31st Annual INCOSE international symposium
virtual event
July 17 - 22, 2021

Unlocking the power of big data within the early design phase of the new product development process.

Haytham B. Ali
University of South-Eastern Norway
haytham.ali@usn.no

Fredrik H. Helgesen
University of South-Eastern Norway
fredrikhhelgesen@gmail.com

Kristin Falk
University of South-Eastern Norway
Kristin.Falk@usn.no

Abstract. This study investigates how to exploit big data to increase knowledge in the early design phase within new product development through an industrial application. There are limited research and a scarcity of empirical studies that show the use of big data within new product development in manufacturing industries. Shorter design cycle demand, rapid decision-making, and the need for data-driven methodology are evident. An increase of knowledge through big data analytics, closing the loop with a knowledge base, has become a critical success criterion within the various industries.

This paper reviews the state-of-the-art in academic literature and investigates a case company. The case uncovered a gap of limited feedback into the early design phase. We developed a generic agile approach to extract value through analyzing big data. To fill in the identified gap, we tested our approach on a sample of big data, including both internal and external user data. Positive feedback from a survey complemented by interviews indicates that our method can aid decision-making within the early design phase by acquiring a more data-driven methodology.

Introduction

In a closed-loop with a knowledge base, big data analytics has become an emerging trend in extracting information patterns. These patterns can be fed back into the New Product development (NPD) process to increase knowledge. The paper highlights the “as-is” situation from an academic “state-of-the-art perspective” and presents an industrial case.

The case is a medium-sized Scandinavian supplier of high-tech components (Company). The manufacturing Company supplies a range of those components to large international systems suppliers. The manufacturing Company’s NPD process covers the complete process from receiving an order to market launch and then phase out. This study mainly focuses on the early phase. It includes the following steps in the NPD process: Design preparation, selection of critical design aspects, and evaluation.

Problem statement. The Company identified a need to integrate big data analytics into its knowledge base to shorten its development cycles. The Company did not use available data to support decisions in the early design phase within the NPD process due to the datas’ complexity. The data varied from structured to unstructured data. Besides, data lacked structure and sufficiency. There was inadequate documentation of what is necessary to be able to exploit the data. We expected that decisions made in the early design phase could be supported by exploiting external data. This expectation was the starting point of our research, and this paper explores the following research question:

How to exploit big data to offer more fact-based design decisions within new product development?

For the case, we find that the manufacturing Company can further enhance design decisions in the early phase of their product development by identifying a pattern between multiple data sources. We recommend using a data-driven feedback loop. Data sources typically include both internal and external data.

Content. The first section of this paper highlights the theoretical foundation. It consists of the following sub-sections: the knowledge framework, big data definition, knowledge increase in new product development, a review of best practices, and customer involvement through user data. There is then a section with the research methodology. Next, the paper identifies gaps by analyzing the “as-is” situation - current practice in the Company and proposing an approach to cover this gap.

Furthermore, the paper shows how we tested our approach on a sample of big data. We evaluated our approach through a Likert scale survey and interviews. The article provides a thorough discussion, including a critical reflection, financial perspective, and suggestions for further research. Finally, the study wraps up with a conclusion.

Theoretical Framework

This section introduces customers as a source of data. It also reviews the state-of-the-art in academic literature on “how to exploit big data in general and customer feedback data in particular.” Furthermore, we look at conceptualization that is how to transform this data into knowledge. We base this transformation on the knowledge management model: Data, Information, Knowledge, and Wisdom (DIKW) hierarchy. Further, we highlight how this knowledge relates to NPD through a Systems Engineering approach, and we present our definition of big data.

The Knowledge Framework: Data, information, knowledge, and wisdom

This study bases data, information, knowledge, and wisdom definitions on the DIKW hierarchy. This hierarchy is a widely recognized concept that constructs our knowledge management model. Figure 2 illustrates the five different layers in the DIKW hierarchy and a conceptualization of the knowledge management model. (Rajpathak & Narsingpurkar, 2013; Unhelkar, 2017) states that this model is decisive in efficiently managing big data to further make it exploitable in decision-making within the NPD process.

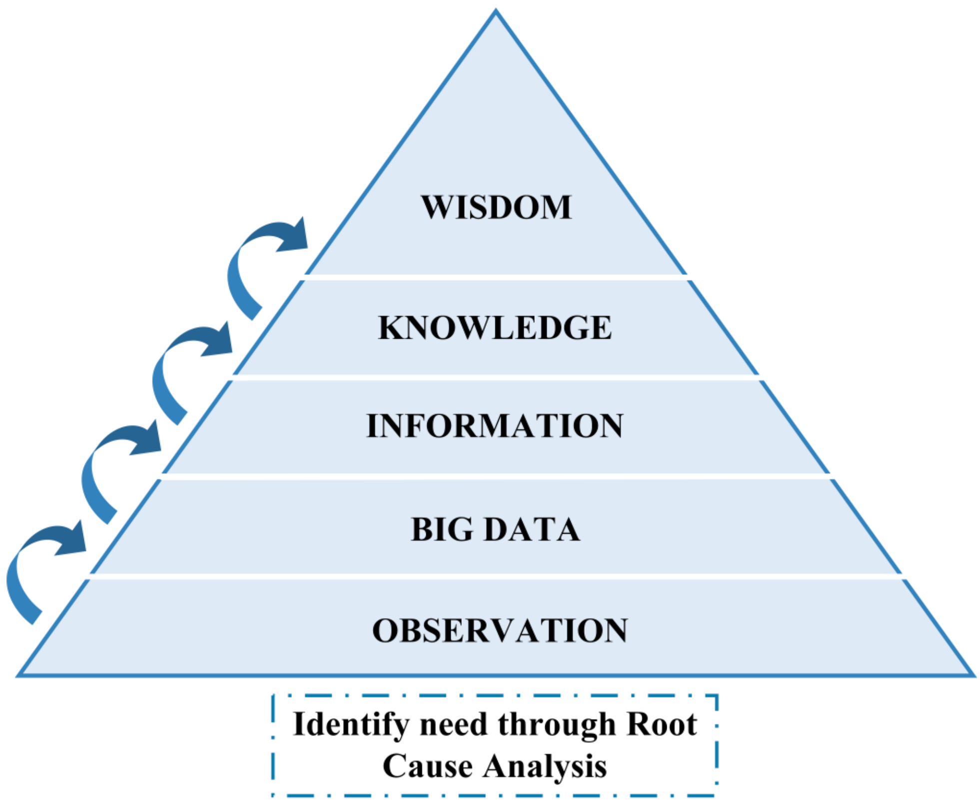

Figure 1. DIKW hierarchy, from observation to wisdom based on Rajpathak & Narsingpurkar, 2013; Unhelkar (2017)

In the context of this study, observations are facts collected and based on the research question derived from a Root Cause Analysis (RCA). Generally, observations are highly affected by the understanding of the observer. The observer usually complements the observations through interviews. A collection of observations constructs data. *Data* varies from structured data (data organized in rows and columns) to unstructured data (such as images). The data we collected from Company includes a combination of both; structured and unstructured data. *Information* is extracted data that is processed to be valuable. This data aims to answer the questions: "who," "what," "where," and "when." *Knowledge* emerges from the application of data and information by answering "how" questions. Knowledge and its reuse form the foundation of the NPD process.

