THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Applying Design Analytics to Understand Engineering Change Request Information

Ívar Örn Arnarsson

CHALMERS

Department of Industrial and Materials Science
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

Applying Design Analytics to Understand Engineering Change Request Information
Ívar Örn Arnarsson

Report No. IMS-2018-10

Department of Industrial and Materials Science
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000
Email: varo@chalmers.se

Printed by
Chalmers Reproservice
Gothenburg, Sweden 2018

Trust me, I’m an Engineer.

- Every Engineer Ever

Applying Design Analytics to Understand Engineering Change Request Information

Ivar Örn Arnarsson
Department of Industrial and Materials Science
Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

Large complex system development projects take several years to carry out. Such projects involve hundreds of engineers who develop tens of thousands of parts and millions of lines of code. During the course of a project, many design decisions often need to be changed due to the emergence of new information. These changes are often well documented in databases but, due to the complexity of the data, few companies analyze engineering change requests (ECRs) in a comprehensive and structured fashion. ECRs are important and plentiful in the product development process in order to enhance a product.

This thesis sets out to explore the growing need of product developers for data expertise and analysis. Product developers are increasingly looking towards analytics for improvement opportunities within business processes and products. For this reason, we look at the three components necessary to perform data mining and data analytics: exploring and collecting ECR data, collecting domain knowledge towards ECR information needs and applying mathematical tools for solution design and implementation.

Results show two software tools including visuals of ECR text mining and design structure matrix. The tools were evaluated using industrial data showing patterns and improvement for products and process. Results also show a list of engineering information needs towards ECRs. New information derived with data mining and analytics can thus support product developers in making better decisions for new designs/re-designs of processes and products that lead to robust and superior products.

Keywords: Product Development, Engineering Change Request, Design Analytics, Design Structure Matrix, Markov chain.

ACKNOWLEDGEMENTS

The research was performed in collaboration between the Department of Industrial and Material Science (IMS) at Chalmers University of Technology in Gothenburg and Volvo Group Trucks Technology (Volvo) in Gothenburg. All of the work from 2015 has been financially supported by Volvo and in 2017 it also received financial support from the Swedish Governmental Agency for Innovation Systems (Vinnova). This support is gratefully acknowledged.

I’ve been surrounded by great people during the process of obtaining my Licentiate degree that have inspired me and supported in times of need. I would first like to thank my main supervisor Professor Johan Malmqvist, who has continuously provided me with guidance, ideas, visionary sketches and feedback, and my assistant supervisor Rikard Söderberg. Next, i would like to thank Mats Jirstrand and Emil Gustavsson at Fraunhofer-Chalmers Research Centre for Industrial Mathematics for collaboration and co-authorship. I would also like to thank everyone at the division of Product Development for a friendly and supportive environment.

I wish to express my deepest gratitude to my industrial sponsor Anders Ydergård, my industrial supervisor Lena Borg and my previous industrial supervisor Lars Börjesson for their corporate insight, guidance and on site support. I’m also greatful to all my colleges at my Volvo that I’ve had great pleaseure working with during the past three years.

I’m greatful for the excellenct language support I got from Gunilla Ramell in the final steps of this thesis.

Finally, I’m greatful for having my family and friends unconditional support in all situations.

Sincerely,

Ívar Örn Arnarsson

Gothenburg, Sweden, 2018.

APPENDED PUBLICATIONS

The following research papers form the foundation of this licentiate thesis.

Paper A

Arnarsson, I.Ö., Malmqvist, J., Gustavsson, E. and Jirstrand, M., 2016. *Towards Big-Data analysis of deviation and error reports in product development projects*. In Proceedings of NordDesign 2016, Trondheim, Norway (pp. 83-92).

Paper B

Arnarsson, Í.Ö., Gustavsson, E., Malmqvist, J. and Jirstrand, M., 2017. *Design analytics is the answer, but what questions would product developers like to have answered?*. In DS 87-7 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 7: Design Theory and Research Methodology, Vancouver, Canada, 21-25.08. 2017.

Paper C

Arnarsson, Í.Ö., Gustavsson, E., Malmqvist, J. and Jirstrand, M., 2018. *Applying Design Analytics to understand Engineering Change Request information*. Proceedings of the 15th International Design Conference on Excellence in Design (Design 2018), Dubrovnik, Croatia.

DISTRIBUTION OF WORK

The work for each paper was distributed among the authors in the following way:

Paper A Arnarsson coordinated the paper, performed literature review, analysis of data and wrote the paper. Malmqvist performed literature review and wrote the paper. Gustavsson coded the software demonstrator used to perform statistical analysis and wrote parts of the paper with the support of Jirstrand who also reviewed the paper.

Paper B Arnarsson planned and coordinated the study. He also conducted semi-structured interviews with AB Volvo, performed literature review and analysis in collaboration with Malmqvist. Emil Gustavsson performed statistical analysis. Arnarsson, Gustavsson, Malmqvist and Jirstrand wrote and reviewed the paper.

Paper C Arnarsson coordinated the paper and performed most of the literature review together with Malmqvist and Gustavsson. Gustavsson coded the Markov Chain DSM with domain knowledge support from Arnarsson. Arnarsson and Gustavsson drafted the paper with help of Malmqvist and Jirstrand who both reviewed the paper.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

DRM	Design Research Methodology
DSM	Design Structure Matrix
ECR	Engineering change request
MC	Markov chain
PD	Product Development
R&D	Research & Development
Volvo	Volvo Group Truck Technology

1 INTRODUCTION

Fast development of data collection and storage is pushing the need for data expertise and analysis. Companies are faced with large volumes of complex data from multiple sources and are looking increasingly towards analytics to improvement opportunities within business and their products (Wu et al., 2014). Product development areas are no exception with large complex development projects lasting several years. Hundreds of product developers making tens of thousands of parts and millions of lines of code that contribute to data growth that is difficult to analyze manually. Hence, there is an opportunity to research the application of data mining and analytic models on product development data to identify patterns and meaningful outputs that can support decision-making for new designs/re-design of products.

1.1 Background

Product changes in product development projects are often logged and stored in databases where structured data (i.e., numerical data) are mixed with unstructured data (i.e., text inserted by engineers). With the progress made within machine-learning and data mining, new techniques for retrieving insights from complex data sets have emerged.

Machine-learning is a form of analytical model in which algorithms are utilized in order to explore data and make predictions from data (for a survey, see Kotsiantis et al. (2007). Data mining is another closely related research field in which the aim is to detect patterns and knowledge in data sets (for a survey, see, e.g., Berkhin, 2006).

Machine-learning and data mining offer quantitative methods for performing data analysis with any system that is generating data. Once an overview of the data has been created, it can be identified and collected from databases, and data analytics can be employed in order to extract insights from historical data within companies (Zheng *et al*., 2014). The traditional way of analysis has been to try to find answers in data using manual exploration, for example by users who have exported data to a spreadsheet software tool and explored it according to their own methods. Now there are opportunities to make these explorations more effective and precise by automating the work. Moreover, data analysis can help shed some light on a variety of complex issues that would not be obtainable through “manual” inspection and analysis.

