and attribute selection, integration of different data sets, data cleaning, data transformation, and feature engineering).
Output: Pre-processed data

(3.3) Model building:
Input: Conceptual workflows (3.1); Pre-processed data (3.2)
Procedure: This sub-process is about implementing the algorithms of the workflows together with an adequate test design [CRISP-DM, KDD, VDI/VDE, P-7]. The latter examines and compares the workflows and their parameter settings.
Output: Analysis results

(3.4) Model validation:
Input: Analysis results (3.3)
Procedure: The models and their results need to be validated [CRISP-DM, VDI/VDE, P-8]. For this, suitable technical metrics and evaluation criteria must be selected. Model building and validation iterate until results no longer improve and reach satisfaction.
Output: Model performance

(4) Exploitation of the data analysis results

(4.1) Results interpretation:
Input: Use case (1.4); Analysis results (3.3); Model performance (3.4)
Procedure: The interpretation of the analysis results seeks to reveal unknown, but valuable insights about the product and its users [P-9]. It covers the evaluation of the analysis results and their transfer into the product context, e.g., the description, explanation, and verification of the results by product experts [KDD, VDI/VDE]. The insights form the starting point for the identification of new success potentials for product improvement [I-1].
Output: Success potentials for product improvement

(4.2) Idea generation:
Input: Investigation needs (1.1); Success potentials for product improvement (4.1)
Procedure: In this step, promising ideas for product improvement are generated [P-10]. They are aimed at exploiting the identified success potentials. From all ideas, the most promising ones are selected [I-1].
Output: Ideas for product improvement

(4.3) Requirements derivation:
Input: Success potentials for product improvement (4.1); Ideas for product improvement (4.2)
Procedure: Building on the success potentials and ideas for product improvement, this sub-process aims at translating the new ideas into requirements for product development [P-11].
Output: Requirements

(4.4) Product strategy update:
Input: Requirements (4.3)
Procedure: In the last sub-process, the product strategy is updated based on the new requirements [I-1]. It specifies the changes planned for new product generations and existing products [P-12].
Output: Updated product strategy

4. Customized process model

To demonstrate how it could be implemented in practice, we customized the reference process model with a machinery and plant engineering company, which produces machine tools and production equipment in the field of forming technology (see section 2.4). The company links nine roles to the implementation of the process: Purchasing agent, sales engineer, service technician, pre-developer, mechanical developer, electrical developer, virtual developer, head of development and CEO. In the following, tasks and responsibilities of each role are described. The resulting RACI-matrix is shown in Figure 4.

Of all roles, the **purchasing agent** is the least involved. He only supports during the *identification of investigation needs* by presenting current and future purchasing challenges.

The activities of the **sales engineer** are limited to the first main process. He is responsible for the *identification of investigation needs*. After that, he is consulted because of his close customer relationships. Later, he receives information about the *results interpretation* and the *product strategy update*.

The role of the **service technician** is especially important in the first two main processes. As he is in close contact with customers, he is consulted during the *planning* and *preparation* main processes. Furthermore, he supports in the *results interpretation*.

The **pre-developer** helps in the *identification of investigation needs*. At the end, he is informed about the *product strategy update* as it affects his future work.

Mechanical and electrical developers accompany the process from the *derivation of use cases* to the *product strategy update*. Early on, their expertise is needed for the *analytics and data preparation*. While they are only informed about the *analytics workflow design and modeling*, they are again consulted for *results interpretation* and *idea generation*.

The **virtual developer** is the most important role for performing usage data-driven product planning in the considered company. He is responsible for all sub-processes of the first three main processes except the *identification of investigation needs* (see sales engineer). For the second and third main process, he also oversees the results obtained. While he has a consulting function for *results interpretation*, he is only informed about the results of the sub-processes *idea generation*, *requirements derivation*, and *product strategy update*.

Sub-processes	Roles								
	Purchasing agent	Sales Engineer	Service technician	Pre-developer	Mechan. developer	Electr. developer	Virtual developer	Head of development	CEO
(1.1) Identification of investigation needs	C	R	C	C				I	A I
(1.2) Analysis of data analytics capabilities		C	C				R	C	A I
(1.3) Definition of boundary conditions & goals		C	C				R	C	A I
(1.4) Derivation of use cases			C		C	C	R	C	A I
(2.1) Specification of analytics objectives			C		C	C	R A	I	I
(2.2) Definition of data needs			C		C	C	R A	I	I
(2.3) Collection of data			C		C	C	R A	I	I
(2.4) Description of data			C		C	C	R A	I	I
(3.1) Workflow design			I		I	I	R A	I	I
(3.2) Data pre-processing			I		I	I	R A	I	I
(3.3) Model building			I		I	I	R A	I	I
(3.4) Model validation			I		I	I	R A	I	I
(4.1) Results interpretation		I	C		C	C	C	R A	I
(4.2) Idea generation					C	C	I	R	A I
(4.3) Requirements derivation					I	I	I	R	A I
(4.4) Product strategy update		I		I	I	I	I	R	A I