Unhelkar (2017) describes an agile knowledge management model that aims to shrink the NPD project time and increase customer satisfaction by improving product quality. According to Ackoff (1989), the development process usually utilizes a combination of data, information, and knowledge. Wisdom is the integrated knowledge that adds value to the organization through best practices and lessons learned. Expertise and insight of the Company's key persons especially initiate wisdom. (Rajpathak & Narsingpurkar, 2013; Ackoff, 1989).

Big data definition

The term big data has become broadly used and has several definitions. Many authors and practitioners have used the notion (V) to define big data. Kwon and Sim (2013); Russom et al. (2011); Laney (2001) define big data in terms of the 3Vs: Volume, Velocity, and Variety. Gantz and Reinsel (2012); Dijcks (2012); Gogia et al. (2012) have extended the definition by adding value as the fourth V (4Vs). Besides, White (2012) suggested adding Veracity as the fifth V (5Vs). In this context, "Volume" refers to the vast amount of data. "Velocity" refers to the speed at which new data is generated, whereas "Variety" represents different types of data. The fourth V, "Value," refers to how we can benefit from big data by turning it into value. "Veracity" includes biases in the data and strives to encompass the level of sufficiency or insufficiency of the data.

Besides the five dimensions, The Method for an Integrated Knowledge Environment (MIKE2.0) project introduces a seemingly contradictory notion of big data: “Big Data can be very small and not all large datasets are big” (MIKE2.0, n.d.). This definition attributes complexity and not size as the dominant factor. However, according to a more anecdotal description, big data is beyond conventional tools’ capability to process. NIST supports this idea and states that big data is data that ‘‘exceed(s) the capacity or capability of current or conventional methods and systems” (NIST, 2017).

Complexity and data size are the main characteristics of the Company’s data. Complexity in terms of Variety (structured and unstructured data) and Veracity (insufficiency and lack of structure). In this study, we, therefore, define big data based on both literature review and on-site observation in Company as the following:

“Big data refers to datasets whose size or complexity exceeds the capability or capacity of current or conventional methods or data management systems in Company.”

Knowledge increase in new product development

Systems Engineering proposes a systematic approach to enhance design options and develop more cost-effective systems based on system thinking (Baxter, 2011; Moser, 2013; Camelia & Ferris, 2016). Systems Engineering aids in identifying gaps through applying an interdisciplinary and holistic approach. In general, integrating big data into various engineering processes, including NPD processes, can fill these identified gaps,

Figure 2 shows the design process paradox (Ullman, 2010). As the Figure indicates, knowledge increases as the project progresses. Unfortunately, design freedom disappears when knowledge rises to its peak.

Thus, it is crucial to consider the whole product life cycle in the NPD process, especially design decision-making, for two main reasons (Ullman, 2010). Firstly, making changes in the later phase rather than in the early phase of the product life cycle results in exponentially increased costs. 80% of a product’s cost can be determined through decisions made during the design phase (Duverlie & Castelain, 1999). Secondly, the product life cycle affects both product quality and customer satisfaction (Ullman, 2010).

Integrating available big data in decision-making contributes to cost reduction, improved product design, quality, and customer satisfaction (Rajpathak & Narsingpurkar, 2013). Additionally, Rajpathak & Narsingpurkar (2013) argue that big data analytics can close the loop with the knowledge base by extracting information patterns that can provide feedback into the design.

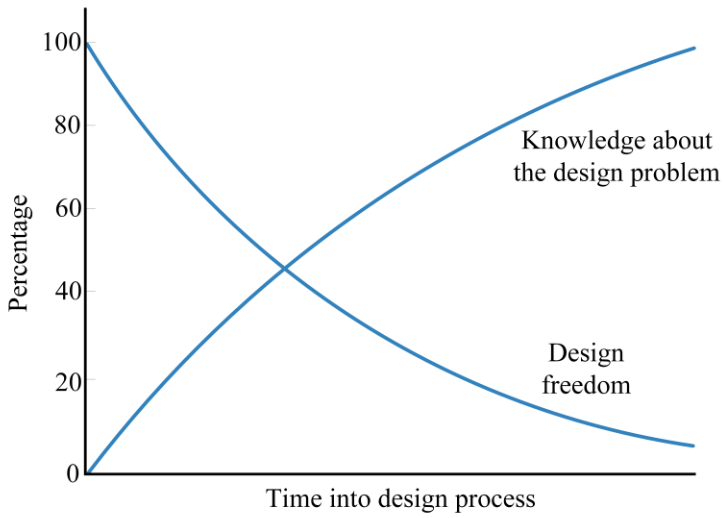

Figure 2. Design process paradox Ullman (2010)

Reviewing state of the art in the industry

We conducted a literature review to gain an insight into how best-in-class high-tech companies benefit from their stored big data. Companies' data are generally not available in a simple format but dispersed across the whole enterprise. Thus, companies need a well-defined strategy to collect, store, synthesize, and disseminate this data in the form of knowledge (Rajpathak & Narsingpurkar, 2013).

To derive maximum business insights from big data, companies' IT infrastructure needs to be modified to manage complex and large volume data. Thus, organizations need to have appropriate insight into their data and a clear strategy for merging it with their IT systems. Furthermore, it is crucial to establish a close collaboration between the large international industry manufacturers and suppliers to increase product quality. Such cooperation is achieved by sharing information and knowledge through cooperative product lifecycle management (PLM) systems. Thus, holistically understanding the value chain is essential (Rajpathak & Narsingpurkar, 2013).

The top U.S.-based manufacturers within the aerospace, automotive, and high-tech electronics industries spent $26 billion on warranty claims. These warranty claims are the all-industry average cost between 2003-2018 (Week, 2019). Best-in-class manufacturers can capture data generated from warranty claims, quality testing, and diagnosis. By converting this data into a valuable knowledge base, they can exploit warranty claims as additional feedback to the R&D department. The collected data is supplemented by further analysis to identify correlations in patterns to enhance the NPD process.

Customer involvement through customer feedback.

There is little research, and an especial scarcity of empirical studies, in utilizing big data within new product development in manufacturing industries (Tiwari, Wee, & Daryanto, 2017; O'Donovan, Leahy, Bruton, & O'Sullivan, 2015; R. G. Cooper & Edgett, 2012). The literature reveals a need to focus not only on the front-end but also on back-end data. For instance, utilizing customer feedback to increase knowledge further supports design decisions (Siva, 2012).

(Tatikonda & Rosenthal, 2000) states that the new product development process, especially the design phase, needs significant information processing from raw data. This data is generally found in large volumes dispersed across the Company. Value is created by intelligently extracting necessary and accurate information from this data that further substantiates more agile and fact-based decisions (DNV, 2016). A survey published by IBM innovation mentioned that using big data and analytics increases, by 36%, the probability for companies to be more successful than their competitors (Marshall, Mueck, & Shockley, 2015).

To a large extent, customer involvement has replaced the role of traditional R&D in the creation and marketing of new design concepts in new product development. This replacement is a consequence of the time-consuming feedback loop between expenditure and production cycles. (Rajpathak & Narsingpurkar, 2013). Cooper proposed customer inclusion as input, and feedback at every step, throughout the whole NPD process (Cooper, 1993). Companies can establish this feedback through, for example, integrating warranty data as back-end data. Through this integration, companies can reduce R&D expenses by understanding unspoken customer needs.

However, companies also collect customer data during the processing of warranty claims and Post-sale services. An analysis of this data can reveal information pertinent to both the product and the customer, such as mode of failure, defective components, usage at failure, customer usage, operating environment, intensity, and maintenance (Siva, 2012).