In the seminal book "Competing on Analytics: The New Science of Winning", Davenport and Jeanne (2007) describe data analytics as using statistical and

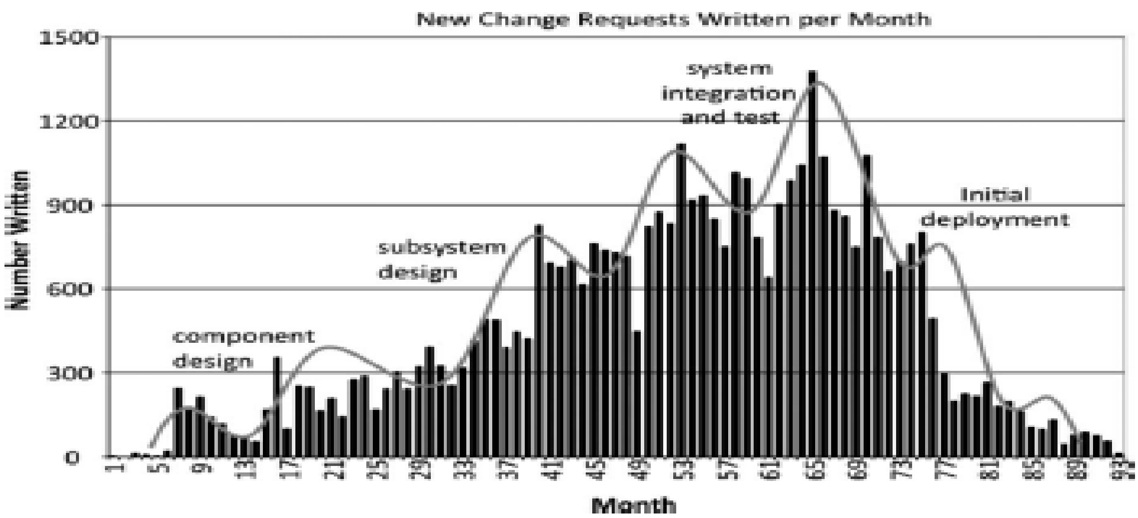

Figure 1. Frequency of change requests during a complex system development project (Giffin et al., 2009).

quantitative analysis of data, combined with explanatory and predictive modeling. The models and analyses provide the basis for fact-based management and decision-making. For a survey on data mining and knowledge discovery on a general level, see Han *et al*. (2011), and for a more technically-oriented survey describing techniques and machine-learning tools for data mining, see Witten et al. (2016).

The term “design analytics” has recently been proposed by Van Horn et al. (2012) to identify the area of research that focuses on processes and tools to enhance the transformation of design-related data into formats suitable for design decision-making. Examples include Tucker & Kim (2011) who applied analytics to consumer trend data in order to inform product design. Bae and Kim (2011) presented a study on how to optimize the development process of a digital camera by using data mining techniques on customer information. Also considering customer data, Lewis & Horn (2013) looked at customer behavior profiles and reflected upon customer needs in the late stages of the development process.

In conclusion, design analytics (data mining, machine-learning and modeling) can provide insights into a wide variety of information needs of many companies.

1.2 Focus on engineering change request data

During a project, many design decisions need to be changed due to the emergence of new information. It is known that changes late in the development process are very costly and run the risk of delaying the project (Clark and

Fujimoto, 1991). Unfortunately, the bulk of engineering needs for changes tends to be discovered late in the process, as shown in Figure 1, which depicts the amount of changes recorded per month through a complex development project (Giffin et al., 2009). These changes are often well documented in databases, but due to the complexity of the data, few companies analyze engineering changes in a comprehensive and structured fashion.

Causes of late changes in product development projects can result in a failure to meet objectives for budgeting, scheduling or technical performance. Weak leadership, lack of planning and rigid processes play important roles (Thomke & Reinertsen, 2012). The problem with late product changes has been known for several decades but recent studies confirm that it remains a challenge associated with high costs, quality problems and development lead time delays (Giffin et al., 2009; The Standish Group, 2014; Fernandes et al., 2015). The root causes of these changes are still poorly understood. It seems that mitigations that were proposed during the 1980's such as concurrent engineering and quality function deployment were not sufficient to solve the problem.

This Licentiate Thesis paper is focusing on engineering change requests (ECRs) or engineering change reports that reside in product development databases. The data itself is based on a large development project with a duration of several years at Volvo Group Trucks Technology (Volvo). Organizations try to continuously improve their products to stay ahead of the competition and retain quality. Product development projects often need to make an enhancement to a product and procedure to initiate such a change process, which is well-known under the name of ECR. The motive for making engineering changes can be an opportunity to improve, enhance or adapt a product (Pikosz & Malmqvist, 1998). ECRs within organizations contain information about desired product changes. The effects of such a change can then be evaluated and the best solution selected. Changes can occur throughout the entire product lifecycle from the concept phase to the after-market.

1.3 Research context

The research is performed at a multi-brand commercial vehicle company, with data from vehicle types such as trucks and buses. The company has R&D operations in more than ten countries globally.

1.4 Purpose and Goals

There exists some previous research within design analytics for product development based on customer data and project but the earlier research has not specifically considered ECR data.

1.4.1 Purpose

The purpose of this research project and licentiate thesis looks is to examine how historical ECR data can be analysed to gain new insights and patterns with which to improve the ECR process and support product developers in their daily work. This licentiate thesis contributes to the research of how companies can explore design analytical models on product development data and gain new insights with which to support the work performed in product development.

1.4.2 Scientific Goals

The scientific goals are as follow:

- Develop and implement methods to work with the analysis of product development data.
- Evaluate the need of product developers for data mining and identify beneficial outcomes.
- Propose, develop and implement an analytical model that matches organizational needs.
- Evaluate the effectiveness of the model together with product developers working in industrial development projects.

1.4.3 Industrial Goals

The intended industrial goals are as follow:

- Identify improvement areas based on research into applied data.
- Carry out a pilot project of the suggested improvements and a full scale implementation at Volvo.
- Decrease number of engineering changes.
- Perform data analysis using new insights for product developers and improvements of tools and processes around ECRs.
- Enhance efficacy of ECR that will correspond to cost savings in large development projects.

1.5 Research questions

The research focus was formulated to fit inside the research questions below with the goal of answering them throughout the licentiate thesis. The questions concern the usage of data in product development, perspectives of product developer and testing with analytical capabilities.

RQ1. Is it possible to use historical ECRs by applying data mining to gain new insights into data for product developers?

The first research question looks at the process of working with data from extraction to visualization. It further illustrates some data visualization and exploration ideas that can lead to new insights into ECR data.

RQ2. What information needs do product developers have regarding ECR data and what methods can support these needs?