Legend: R Responsible   A Accountable   C Consulted   I Informed

Figure 4: Results of the customization in a RACI-matrix

The **head of development** has a consulting role during the *planning* main process. While he is only informed about the second and third main process, he is responsible to exploit the identified insights and find promising product improvements.

The last role is the **CEO**. He is informed about all process activities. Furthermore, he oversees the *planning* and *exploitation* main processes.

5. Conclusion

The result of this paper is a reference process model for usage data-driven product planning. It consists of four main processes: (1) *Planning of the usage data analysis*, (2) *Analytics and data preparation*, (3) *Analytics workflow design and modeling* and (4) *Exploitation of the data analysis results*. Each main process is divided into four sub-processes, which further explain the tasks to be performed. The model's customization with a machinery and plant engineering company shows how the reference process model could be implemented in practice. In the following, our work's contributions to theory and practice are described. Finally, its limitations as well as recommendations for future research are presented.

5.1 Contributions to theory and practice

Usage data-driven product planning is a new and promising research field. While there are numerous established reference process models for performing data mining or big data analyses, none of them can easily be applied to product planning. The reference process model presented in this paper addresses and fills this gap. Thereby, the developed artifact represents a valuable addition to the knowledge base as requested in design science research [12]. The reference process model contributes to the scientific discourse by describing the utilization of data analytics approaches in the early phase of innovation.

For the practical contribution, the model provides managers an overview about how to perform usage data-driven product planning. Managers can confidently build on the model as it is based on three process models that are widely used in practice. For the implementation, our paper lays out how to customize the model by assigning responsibilities to each sub-process.

5.2 Limitations and recommendations for future research

The result of this work is subject to four main limitations. (1) When analyzing existing reference models, we focused on three models: CRISP-DM, KDD, and VDI/VDE 3714. However, there are numerous further reference models, many of which are derivatives of the original CRISP-DM and KDD models [34]. These models might have included further aspects for our process. Yet, we are convinced that we captured all critical aspects with our selection. (2) The validation was performed with three interviews. A higher number of interviews would probably have generated further

valuable suggestions. Yet, as our interviewees have diverse experiences and perspectives, we are confident that they pointed out the most important points. (3) The reference process model conceptually describes the required steps and tasks of usage data-driven product planning. However, concrete methodical approaches are not yet included in the reference process model. (4) Until now, we have not yet fully applied the reference process model in practice, but only parts of it. Therefore, some real-world challenges could still be unaddressed.

Considering the limitations, future research should focus on the following three recommendations: (1) Methodical approaches need to be developed for each sub-process. (2) The whole process should be performed with multiple companies to find more practical challenges. (3) The reference process model should be regularly updated and improved, e.g., after its first complete application in practice.

Acknowledgements

References

[1] A. Albers, N. Bursac, J. Urbanec, R. Lüdcke, and G. Rachenkova, "Knowledge Management in Product Generation Development - an empirical study", in Design for X – Proceedings of the 25th DfX-Symposium, TUHH Universitätsbibliothek, 2014, pp. 13–24.

[2] Q. Deng, S. Wellsandt, K. Hribernik, and K.-D. Thoben, "Understanding Users and Products in Product Development: The Application of Product Usage Information and its Challenges", in Proceedings of the International Conference on Engineering Design (ICED21), Gothenburg, Sweden, 2021, pp. 3299–3308.

[3] M. E. Porter and J. E. Heppelmann, "How Smart, Connected Products Are Transforming Competition", Harvard Business Review, vol. 92, no. 11, pp. 64–88, 2014.

[4] M. Holler, F. Uebernickel, and W. Brenner, "Understanding the Business Value of Intelligent Products for Product Development in Manufacturing Industries", in Proceedings of the 8th International Conference on Information Management and Engineering, Istanbul, Turkey, 2016, pp. 18–24.

[5] J. Gausemeier, R. Dumitrescu, S. Kahl, and D. Nordsiek, "Integrative development of product and production system for mechatronic products", Robotics and Computer-Integrated Manufacturing, vol. 27, no. 4, pp. 772–778, 2011.