The large international industry manufactures, and their manufacturing suppliers can utilize this customer feedback to create problem-solving tools. Companies can establish such tools by first gathering the data and analyzing it to identify the problem. This problem can then be narrowed down to its source, i.e., customer-related, production-related, or design-related. Furthermore, companies should focus on solving the issue in isolation to create a better product for the customer (O'Donovan et al., 2015).

Research Methodology

The research methodology is based on industry-as-laboratory as we stayed in the Company while doing this research (Potts, 1993). We started relatively broad and tested our method on a specific case. The study consists of five significant steps: a preliminary study, current practice in the Company, data-analysis model, iterative case-based data-analysis for specific sub-system (Part), and verification and validation of the data-analysis model. Figure 3 shows a schematic view of our research process.

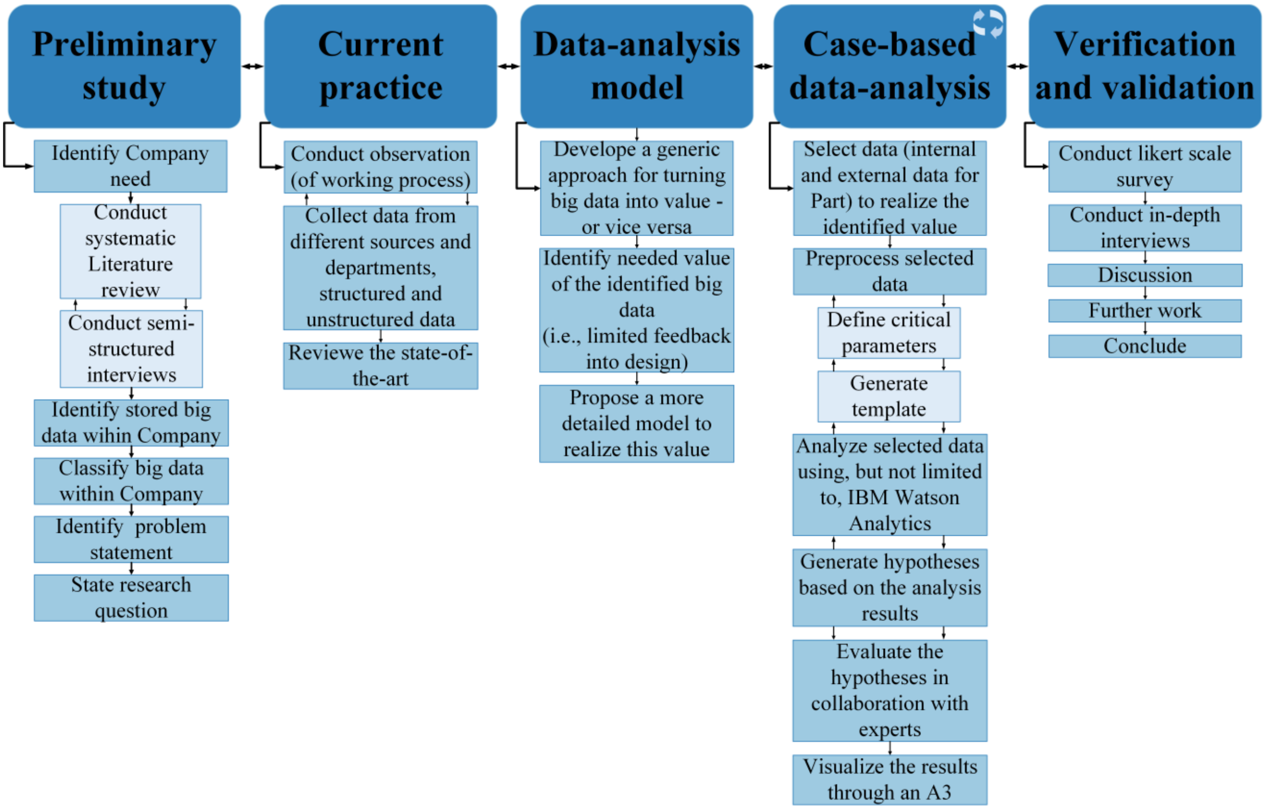

Figure 3. The research approach visualized by the performed steps

Preliminary study. As preparation, we conducted a preliminary study. We started by trying to understand the Company's needs and conducted a systematic literature review and semi-structured interviews. We identified and classified sources of big data within the Company. This study enabled us to shape the problem statement, as well as determine the research question.

Current practice. To define the current practice, we applied unstructured and semi-structured observations. That is, we asked both specific and open-ended questions simultaneously as we worked in the Company's office on the project for almost a year. This type of observation provided us with a thorough understanding of the Company's current state, which helped us identifying business opportunities within the Company. We collected archive data from Company's databases and departments that constitute our big data. This data included both external and internal data, which also was both digitally and not digitally documents. However, the collected data vary between structured and unstructured data. We reviewed state-of-the-art continuously during the year of observations while we were collecting data and conducting the observations.

Propose a model for data analysis. We develop a generic agile approach to turn big data in the Company to value or vice versa. Then, we identified a needed value (i.e., limited feedback into design) based on observing the current practice and the identified big data within the Company. Based on the generic agile approach, we developed a more detailed model to realize the recognized value.

Data analysis & case. Further, we selected a specific case that we analyzed in detail. The case or Part had a sample of identified big data within the Company. This data was needed to realize the recognized value. We preprocessed and then analyzed the sample. This sample's size is mentioned under the subsection "Exemplifying our approach on a Case" as the 5Vs make it characterized as big data, emphasizing Variety (Mohanty & Srivatsa, 2013). To generate and formulate the hypotheses, we continuously iterated all steps from the "State research question (the last step in the preliminary study) to " Analyze selected data using, but not limited to, IBM Watson Analytics." These iterations consisted of reviewing the state-of-the-art of academic literature. Further, we evaluated these hypotheses in collaboration with experts within the Company before visualizing the analysis results through an A3. Iterating the process at least three times allowed us to narrow down from a broad understanding to concrete research claims (Riel, 2010).

Verification and validation. Our study conducted a five-point Likert scale survey to verify and validate our recommendations (Jamieson et al., 2004). We collected input from 12 respondents, consisting of a diverse distribution of disciplines involved in the early design phase in NPD, three Test Engineers, three Engineering Managers, four Project Engineers, and two Quality Engineers. The survey consisted of the following six elements:

1. Personal review of the respondents – qualitative
2. Verification of identified gaps of current practice in Company – quantitative (Likert scale)
3. Measurement of current insight regarding big data – qualitative

   Presentation of the results (here we presented the results including, but not limited to, figure 6 to figure 9 before we were continuing the survey)

4. Measurement of the new insight regarding big data – qualitative
5. The respondent's opinions about the use of big data – quantitative (Likert scale) & qualitative

To evaluate the survey, we applied Net Promoter Score (NPS). We calculate the NPS scores using the following formula (Muller, 2013):

NPS = #strongly agree - (#neutral + #disagree + #strongly disagree)

Gillham (2008) stated that one of the main limitations when applying observations is the need for a complementary research method. To close the loop of the study, we supported the survey by interviews. We conducted in-depth, semi-structured interviews with an open framework. This type of interview can provide us with a focused, conversational, and two-way communication dialog to explore issues in-depth and to a further extent (Boyce, Carolyn, Neale, & Palena, 2006). Hence, both qualitative and quantitative data support our conclusion.