The aim here is to gain domain-specific knowledge about product developer needs from ECR data and the kind of analysis they would like to see to help them in making better decisions in new product development projects. Also, to briefly reflect on methods or tools that can support analysis of ECR data in the way described by the product developers.

RQ3. What are the efforts, benefits and limitations of using the Markov chain Design Structure Matrix for ECR process analysis?

The research question explores a statistical probability method known as Markov chain that has potential for supporting information needs listed as outcomes of the previous research paper. These outcomes are then visualized in a Markov chain DSM matrix used for drawing conclusions on patterns and improvements together with product developers.

1.6 Delimitations of the research

ECR data produced in a product development project can be gathered from a broad source of information. This research focuses on the report data tracking the progress or overall case for change, leaving out any attached documents. Common documents that are attached to an ECR include pictures, drawings and tests results. The focus is on a single large development project with duration of several years.

1.7 Thesis structure

The structure of the licentiate thesis chapters is outlined as follows:

Chapter 1 introduces the topic, analyses of the problem and outlines the research questions.

Chapter 2 provides a framework for this research and the state-of-the-art literature related to the research topic. Further, the chapter identifies research gaps in the area.

Chapter 3 accounts for the research strategy and methodology used in this research.

Chapter 4 summarizes the results and findings of each paper that is appended.

Chapter 5 discusses the results in relation to goals and research questions.

Chapter 6 outlines the conclusions from earlier chapters and the future direction of this research.

Appendix includes the full versions of the three published papers that from the basis for this licentiate thesis.

> **Paper A** – Towards big-data analysis of deviation and error reports in product development projects.
>
> **Paper B** – Design analytics is the answer, but what questions would product developers like to have answered?
>
> **Paper C -** Applying Design Analytics to Understand Engineering Change Request Information.

2 FRAME OF REFERENCE

A theoretical framework for this research is presented which describes the general definitions and work performed in the fields of product development, engineering changes and design analytics.

2.1 Product development process

In engineering, product development covers all processes for bringing a new product or modifications of existing products to market. The incentive for product development is often the customers and satisfying them through new or additional benefits. Ulrich and Eppinger (2012) define product development as a couple of activities, beginning with the perception of a market opportunity and ending with production, sales and delivery of the product. A generic product development process proposed by Ulrich and Eppinger (2012), can be seen in Figure 2, where major activities in the process include planning, concept development, system design, detailed design, testing & refinement and production ramp-up. Similar methodologies for product development have been proposed by Hubka and Eder (1996), Pahl and Beitz (1996), Roozenberg and Eekels (1995), Ullman (1992) and Andreasen and Hein (1987).

Pahl and Beitz (1996) outline four main phases (Figure 4) in engineering design: product planning and clarification of the task, conceptual design, embodiment design and, finally, detail design. Hubka and Eder (1996), present a design process with a series of stages that creates information about the design.

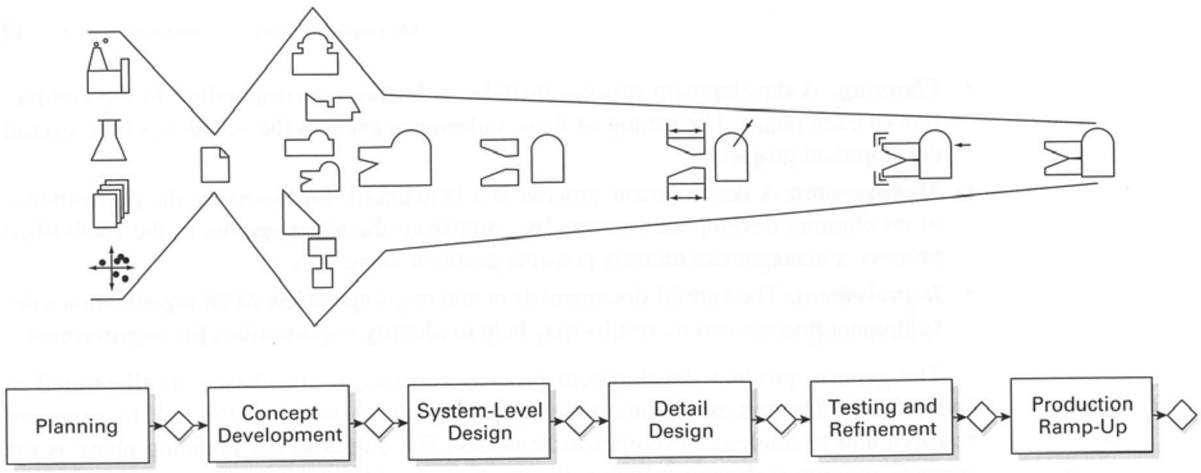

Figure 2. Generic product development process with six phases (Ulrich and Eppinger, 2012).

A variety of design process methodologies have also been proposed by Roozenburg and Eekels (1995) and Andreasen and Hein (1987). Ullman (1992) takes into consideration the manufacturing problems that may arise if manufacturing is not involved during the design process.

Product iterations are common as information flow is constant over the problem-solving phase. Pahl and Beitz (1996) highlight information that is processed using analysis and synthesis while developing a solution: concept, calculation, experiment, elaboration of drawing layout and evaluation of a solution. Iteration is a step-by-step process with which to approach a solution. Steps are repeated with a higher level of information based on the results of previous loops (Figure 3). In this way the solution can obtain information for refinement and ensure continuous improvement.

Prototypes are product iterations where time and money of building and evaluating the prototype must be weighed against anticipated benefits. Products high in risk and uncertainty due to the high cost of failure, new technology or revolutionary aspects should be considered for such prototyping (Ulrich and Eppinger, 2012).

There used to be great emphasis in the academic design literature on the design of novel products [Ulrich and Eppinger (2012), Otto and Wood (2001), Wright et al. (2005), Pahl and Beitz (1996) and Cross and Roy (1989)]. Around the last millennium, design reuse started to appear and has been showing up more frequently in literature ever since.

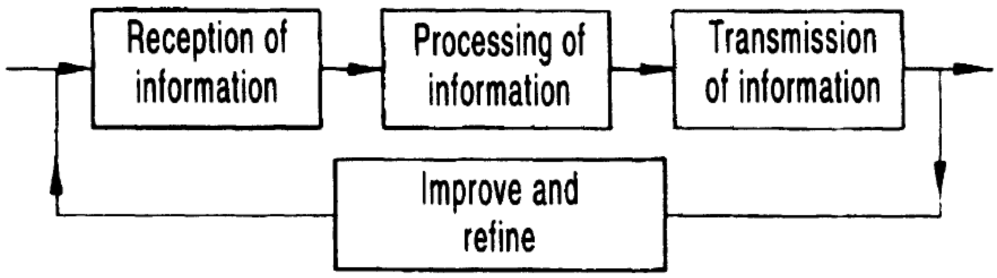

Figure 3. The conversion of information with iteration (Pahl and Beitz, 1996).