[6] K. T. Ulrich and S. D. Eppinger, Product design and development, 6th ed. New York, NY: McGraw-Hill, 2016.

[7] M. Meyer, I. Wiederkehr, C. Koldewey, and R. Dumitrescu, "Understanding Data-Driven Product Planning: A Systematic Literature Review", in Proceedings of the International Conference on Engineering Design (ICED21), Gothenburg, Sweden, 2021, pp. 3289–3298.

[8] A. Bertoni, "Data-Driven Design in Concept Development: Systematic Review and Missed Opportunities", Proceedings of the Design Society: DESIGN Conference, vol. 1, pp. 101–110, 2020, doi: 10.1017/dsd.2020.4.

[9] J. Wilberg, T. Kalla, M. Fetscher, F. Rimbock, C. Hollauer, and M. Omer, "Development of a Use Phase Data Strategy for Connected Products: A Case Study in Industry", in 2018 Portland International Conference on Management of Engineering and Technology (PICMET), Honolulu, HI, 2018, pp. 1–12.

[10] J. Wilberg, L. Fahrmeier, C. Hollauer, and M. Omer, "Deriving a Use Phase Data Strategy for Connected Products: A Process Model", in Proceedings of the DESIGN 2018, 2018, pp. 1441–1452.

[11] J. Ling and L. Zhang, "Generating Hierarchical Reference Process Model Using Fragments Clustering", in Asia-Pacific Software Engineering Conference, 2015, pp. 40–47.

[12] A. Hevner, S. T. March, and J. Park, "Design Science in Information Systems Research", MIS Quarterly, vol. 28, no. 1, pp. 75–105, 2004.

[13] F. Ahlemann and H. Gastl, "Process Model for an Empirically Grounded Reference Model Construction", in Reference modeling for business systems analysis, P. Fettke and P. Loos, Eds., Hershey, PA., London, Melbourne, Singapore: Idea Group Publishing, 2007, pp. 77–97.

[14] A. N. de la Hidalga, A. Hardisty, P. Martin, B. Magagna, and Z. Zhao, "The ENVRI Reference Model", in Lecture Notes in Computer Science, Towards Interoperable Research Infrastructures for Environmental and Earth Sciences, Z. Zhao and M. Hellström, Eds., Cham: Springer International Publishing, 2020, pp. 61–81.

[15] M. Frank, J. Gausemeier, N. Hennig-Cardinal von Widdern, C. Koldewey, S. Menzefricke, and J. Reinhold, "A reference process for the Smart Service business: development and practical implications", in Proceedings of the ISPIM connects, International Society for Professional Innovation Management (ISPIM), 2020, pp. 1–19.

[16] J. vom Brocke, "Design Principles for Reference Modeling: Reusing Information Models by Means of Aggregation, Specialisation, Instantiation, and Analogy", in Reference Modeling for Business Systems Analysis: IGI Global, 2007, pp. 47–76.

[17] J. Wilberg, I. Triep, C. Hollauer, and M. Omer, "Big Data in product development: Need for a data strategy", in Proceedings of PICMET '17: Technology Management for Interconnected World, Portland, USA, 2017, pp. 1–10.

[18] Hofmann, P., Jöhnk, J., Protschky, D., & Urbach, N., "Developing purposeful AI use cases – a structured method and its application in project management", in

15th International Conference on Wirtschaftsinformatik (WI), Potsdam, Germany, 2020.

[19] L. Hou and R. J. Jiao, "Data-informed inverse design by product usage information: a review, framework and outlook", J Intell Manuf, vol. 31, no. 3, pp. 529–552, 2020, doi: 10.1007/s10845-019-01463-2.

[20] Y. M. Goh and C. McMahon, "Improving reuse of in-service information capture and feedback", Journal of Manufacturing Technology Management 20(5), 2009.

[21] M. Abramovici, P. Gebus, J. C. Göbel, and P. Savarino, "Utilizing Unstructured Feedback Data from MRO Reports for the Continuous Improvement of Standard Products", in Proceedings of the 21st International Conference on Engineering Design (ICED17), Vancouver, Canada, 2017.

[22] M. E. Porter and J. E. Heppelmann, "How Smart, Connected Products Are Transforming Companies", Harvard Business Review, vol. 93, no. 10, pp. 96–114, 2015.