We applied Root Cause Analysis (RCA) through the ''Five Whys" technique. This research adapted this technique by repeating the question ''why?" five times in the context of interviews, observations, and meetings with the Company. The ''Five Whys" technique enabled us to identify the root cause of an event, to offer further realizable and implementable recommendations (Andersen & Fagerhaug, 2006).

This research made use of a sample of stored data from the Company. There are two sources of data: internal data and external data (in the form of user data). Due to limited time and resources, we have focused our attention on one customer and one system only. In the further, the paper refers to this system as Part.

Results and Analysis

One of our main results is the architecture model. This model describes and visualizes current practice within the early phase of the NPD process in the Company. The model considers gaps that we have identified. This study covers what we identify with the main gap; limited feedback of external data into design.

The current practice of new product development process in Company – related to the use of data

This study focuses on the early design phase of the NPD process. Based on our research, we have developed a model for the "as-is"-situation - the current practice of the NPD process in the Company – related to the use of data, Figure 4.

When Company receives an order from its customers, Company initiates the first phase of the NPD process, i.e., know-how (a feasibility study). This phase encompasses two steps: (1) Design preparation, including identification of design aspects; (2) Selection of critical design aspects, including gathering facts with regards to identified tensions and conflicts and building miniature models, e.g. ''prototyping", based on stakeholder requirements (Heemels, vd Waal, & Muller, 2006). To evaluate the critical design aspects, Company verifies these aspects through testing. The "Evaluation" supports the ''know-how (feasibility study)" phase through a feedback loop including data, models, and experience.

We have identified two gaps within the early phase of the Company's NPD process. This study focuses on the primary gap.

Primary gap: Limited feedback of external data into the design phase. The persons making the design decisions within the R&D department in Company are not adequately introduced to external data. These decision-makers include designers, managers, systems engineers, and functional safety specialists. Our observations reveal that external data is available in the aftermarket department. However, it is not dispersed across the Company.

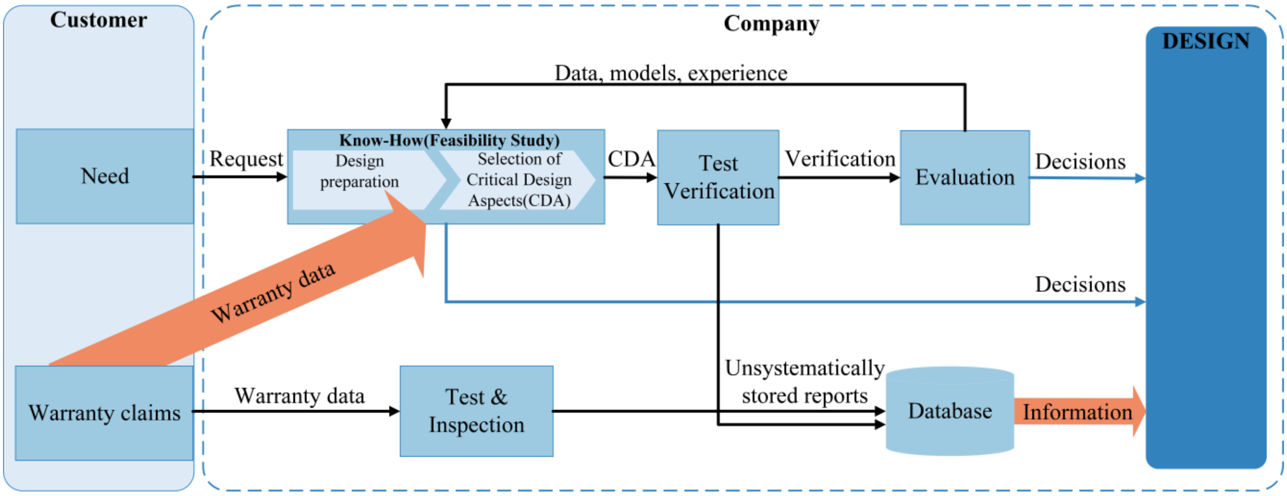

Figure 4. The current practice of the NPD process in the Company – related to the use of data

Secondary gap: Limited use of stored data in the Company's database. Employees store reports generated from test-verification results in an internal database within the Company. Additionally, this database includes test and inspection reports. However, our observations indicate a lack of guidelines in storing information. We have also observed that this lack of guidelines makes it difficult to retrieve and trace stored data. Employees share knowledge orally. The mentioned phenomena (missing guidelines, lack of traceability, and lack of retrievability) results in the following challenges: (1) Limited use of stored data; (2) Unstructured stored data. These challenges lead to our secondary identified gap. Additionally, we have noticed that some key persons establish their own Excel sheets as supportive tools to trace the documents' history. Figure 4 strives to highlight this gap through the orange arrow connecting "Database" and "Design."

We have iteratively reviewed state-of-the-art and listed the results of best practices. Then, we have mapped these results to Company's current practice. This process, alongside our on-site observation, has led us to suggest integrating external data as an input to "know-how" to cover the primarily identified gap; limited feedback of external data into the design phase. The orange arrow represents external data in Figure 4 and visualizes how to cover the identified gap. By mapping external data into the "know-how" in terms of feedback into the design, we expect that knowledge will increase within the early design phase, as Ullman (2010) describes. The orange curve in Figure 5 shows a "To-Be" state of this knowledge increase.

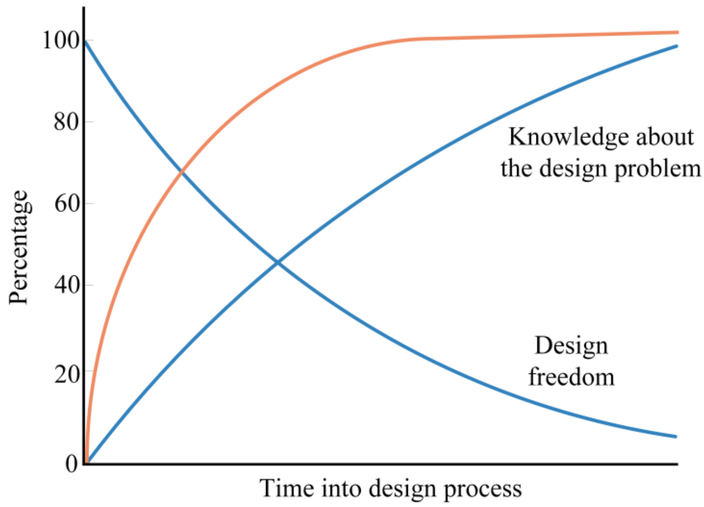

Figure 5. "To-Be" state of knowledge increase to break the design process paradox Ullman (2010)

Turning big data into value - or vice versa

This section highlights a generic approach to exploit big data. Based on our approach, we introduce a more detailed model on how to utilize external data. We exemplify the detailed model on a sample of big data to demonstrate our approach's adaptability and value.

A generic approach. We develop a generic agile approach to exploit big data in Company, which the DIKW model greatly inspires. Figure 6 shows a two-way decomposition approach. This framework provides an agile approach to exploit big data. Hence, we can apply the model in both top-down and bottom-up manners. For instance, we apply the top-down approach when we have identified stored and available big data. The aim is to add value to the Compay from the stored data.

In contrast, when we have already identified a required value of the big data (e.g., covering identified gap(s)), we apply the bottom-up approach. The iterative circle between "preprocess" and "analyze & visualize" in Figure 6 illustrates that we propose several iterations. We repeat this step until a value is realized.