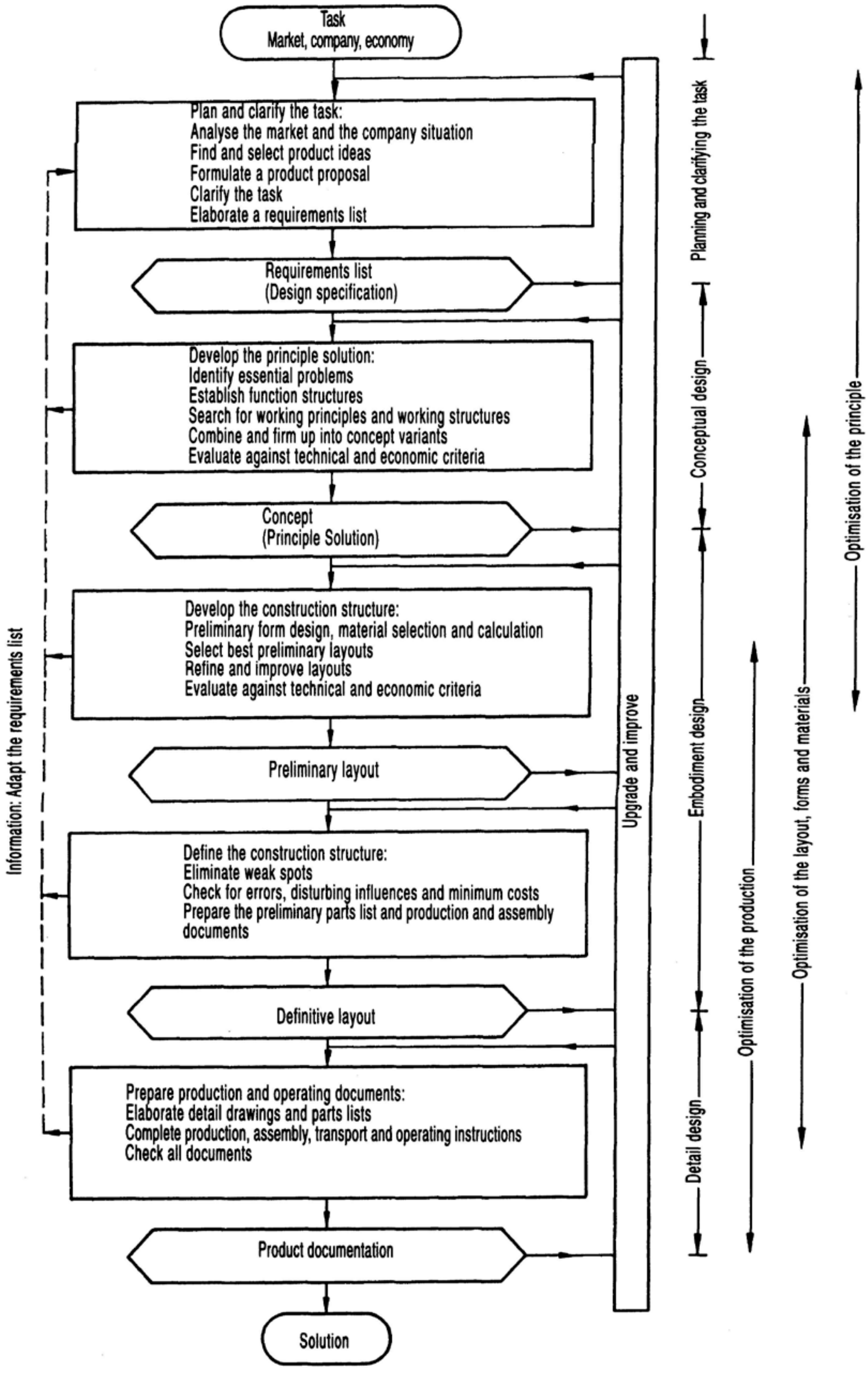

Figure 4. Four phase design process model (Pahl and Beitz, 1996).

2.2 Engineering changes in the product development process

The topic of engineering change started to gain popularity soon after the millennium due to the rise of concepts such as concurrent engineering, product platform design and simultaneous design.

Product development projects often need to make an enhancement to an existing product. The documents used to initiate such a change process are well known under the name of Engineering Change Requests (ECRs). Companies regard engineering changes as sources of problems in the product development process, both during design and manufacturing (Acar et al., 1998). The motive for making engineering changes can be to improve, enhance or adapt the product to opportunities or issues identified (Pikosz & Malmqvist, 1998). ECRs are used to specify desired product changes and keep track of the evolution of a requested change from initiation, search for a solution, verification and decision acceptance. ECRs thus contain both product and process related information. When existing products are adapted to new designs, Cross and Roy (1989) elaborated on the cost and risk in the engineering design of products.

A high-level overview of an engineering change process can be seen in Leech and Turner (1985), who compare this change process to a project that should only be taken on if the value is greater than the cost. Engineering change processes are similar at a high level, but slight variations can be seen with regard to product characteristics. The change process for safety critical products focuses more on quality than low cost (Pikosz and Malmqvist. 1988).

The management of ECRs is part of the engineering change process, corresponding to the first four stages the generic engineering change management (ECM) process of Jarratt et al. (2011) (Figure 5). The final two stages are known as the engineering change order process. The process suggests a six-stage engineering change process that begins with the ECR, identification of solutions, risk assessment, selection, approval and implementation of solution, followed by a review of the change. Hamras et al. (2013) review methods for ECM and identify 25 key requirements of such methods, including various components of process model building and use. Previously, Maull et al. (1992) proposed a five-step process and Dale (1982) suggested two main process phases.

Detailed steps of the Jarratt et al. (2011) process are:

1. An Engineering change is requested on paper or electronic form. The requester of the change outlines the reason, priority level, type of change and component or system involved.
2. Potential solutions to the change are listed to reduce investigation time or state a known solution. Only one solution is chosen with which to move forward.
3. The impact of implementing a new solution is assessed keeping in mind such factors as design, production, suppliers and budget. Later in the change process, the selected solution is implemented.
4. For approval of a solution takes place by a change committee before final implementation in which a cost-benefit analysis is performed and key stakeholders involved.
5. Implementation of the change takes place immediately or is phased in later, depending on the criticality of the change if it is a safety issue or if it can be implemented somewhere in the product life cycle.
6. Finally, an evaluation of the change takes place later to see if the intended effects have been achieved. Lessons learned are documented for future action.