[23] M. Shahbaz, M. Srinivas, J. A. Harding, and M. Turner, "Product Design and Manufacturing Process Improvement Using Association Rules", Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, vol. 220, no. 2, pp. 243–254, 2006, doi: 10.1243/095440506X78183.

[24] M. Cantamessa, F. Montagna, S. Altavilla, and A. Casagrande-Seretti, "Data-driven design: the new challenges of digitalization on product design and development", Design Science, vol. 6, 2020, doi: 10.1017/dsj.2020.25.

[25] J. Igba, K. Alemzadeh, P. M. Gibbons, and K. Henningsen, "A framework for optimising product performance through feedback and reuse of in-service experience", Robotics and Computer-Integrated Manufacturing, vol. 36, pp. 2–12, 2015, doi: 10.1016/j.rcim.2014.12.004.

[26] D. van Horn, A. Olewnik, and K. Lewis, "Design Analytics: Capturing, Understanding, and Meeting Customer Needs Using Big Data", in Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Chicago, Illinois, USA, 2012, pp. 863–875.

[27] J. Wilberg, F. Schäfer, P. Kandlbinder, C. Hollauer, M. Omer, and U. Lindemann, "Data Analytics in Product Development: Implications from Expert Interviews", in IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 2017, pp. 818–822.

[28] L. Wu, L. Hitt, and B. Lou, "Data Analytics, Innovation, and Firm Productivity", Management Science, vol. 66, no. 5, pp. 2017–2039, 2020, doi: 10.1287/mnsc.2018.3281.

[29] T. Wuest, K. Hribernik, and K.-D. Thoben, "Capturing, Managing and Sharing Product Information along the Lifecycle for Design Improvement", in Proceedings of the 10th International Workshop on Integrated Design Engineering. IDE'14, pp. 107–115.

[30] M. Holler, G. Neiditsch, F. Uebernickel, and W. Brenner, "Digital Product Innovation in Manufacturing Industries – Towards a Taxonomy for Feedback-driven Product Development Scenarios", in Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS), Hawaii, USA, 2017, pp. 4726–4735.

[31] H. Holmström Olsson and J. Bosch, "Towards Data-Driven Product Development: A Multiple Case Study on Post-deployment Data Usage in Software-Intensive Embedded Systems", in vol. 167, Lean Enterprise Software and Systems: LESS 2013. Lecture Notes in Business Information Processing, B. Fitzgerald, K. Conboy, K. Power, R. Valerdi, L. Morgan, and K.-J. Stol, Eds., Berlin, Heidelberg: Springer-Verlag, 2013, pp. 152–164.

[32] S. A. Chowdhery, M. Bertoni, J. Wall, and T. Larsson, "A data-driven design framework for early stage PSS design exploration", Design Science, 2020.

[33] G. Mariscal, Ó. Marbán, and C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies", The Knowledge Engineering Review, vol. 25, no. 2, pp. 137–166, 2010, doi: 10.1017/S0269888910000032.

[34] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: a systematic literature review", PeerJ Computer Science, vol. 6, e267, 2020, doi: 10.7717/peerj-cs.267.

[35] P. Chapman et al., "CRISP-DM 1.0 – Step-by-step data mining guide", SPSS, vol. 9, 2000.

[36] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining", Journal of Data Warehousing, vol. 5, no. 4, pp. 13–22, 2000.

[37] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", AIMag, vol. 17, no. 3, p. 37, 1996, doi: 10.1609/aimag.v17i3.1230.

[38] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data", Communications of the ACM, vol. 39, no. 11, pp. 27–34, 1996, doi: 10.1145/240455.240464.

[39] VDI/VDE 3714 Blatt 1, Implementation and operation of big data applications in the manufacturing industry – Implementation of big data projects. Berlin: Beuth Verlag, 2020.

[40] Project Management Institute, A Guide to the Project Management Body of Knowledge (PMBOK® Guide), 6th ed. Newtown Square, PA: Project Management Institute, 2017.

[41] T. Costello, "RACI—Getting Projects "Unstuck"", IT Professional, vol. 14, no. 2, 64-63, 2012, doi: 10.1109/mitp.2012.41.

[42] J. Gausemeier, R. Dumitrescu, J. Echterfeld, T. Pfänder, D. Steffen, and F. Thielemann, Innovationen für die Märkte von morgen: Strategische Planung von Produkten, Dienstleistungen und Geschäftsmodellen. München: Hanser, 2019.

[43] R. G. Cooper, "The drivers of success in new-product development", Industrial Marketing Management, vol. 76, pp. 36–47, 2019, doi: 10.1016/j.indmarman.2018.07.005.