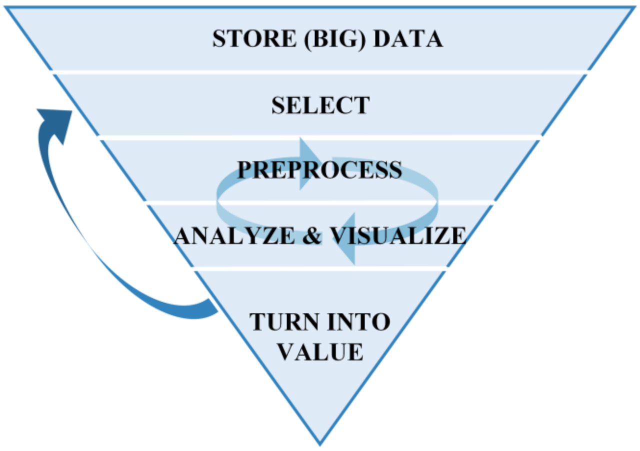

Figure 6. A generic approach

A more detailed model. Based on the generic agile approach, we developed a more detailed model. Figure 4 visualizes the primary gap of limited feedback of external data into the design phase. The architecture model we present in Figure 7 strives to cover this gap by establishing a feedback loop from external data into design.

The letters in Figure 7, A-K, represent the steps alphabetically ordered, whereas arrow-loops demonstrate iterative steps. Internal data contains a higher number of technical details than external data. We vertically visualize this through the level of detail in Figure 7 (Muller, 2005; Muller, 2011; Heemels, vd Waal, & Muller, 2006).

We collect external data received from customers, as the first step in Figure 7. Letter C brings us to the second step, data preprocessing. We refined external data through templates that include only critical parameters. Through letter D, we feed these templates into a data analysis tool. Based on the output of the data analysis, we generate hypotheses.

To support and further evaluate these hypotheses, we need to include a second source of data. For this purpose, letter E leads us to collect internal data generated in Company. Internal data includes a higher level of detail than external data. By defining and selecting critical parameters, we thoroughly refined this data. Like external data, we developed templates based on the defining and selecting parameters from internal data - see letter F in Figure 7. Therefore, we continue the process through letter G by feeding the refined internal data into an analysis tool. The hypothesis generated from the internal data can now support and evaluate the external data's hypothesis. Additionally, we have evaluated the hypothesis in collaboration with experts within the Company.

Through letter H, we visualize the results, including the hypothesis, through an A3 overview. This A3 overview allows us to extract value by efficiently detecting patterns among the results. We can integrate this A3 overview into Company's knowledge base by following the letter J. However, we follow the letter K in cases where the results add no value. This scenario (no added value) can occur due to many reasons; data insufficiency is the most significant. Thus, we need to conduct more iterations to preprocess the collected data or collect more data and then preprocess it. Letter K represents this in the model (see Figure 7). An additional iteration includes retrieving external data or internal data, depending on whether the data is sufficient to add value or not.

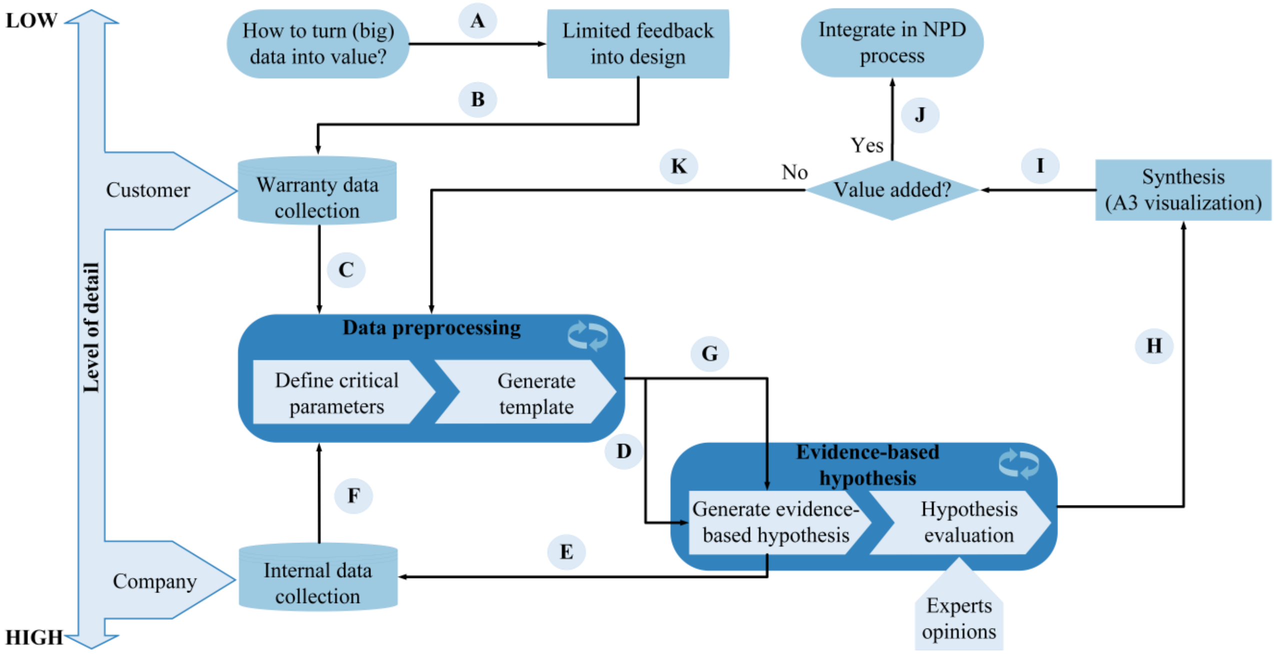

Figure 7. Exemplifying our generic approach through a more detailed model - From raw data to verified recommendation

Exemplifying our approach on a Case

We test the approach, explained in Figure 7, on a sample of big data. The sample consists of two data sources: external data (i.e., feedback from customers) and internal data. The two sources of data are composed of approximately 28 thousand rows and 48 columns. These columns and rows represent only the size of the structured data of our sample. The sample also includes unstructured data that we also collected and analyzed.

External data contains all warranty claims received from two customers in the last decade. Due to a limited scope in terms of resources and time, we narrowed down the sample even more. We focused only on the specific Part as our system of interest. To increase the data analysis quality, we preprocessed the data by selecting and defining critical parameters. We conducted, in total, 19 iterations to preprocess, analyze, and visualize the data. Through these iterations, we excluded one customer due to insufficient data. In this context, insufficient data means that some of the selected critical parameters were not documented thoroughly. These parameters include, but are not limited to, “production year,” “repair year,” and “delivery year” of the Part. Our sample of big data contained, after preprocessing, approximately five thousand rows and 21 columns. The 21 columns represent critical parameters. We further fed these parameters, together with their rows, into the data analysis through the generated templates.

We have primarily applied IBM Watson Analytics (IBMWA), data analysis and visualization service, to discover patterns and meaning in our sample of big data. After feeding Watson Analytics with the preprocessed data, we could gain an insight into specific patterns within this data and further generate hypotheses.

To evaluate the hypotheses, we continued to the next step (see letter E in Figure 7.) We collected internal data in terms of both inspection reports (reports generated by testing & inspecting external data), Engineering Change Notifications (ECN) data, and product line information, which includes process changes data. We extracted internal data from the Company’s database to preprocess it further with a similar approach as applied to the external data. We then analyzed and visualized the internal data using IBMWA. We managed to conduct the same process using the Microsoft Excel Analysis ToolPak. We also used this program to analyze external data.