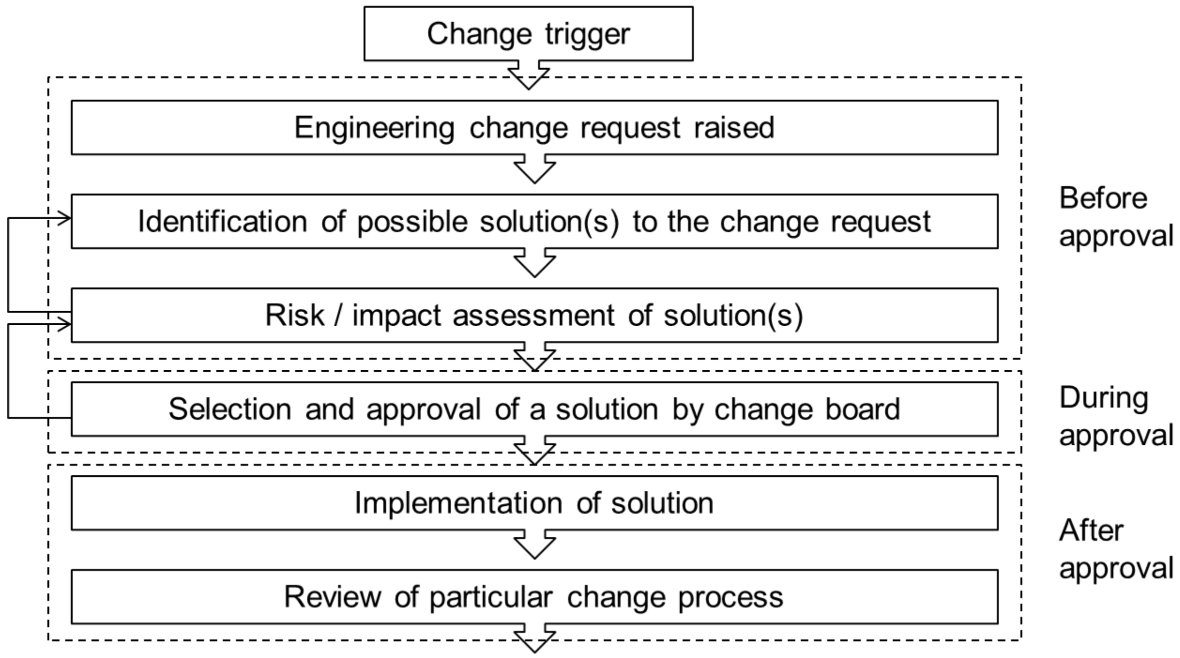

Figure 5. A generic engineering change process (Jarratt et al., 2011).

2.2.1 Engineering change data

ECR data in PD projects is often logged and stored in databases where structured data (i.e., numerical data and time stamps) is mixed with unstructured data (i.e., free text descriptions). Large development projects can contain up to tens of thousands of ECRs (Arnarsson et al., 2017).

ECR data needs to be collected carefully as key factors must be specified for the data gathering process so that those supplying, validating and analyzing the data gain a consistent view (Basili and Weiss, 1984). The ECR captures a description of the problem why a change is needed, product/part description, the name and department of the originator and date. Basili and Weiss (1984) proposed six criteria for data collection in order to identify troublesome issues and efforts to make such changes:

1. Data must contain information permitting identification of the types of errors and changes made.
2. Data must include the cost of making changes.
3. Data to be collected must be defined as a result of clear specification of the goals of the study.
4. Data should include studies of projects from production environments.
5. Data analysis should be historical; data must be collected and validated concurrently with development.
6. Data classification schemes to be used must be carefully specified for the sake of repeatability in the same and different study environments.

Change severity level in relation to customer impact is stored in the data. Jarratt et al. (2011) listed four groups of change properties that contribute to understanding how urgent a change can be:

- *Error correction:* mistakes discovered during the development life cycle, ranging from minor drawing errors to issues that affect operation of product.
- *Change of function:* required when design does meet its functional requirements. Causes can be an incorrect initial assessment or expansion of the operating environment during the design process.
- *Product quality problems:* issues regarding rework and scrap can sometimes be tracked back to poor design, incorrect assembly or manufacturing instructions.

- *Safety:* issue with respect to non-commercial boundaries (Inness, 1994). Changes must occur if a product does not meet expected safety requirements which may lead to death, injuries and property or commercial damage. Hazardous unintended product usage must also be limited.

Examples of common data stored in ECRs include the change motive, root causes, solutions, parts affected, responsible individual and department, part name and number, report status, severity points, part version, product class, date issued, date of incident, planned closure date, project number and test information. ECRs also contain transition stages including time stamps (data and time) and have the capacity of taking on more than 30 unique stages. Each ECR takes up a different stage in the resolution process starting from "ECR created" and ending with "ECR solved". The ECR stages can be categorized into eight groups (Arnarsson et a., 2018). ECR data includes all historical stage transitions that ECRs have assumed under the resolution process and whether or not an ECR has changed owner (Arnarsson et al., 2018).

2.2.2 Problems in ECR process

Recent studies have identified poor management of requirements (Fernandes et al., 2015) and difficulties in predicting the impact of design changes resulting in the late discovery of problems (Eger et al., 2007, Giffin et al., 2009) as causes for late changes. Closely related to that reason, Thomke & Reinertsen (2012) maintain that companies tend to need more time to adjust to constantly evolving market needs, which can lead to overdue detection of product weaknesses. Thomke & Reinertsen further claim that many companies try to over-utilize their product development resources. When product development employees are nearly fully utilized, speed, efficiency, and output quality decreases (Thomke & Reinertsen, 2012). When resources are highly utilized, queues in projects tend to appear. Queuing can result in the unavailability of resources, longer duration of projects, delayed feedback, and unproductive developers. However, there are many more potential causes, including the lack or poor use of simulation tools (Söderberg et al., 2013), the use of too few physical prototypes, too few milestones/lack of continuous follow-up, and reporting systems that are too cumbersome, resulting in the avoidance of reporting errors.

2.3 Process models

The ECR process is a type of design process. Wynn & Clarkson (2017) surveyed available design & simulation models to illustrate the rich variety of models. Wynn & Clarkson argue that detailed, task-based models of design processes can support design, management and improvement of “meso-level” processes including the ECR process. Due to the complexities of processes, no single model can fit all. However, design structure matrices (DSM) [Steward (1981), Eppinger & Browning (2012), and Browning (2016)] have successfully been used to construct task-based models of design processes, including stochastic factors. Design structure matrices offer support for both the qualitative and quantitative analysis of processes, e.g., visualization of processes, as well as the computation of process lead times.

On the other hand, Markov chain models (Norris (1998) or Gilks et al. (1995)) have many applications to real life situations where one wants to investigate and understand processes evolving between different discrete stages. Markov chain models have previously been utilized for analyzing product development processes (Figure 3) in, for example, Ahmadi et al. (2001), where the authors employ Markov chains for developing procedures to minimize iterations during the development process which adversely affect development time and costs. Cho and Eppinger (2001) use Markov chains to simulate a product development process with the aim of providing better project planning and control and Dong (2002) tries to employ ideas from Markov chain models to understand organizational interactions during product development processes.

However, earlier work on DSM and Markov chain have typically been applied to situation- or system-specific design processes for example a brake design process (Smith & Eppinger, 1997). Smith & Eppinger (1997) further note that generating reliable data for a DSM is challenging and will require additional effort for each new system-specific design process modeled.

2.4 Data mining and Design Analytics

2.4.1 Data mining

Focusing on the computational support for such an analysis, some researchers are developing big data mining methods to identify structures or patterns in engineering information. Researchers have developed methods to analyze e-mail databases and social media tools for making inferences about project status and for connecting relevant specialists to queries and issues (Hicks et al., 2013). Earlier studies have also shown that change requests can be analyzed using network graphs (Giffin et al., 2009). Network graphs can assist in visualizing how change requests are related to one another, emerging from a single parent or whether they are disconnected. Change analysis during ongoing product development (Eger et al., 2007) use a node-link diagram to allow the designer to monitor the progress of the project. The tool makes a change propagation tree to provide an exploded view of the design links, aiding in identifying change paths. Tree diagrams and scatter graphs have also been used to analyze data (Giffin et al., 2009). On a more general level, emerging technologies, in particular those for searching and browsing, focus on visualization and other structured presentations of information as key to efficient data analysis. Frameworks such as d3js (D3JS.org, 2013) provide building blocks that support tailor-made solutions for visual representations of text data, word clouds, patent information, as well as data driven dynamic manipulation of documents.

Figure 6 illustrates the three components needed for data mining & data analytics according to Fayyad et al. (1996): i) data, ii) domain knowledge and iii) mathematical tools, such as algorithms, optimization, and statistical models.

Fayyad et al. (1996) provided an overview of the field of data mining and knowledge discovery in databases, elaborating on how the two topics are related to each other and to fields such as statistics, machine learning and databases. They demonstrate an overall process (Figure 7) for finding data patterns with process iterations to determine which patterns can be considered new knowledge.

Arnarsson *et al*. (2016) showed how data mining and visualization tools could be applied to exploring a database consisting of ECRs from a complex truck development project. The study demonstrated a process for compiling and cleaning the data along with methods for numerical and text data analysis, for data visualization and exploration, and for pattern identification and analysis.

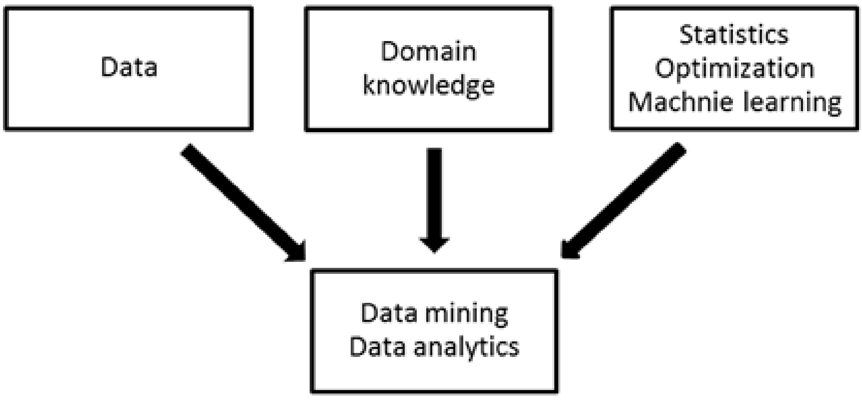

Figure 6. Schematic illustration of data mining and data analytics components (Fayyad et al. 1996).

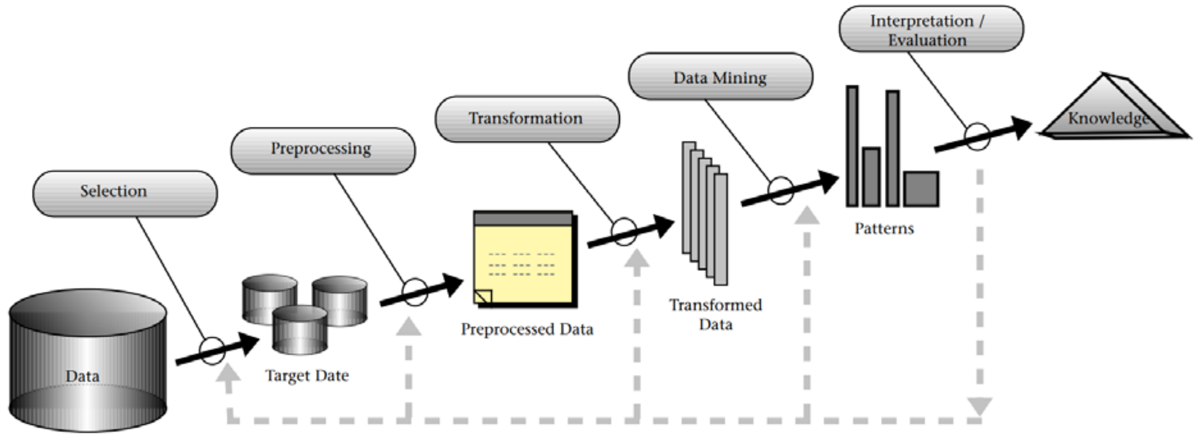

Figure 7. An overview of the process steps that compose knowledge discovery in databases (Fayyad et al., 1996).

2.4.2 Design Analytics

In the seminal book "Competing on Analytics: The New Science of Winning", Davenport and Jeanne (2007) describe data analytics as using statistical and quantitative analysis of data, combined with explanatory and predictive modeling. The models and analyses provide a base for fact-based management and decision-making. For a survey on data mining and knowledge discovery on a general level, see Han et al. (2011), and for a more technically oriented survey describing techniques and machine-learning tools for data mining, see Witten et al. (2016).

Previous information need-focused studies on analytics include Bichsel (2012), who interviewed four focus groups on how they related to analytics. Bichel's interviews covered data analyses, strategic decisions, decision-making, and the culture and politics surrounding analytics. Bichel highlighted the balance between benefits and challenges that people face when working with analytics. Bichel argued that analytics should start with a strategic question and a plan to address that question using data. Analytics should be viewed as an investment and not as an expense and does not require perfect data, but should be initiated when there is a corporate commitment and readiness. LaValle et al. (2011) conducted interviews with over 3,000 business managers and analysts from different industries in order to understand the challenges they are facing and how analytics can be used to aid their decision-making. They concluded that the use of analytical techniques is growing, but that people relate to them in different ways. Similar to Bichsel, LaValle et al. argued that when starting an analytical implementation project, there needs to be a clear question backed up with an organizational readiness and commitment to guide the journey, rather than perfect data.

In the engineering design domain, early studies related to data analytics include Kuffner and Ullman (1991) who chartered out the needs of design engineers for more design information than only standard design documents when developing complex products. Reich (1997) proposed a seven-step process for developing machine-learning tools supporting civil engineering tasks. Menon *et al*. (2005) developed data mining tools for analyzing textual databases to enable faster product development processes.

Recently, the term “Design analytics” has been proposed by Van Horn et al. (2012) to identify the area of research that focuses on processes and tools to enhance the transformation of design-related data into formats that suit design decision-making. Examples include Tucker & Kim (2011) who applied analytics to consumer trend data (product reviews) in order to inform product designers. Bae and Kim (2011) presented a study on how to optimize the development process of a digital camera by using data mining techniques on customer information. Also considering customer data, Lewis & Horn (2013) looked at customer behavior profiles and reflected upon customer needs in the late stages of the development process.