We spent approximately 80% of seven weeks preprocessing our sample of big data. This time estimation also coincides with what is reported in the literature (Hoyt, Snider, Thompson, & Mantravadi, 2016). Various aspects can increase the time consumed in preprocessing data:

(1) Data is dispersed across the Company and is therefore not available within one database;

(2) Data contains different types of document formats, including Excel, Word, and PDF;

(3) Data has poor quality in terms of structure;

(4) Data includes insufficient parameters; and

(5) Data mapping -the different data sources included different names for the same parameter.

We generate four hypotheses based on the results of our sample of big data. We visualize the results, including the hypotheses, through an A3 overview. However, we have been mainly focusing on hypothesis H1. This hypothesis indicates that the number of warranty claims has a positive correlation related to total changes.

Figures 8 and 9 represent an extraction of our visualization of analyzing our sample of big data. This extraction shows the Part that has the highest warranty claims in the last decade for one customer. Figure 8 refers to one of our critical parameters – “production year.” The columns in the Figure represent claims percentage, sales percentage, and failure ratio. The years represent the production year for the Part that Company produced from the years 1 to 8. We calculated the percentage of the claims by dividing the number of warranty claims by the total number of warranty claims. As for the claim percentage, we calculated the sales percentage by dividing the sales quantity by the total sales. Failure ratio refers to the ratio of the number of warranty claims to sales quantity.

Figure 9 provides a timeline representing all changes, including changes in supplier, design, production, and process on the Part. We have developed this timeline based on internal data, particularly from ECN data and production line information data. The analysis of our sample of big data indicates that Parts produced in year 6 generate most warranty claims (see the column for year 6). These warranty claims decreased in the seventh production year (see the column for year 7).

Our analysis of internal data on Part indicates that the sum of all changes from years 1 to 6 has a positive correlation to the total number of customer feedback data in production year 6 (in the form of warranty claims). Besides, design changes in year 6 have a positive correlation to total data in production year 7. The analysis indicates that design changes interface with other changes such as supplier, production, and process. This correlation derives from design decisions’ cumulative effect with other changes such as supplier, process, and production changes. In other words, a change in design may consequently require changes in supplier, process, and production. Therefore, we conclude our big data analysis with our hypothesis indicating that design decisions positively correlate to external data.

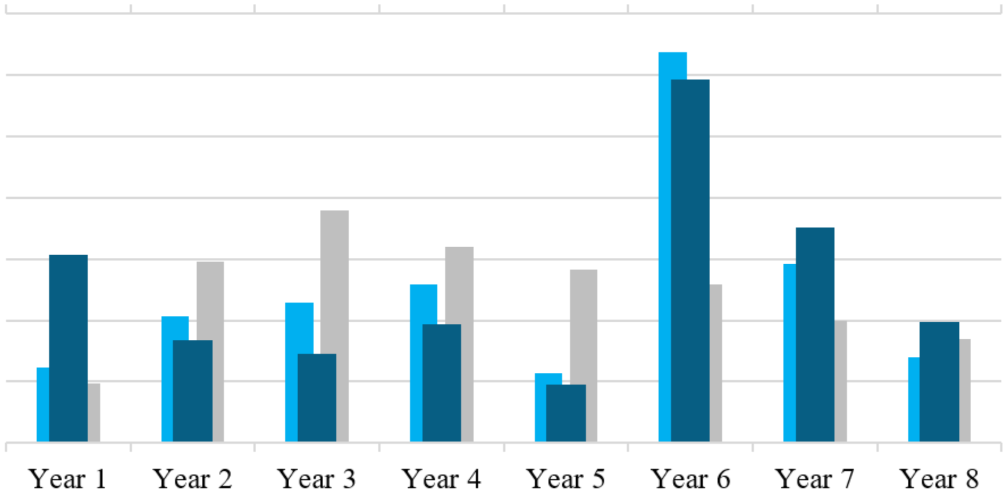

Figure 8. The yearly number of claims in percentage (light blue), failure ratio (dark blue), and sales in percentage (grey) for the Part. The first two bars have the same scale, where the last one (sales in percentage) does not have the same scale.

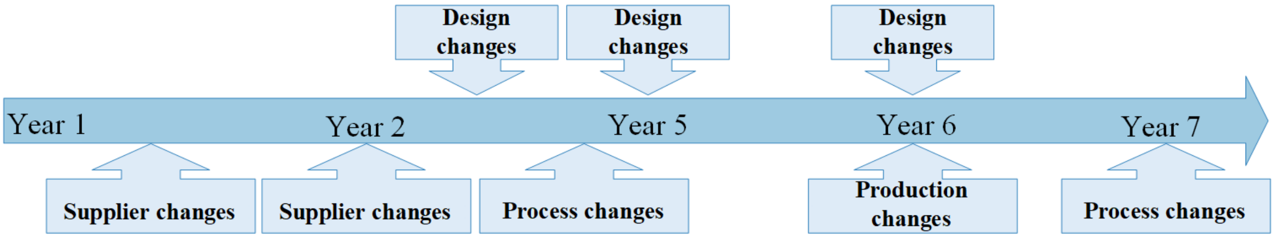

Figure 9. A timeline that evolved from ECN data for Part the last decade.

We have estimated that the cost of customer feedback is approximately 20 Million Norwegian Krone (MNOK). This cost includes only customer feedback on the Part received from one customer for the last decade.

Verification and validation of the findings

We conducted a survey supported by interviews to verify our results further.

Likert scale survey. This study conducted a Likert scale survey for two main reasons. The first reason was to verify the identified gaps in current practice for Company's NPD process. The second reason was to measure the "value" of the study.

Feedback from the survey resulted in a positive NPS for 71% of the statements, whereas the remaining 29% resulted in a negative NPS. A positive NPS indicates a high satisfaction level, whereas a negative NPS illustrates the opposite; unused potential. Even though we could not illustrate our approach's complexity and details due to the time constraint, the findings we presented have 39% more promoters than detractors. Where 45% of all replies were ‘‘strongly agree”.

Figure 10 shows replies from the survey. The two Likert scale questions 2.1 and 2.2 cover the verification of the identified gaps in the Company's current practice. The Figure shows that Question 2.1, "Limited feedback of warranty data… " has a negative NPS of -1. See Figure 10. The negative NPS of -1 indicates that there are more detractors than promoters for this statement. Question 2.2, “Limited use of stored data…” has a positive NPS of 6.