Ma et al. (2014) proposed a new demand modeling technique in order to help design engineers extract knowledge from large-scale data. The model, Demand Trend Mining (DTM), is an analysis tool to *“capture the trend of demand as a function of design attributes”*. DTM can realize Predictive-Life-Cycle-Design for manufacturing, re-manufacturing and recycling stages (Figure 7).

Ma and Kim (2014) proposed an application of a Continuous Preference Trend Mining (CPTM) to address challenges in product and design analytics. Similar to Arnarsson et al. (2018), CPTM used time stamped data and a predictive model. The CPTM methodology can be seen in Figure 8.

Zhang et al. (2016) performed a literature overview of big-data analytics for product lifecycle, with respect to product lifecycle management and a cleaner production. They focused on large amounts of real-time and multi-source lifecycle data that is being collected now in relation to the manufacturing and maintenance process of the product lifecycle.

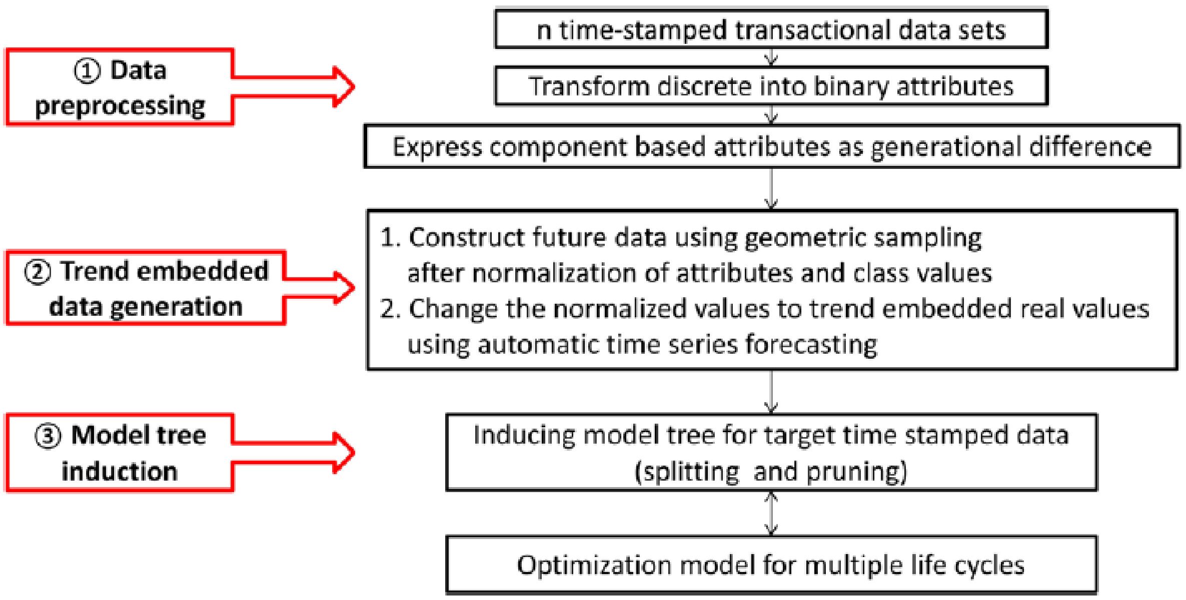

Figure 7. Overall flow of methodology for CPTM (Ma and Kim, 2014).

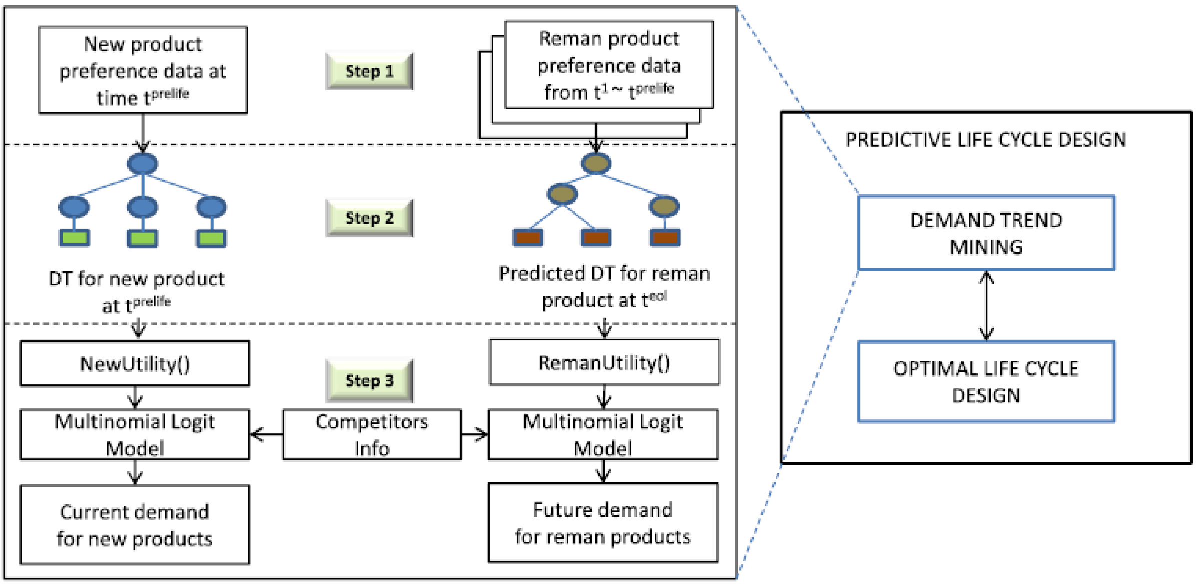

Figure 8. Framework of Predictive Life Cycle Design (Ma et al., 2014).

2.5 Research gaps

Late Changes. Previous work has only considered "single" products/systems. Variant rich, platform-based products such as trucks introduce an additional level of complexity to the task of understanding causes and effects of changes and errors. There is thus a need to look at this situation in a more detailed and comprehensive way, consider a larger set of causes and mitigations, as well as the more complex process of platform-based development.

Design Analytics. There exists some previous research within design analytics for product development based on project and customer data, but earlier research has not specifically considered ECR data. The literature cited above further proposes that analytics should start with a process to determine the relevant strategic questions before any analysis can take place. This process is not trivial as the data is so abundant and not necessarily produced with the intent of ultimately answering a specific question. The research gap is to identify strategic questions (hypotheses, information needs) in the minds of product developers towards ECR data in order to apply design analytics.

DSM. Statistical ECR DSM analyses have not been performed so far despite the availability of the data. Engineers would like to identify how the resolution process of an ECR looks like from a big picture perspective to improve the process. There exists no previous research that specifically analyses the data from ECR stages in product development. The literature cited above has identified questions that can aid the analysis of ECRs. This paper addresses this research gap and aims to apply Markov chain modeling and analysis to ECR databases in product development.