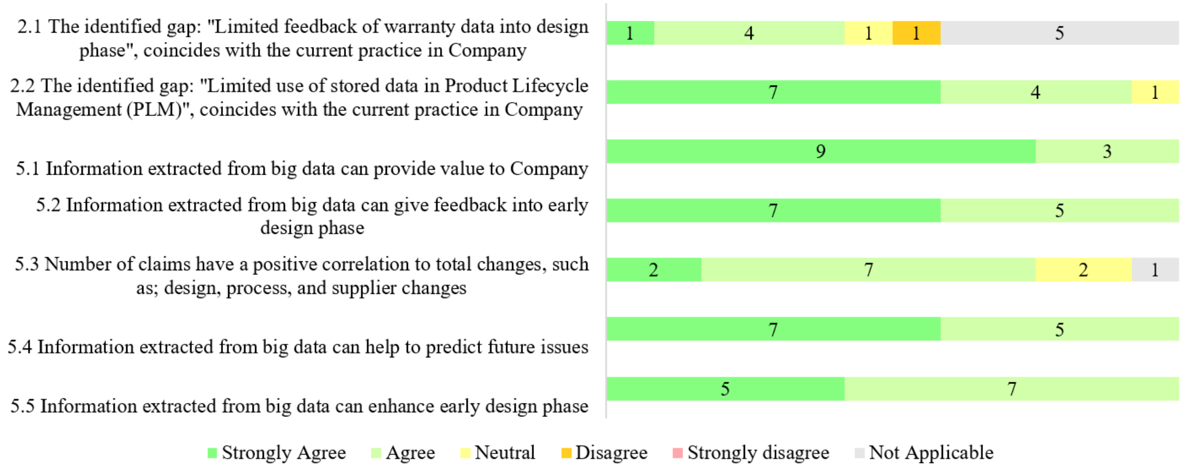

Figure 10. Net Promoter Score results from the Likert Scale questions

Table 1 proposes an overview of the increased knowledge of big data. Questions 3 and 4 (ref. Table 1) aim at measuring the difference of insight regarding big data before and after presenting the research, respectively. Table 1 shows the results of our perception of increased knowledge, in terms of new insight, in percent, after presenting our findings at a high level. We calculate this percentage as follows: perception of increased knowledge (%) = (Number of positive answers/Total number of respondents)

For instance, we analyzed the answers to questions 3.1 and 4.1 (ref. Table 1) by comparing the respondents’ answers and investigating if the respondents’ knowledge increased as an effect of presenting the research results between these two questions. For this example, we found that 10 out of a total of 12 answers showed an indication of increased knowledge (see Table 1). Thus, perception of increased knowledge (%) =(10/12) 100 = 83.3%.

Table 1: Increased insight regarding big data after a presentation of research findings

Question		Increased knowledge (%)	What
3.1	**4.1**	83,3	Complexity of data, difficult to analyse with available tools, unstructured data, data that can improve decisions, different types of data.
From your point of view, what is (big) data?			
3.2	**4.2**	58,3	Test & inspection reports, design, supplier & production changes, warranty data.
Which data can be considered as (big) data in Company?			
3.3	**4.3**	50	Discover trends, performance indication, monitoring, improve NPI and design phase.
How can available (big) data be exploited in Company?			
3.4	**4.4**	75	No failure found, part failure
List most common root cause failure description for Part			

The last part of the survey, part five, aimed at covering the respondents’ opinions regarding the use of big data in the Company. The feedback from the survey on this part of the results had a positive NPS for 80% of the statements, i.e., from 5.1 to 5.5 questions (see Figure 10). In contrast, the remaining 20% of results had a neutral NPS (cf. question 5.3 in 10).

An interpretation of the answers for question 5.6 ‘‘What do you think Company can get from big data in the early design phase?”) shows that all the respondents agree that exploitation of big data in the early design phase would yield value in terms of knowledge, reuse of knowledge, or both. Likewise, the answers for question 5.7 (‘‘From your point of view, what are potential challenges by implementing big data in Company?”), highlights the potential challenges of implementing big data in the Company. All respondents agreed on the necessity of a template to structure the defined critical parameters from the data of interest and the organization of commonly available data.

Closing the loop with interviews. To ‘‘close the loop”, we conducted in-depth interviews. We selected the most appropriate informants for the interviews, in particular four key persons in Company, with the following disciplines: (1) Chief Engineer; (2) Manager R&D; (3) Global Product Improvement Process (PIP) & Warranty Manager; (4) Functional Safety Specialist.

The purpose of the interviews was to provide further insights into the verification of the results. We covered the negative Part of the NPS (cf. questions 2.1 and 5.3 in Figure 10). All informants, through the interviews, agreed on the identified current gaps: (1) Limited feedback of external data into the design and (2) Limited use of PLM due to unsystematically stored reports. The informants have emphasized that they usually gather information, including documents, orally and through key persons. Due to the unsystematically stored documents in PLM, employees developed specific sheets to retrieve and trace required documents.

Furthermore, the informants agreed on the positive correlation between the number of claims and the total changes portrayed through the A3. The interviews end by discussing potential ways of integrating our approach in the NPD process. We review these discussions as a part of further work and discuss possibilities for Company improvements in the next Section - Discussion.

Discussion

Through our study, we have developed an approach for exploiting stored and available big data. Our observations have revealed multiple sources of data within the Company, such as feedback from users, customer feedback on prototypes, end of line test data, production line information, test reports, log data from testing, and ECN. We have focused on users’ feedback, complemented by internal data in terms of ECN, test & inspection reports, and production line information, to cover the identified gap by mapping external data into the NPD process.

Theoretical and industrial implications

Our research has both theoretical and industrial implications. We have found a scarcity in academic literature, especially empirical studies, regarding big data research. This scarcity includes a lack of using multiple sources of data within NPD. We have included two sources of data in our research; warranty claims in terms of external data and internal data. Many authors have mainly based their research on theoretical studies, whereas empirical research is of a minority.

We found a contradiction between the academic literature and the way the Company was storing data. The academic literature states that storing data takes place for a specific purpose. However, the emerging challenge for organizations is to derive meaningful insights from their stored data. We have observed that the Company collects and stores data without any specific purpose, except that they expect to exploit the data for value in the future. We have developed an approach for how to exploit this big data. Our study exemplifies our approach through a sample of big data and visualizes our results through an A3 overview. The Company can integrate this A3 into the early phase of their NPD process through, for example, knowledge briefs (see Appendix A).

We observed that the employees did not have a complete overview of which data is available within the Company through the preliminary study. Thus, we investigated what is available data within

Company and classified it within this study. While we are conducting our research, we have also observed that the Company's employees have become more concerned with big data as a term characterized by the 5Vs within the Company's stored big data. Mainly, value (one of the 5Vs) in terms of, but not limited to, applying big data analytics to study design choices and improving customer participation in the design in line with design innovation.

We have stimulated the employees through our work within the Company. For instance, the Aftermarket department has implemented a classification approach on external data by categorizing different customer-feedback data types to share across the Company. We have observed insufficient data within Company's received external data. In this context, insufficient data can be understood as data that is missing parameters, e.g., ‘‘production year”, ‘‘repair year,” or ‘‘delivery year” of a product. Thus, we suggest that Company collects more data from their customers. Also, Company has expressed its interest in implementing our approach in future projects. This implementation includes additional sources of data and processes.

Furthermore, we have applied "systems thinking" by identifying Company's gaps through our architecture model. This model visualizes current practice within Company's NPD process. "Systems thinking" has enabled us to establish a holistic understanding in addressing the complexity of integrating big data into the early phase of the NPD process. We have identified and mapped data sources and our main gap ("*Limited feedback of external data into the design phase*") through root cause analysis, models, loops, and layers. This study primarily used Microsoft Visio to model our figures.

Further, we applied both IBM Watson Analytics (IBMWA) and Microsoft Excel Analysis ToolPak to analyze big data. A brief comparison of these two tools reveals that they mainly differ in speed, functionality, and ease of use in IBMWA's favor. Thus, the Company can implement our approach and apply Excel without any licensing with IBM.

Further work.

This study exemplified an agile generic approach to exploit big data. We apply this exemplification through a more detailed model to cover the identified gap, i.e., "*Limited feedback of external data into the design phase*," using a sample of big data. This sample includes both external data (feedback from users and customers) and internal data. To increase our approach's validity and reliability, we suggest that further work should include other samples and sources of big data. Besides, the approach needs to be applied to other sources of data to cover similar gaps.

Conclusion

Our research question sought to investigate how to exploit big data to offer more fact-based design decisions within new product development. We analyzed the current practice – "as-is" situation in the Company through on-site observations and interviews, accompanied by reviewing the state-of-the-art in academic literature. We have identified a gap in integrating customer-feedback data into the early design phase.

To cover this gap, we have developed a generic agile top-down and bottom-up approach inspired by the DIKW-hierarchy. This study exemplifies our approach through a more detailed model on a sample of big data. Through this sample, we analyzed two sources of big data: internal data and external user data. The central hypothesis that emerged from analyzing this sample of big data indicates that design decisions positively correlate to external data. This correlation derives from the cumulative effect of design decisions on other changes, such as supplier, process, and production changes. Positive feedback from a survey, complemented by interviews, indicates that our approach can enhance early-phase decision-making. Through our contribution, we can achieve this enhancement by closing the loop with a knowledge base. Thus, our results emphasize the crucial role of the big data analytics closing loop, with a knowledge base trend.

References

Ackoff, R. L. (1989). From data to wisdom. Journal of applied systems analysis, 16(1), 3-9.

Andersen, B., & Fagerhaug, T. (2006). Root cause analysis: simplified tools and techniques. Quality Press.

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. Interacting with Computers, 23(1), 4-17.

Boyce, C., & Neale, P. (2006). Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input.

Cooper, R. G. (1993). Winning at new products : accelerating the process from idea to launch (2nd ed., pp. XIII, 358). Addison-Wesley.

Cooper, R. G., & Edgett, S. J. (2012). Best practices in the idea-to-launch process and its governance. Research-Technology Management, 55(2), 43-54.

Camelia, F., & Ferris, T. L. (2016, July). Systems thinking in systems engineering. In INCOSE International Symposium (Vol. 26, No. 1, pp. 1657-1674).

DNV, G. (2016). Are you able to leverage big data to boost your productivity and value creation.

Dijcks, J. (2012). Oracle: Big data for the enterprise. Oracle White Paper,(June), 16.

Duverlie, P., & Castelain, J. M. (1999). Cost estimation during design step: parametric method versus case based reasoning method. The international journal of advanced manufacturing technology, 15(12), 895-906.

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future, 2007(2012), 1-16.

Gillham, B. (2008). Observation techniques: structured to unstructured. Education Review.

Gogia, S., Barnes, M., Evelson, B., Hopkins, B., Kisker, H., Yuhanna, N., ... & Malholtra, R. (2012). The Big Deal About Big Data For Customer Engagement: Business Leaders Must Lead Big Data Initiatives To Derive Value [verkkodokumentti].[Viitattu 14.10. 2016] Saatavilla https://www. forrester. com/report/The+ Big+ Deal+ About+ Big+ Data+ For+ Customer+ Engagement/-. E-RES72241.

Heemels, W. P. M. H., vd Waal, E., & Muller, G. J. (2006). A multi-disciplinary and model-based design methodology for high-tech systems. Proceedings of CSER.

Hoyt, R. E., Snider, D. H., Thompson, C. J., & Mantravadi, S. (2016). IBM Watson analytics: automating visualization, descriptive, and predictive statistics. JMIR public health and surveillance, 2(2), e157.

Jamieson, S. (2004). Likert scales: How to (ab) use them?. Medical education, 38(12), 1217-1218.

Kwon, O., & Sim, J. M. (2013). Effects of data set features on the performances of classification algorithms. Expert Systems with Applications, 40(5), 1847-1857.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and Variety. META Group research note, 6(70), 1.

Marshall, A., Mueck, S., & Shockley, R. (2015). How leading organizations use big data and analytics to innovate. Strategy & Leadership.

MIKE2.0. (n.d.). Big data definition- the open-source standard for information management. Online: http://mike2.openmethodology.org/wiki/Big_Data_Definition. (Accessed: 14-11-2020).

Mohanty, S., Jagadeesh, M., & Srivatsa, H. (2013). Big data imperatives: Enterprise 'Big Data'warehouse, 'BI'implementations and analytics. Apress.

Moser, H. A. (2013). Systems engineering, systems thinking, and learning: a case study in space industry. Springer.

Muller, G. (2005). System architecting.

Muller, G. (2011, June). Systems architecting: A business perspective. In INCOSE International Symposium (Vol. 21, No. 1, pp. 1845-2142).

Muller, G. (2013). Systems engineering research methods. Procedia Computer Science, 16, 1092-1101.

NIST. (2019). NIST Big Data Public Working Group (NBD-PWG). Online:

https://bigdatawg.nist.gov/home.php. (Accessed: 14-11-2020)
O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. (2015). Big data in manufacturing: a systematic mapping study. Journal of Big Data, 2(1), 20.
Potts, C. (1993). Software-engineering research revisited. IEEE software, 10(5), 19-28.
Rajpathak, T., & Narsingpurkar, A. (2013). Managing knowledge from Big Data analytics in product development. White Paper, Tata Consultancy Services.
Riel, M. (2010). Understanding action research. Research Methods in the Social Sciences, 17(1), 89-96.
Russom, P. (2011). Big data analytics. TDWI best practices report, fourth quarter, 19(4), 1-34.
Siva, V. (2012). Improvement in Product Development: Use of back-end data to support upstream efforts of Robust Design Methodology. Quality Innovation Prosperity, Kristina Zgodavova, 16(2), 84.
Tatikonda, M. V., & Rosenthal, S. R. (2000). Technology novelty, project complexity, and product development project execution success: a deeper look at task uncertainty in product innovation. IEEE Transactions on engineering management, 47(1), 74-87.
Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, 319-330.
Ullman, D. G. (2010). The mechanical design process: Part 1. McGraw-Hill.
Unhelkar, B. (2017). Big Data Strategies for Agile Business. CRC Press.
Week, W. (2019). Auto parts supplier warranty report. Online: https://www.warrantyweek.com/archive/ww20190627.html. (Accessed: 14-11-2020)
White, M. (2012). Digital workplaces: Vision and reality. Business information review, 29(4), 205-214.

Biography

Haytham B. Ali is employed as Assistant Professor at the University of South-Eastern Norway (USN), where he is working at connecting engineering with science, focusing on mathematics. He holds a Master of Science in Systems Engineering with Industrial Economics degree and a Bachelor degree in Mechanical Engineering with a specialization in Product Development, both from USN.

Fredrik Hoel Helgesen is currently working as senior consultant at Capgemini, where he advise a variety of clients to enable digital transformations. He holds a master of Science in Systems Engineering with Industrial Economics degree from the University of South-Eastern Norway (USN). During the fall of 2017, he studied Industrial engineering at Universidad Pontificia Comillas ICAI. Further, he completed his Bachelor's degree in Automation Engineering at NTNU in Ålesund in June 2015.

Prof. **Kristin Falk** entitles a master's in applied mathematics and a Ph.D. in petroleum-production with a focus on simulations. She teaches at the University of South-Eastern Norway and is responsible for the research group in Systems Engineering. She has led technology teams in start-ups, SME and large corporations primarily in the energy industry, and interdisciplinary research projects in academia. Her research focus is 'how to create systems fit for purpose in a volatile, uncertain, complex, and ambiguous world.'

Appendices

Appendix A: Turning (big) data into value-A3visualization