3 RESEARCH APPROACH

The purpose of this chapter is to motivate and describe the applied research approach methodology.

3.1 Design research methodology

Research methodologies are chosen to account for research gaps and questions at hand. Blessing and Chakrabarti (2009) point out that one of the main issues when conducting design research is the diversity of design activities. The chosen research methodology should enable data collection to discuss and answer the research questions. A risk when conducting design research is that topics can lead to multiple pathways, ending in unconnected streams of research (Eckert et al., 2013). The methodology for this licentiate research is related to Blessing's and Chakrabarti's (2009) Design Research Methodology (DRM), a methodology used for design research to ensure scientific validity and overcome lack of scientific rigor. DRM is described as *"an approach and a set of supporting methods and guidelines to be used as a framework for doing Design Research"*. Related research methodologies are qualitative study theory (Maxwell, 2012) and case study theory (Yin, 2017).

DRM strives to fulfill two purposes: to understand the study objective and to propose useful tools and methods to be applied. Further the DRM consists of four stages (Figure 9):

1. *Research Clarification*: Identifies and clarifies research problems and goals that will determine successful research. The main source of information at this stage is a literature study together with scenarios of desired outcomes.
2. *Descriptive Study I:* Empirical studies are used to create increased understanding of research problems and goals. The main outcome is to formulate influencing factors, models and theories under study.
3. *Prescriptive Study*: Identifies research gaps to get from the current to the desired situation. The focus of the study is to enhance *Descriptive Study I* with the use of supportive guidelines designed to evaluate of previous assumptions and concepts.
4. *Descriptive Study II:* The impact of the proposed study is evaluated to see if the supportive guidelines improve the current situation through measurable criteria.

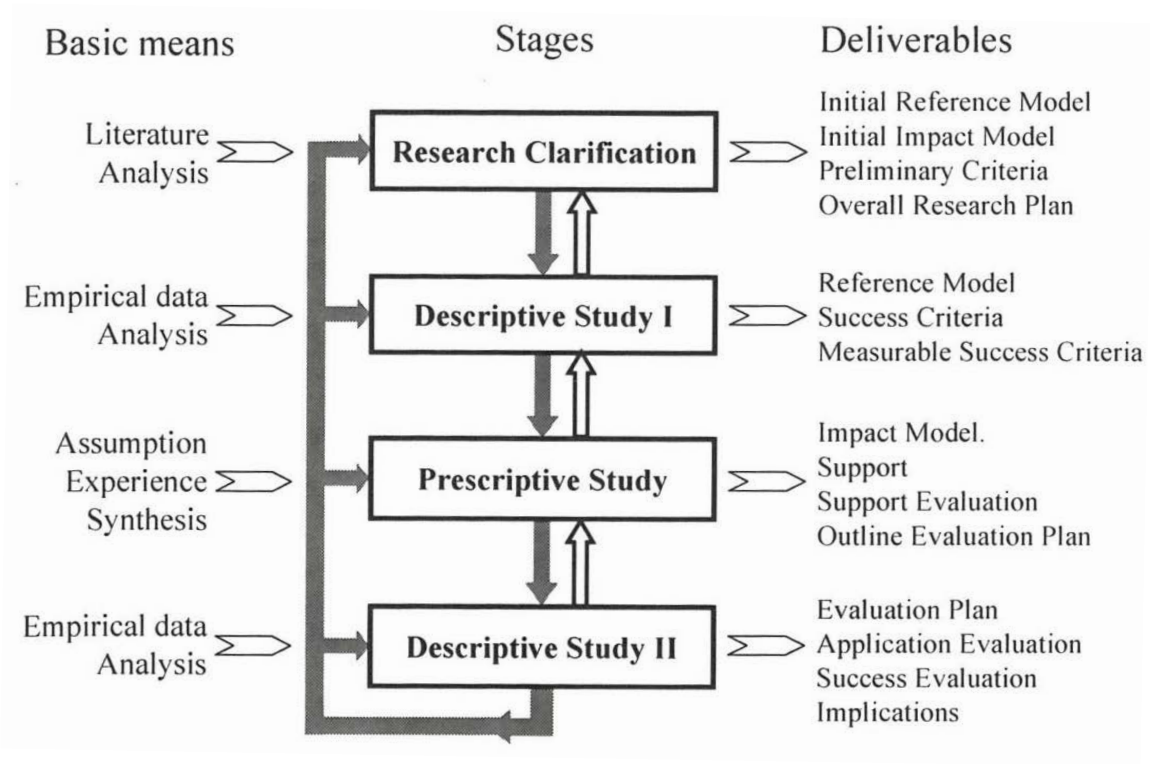

Figure 8. Design Research Methodology framework (Blessing and Chakrabarti).

All three appended papers are connected together as components of data mining and data analytics. Paper A focused the first two stages (research clarification and descriptive study I) of the DRM with literature analysis and empirical analysis. Paper B (prescriptive study) connected assumption experience synthesis. Paper C (descriptive study II) continued with a more focused empirical data analysis base on the outcomes of previous steps. Methods use for each research question can be seen in Table 1.

	Research Question	Method
1	Is it possible to use historical ECRs by applying data mining to gain new insights into data for product developers?	Cast study demonstrator
2	What information needs do product developers have regarding ECR data and what methods can support these needs?	Interview study
3	What are the efforts, benefits and limitations of using the Markov chain Design Structure Matrix for ECR process analysis?	Cast study demonstrator

Table 1. Research methods connected with research questions.

3.1.1 Paper A

In Paper A, a case-based method was applied suitable for investigation of complex configurations of events and structures (Brady and Collier, 2010). An in-depth study was conducted to identify specific behavior in the PD project selected. A demonstrator was developed to analyze data from a large recently concluded truck development project. We selected a single large project to analyze what data was available and data sources included a prototype build and test report database, a product documentation database, records of steering committee meetings and other individual documents, such as time plans and department descriptions. The data set consists of around 4,000 ECRs. The current paper reports on the encircled elements of the methodology (Figure 10).

The objective was to analyze data in order to understand the root causes of changes made during a PD project. *Numerical* data (i.e., time stamps and number of severity points) needs to be coupled with both *categorical* data (i.e., which component caused the issue) and *unstructured* data (i.e., free text inserted by engineers). The demonstrator was programmed and tested with text mining tool based on Python and ECR data.

In this thesis, these data science capabilities are placed into a methodology that supports the wider process of extracting data from a set of databases used to document a product development project. Data analyses are performed in order to find the root causes of failures, design solutions as well as actions that eliminate the root causes of those failures and then implementing these solutions/actions in the next product (Figure 10). More specifically, the process consisted of the following steps:

1. Identify relevant data sources.
2. Export data from data sources and compile this data in a suitable form, such as Excel file.
3. Evaluate data quality and clean up if necessary.
4. Formulate hypotheses on how data is connected and explored.
5. Create an Initial set of graphs to obtain an overview of the data, for example:
    - Pareto chart
    - “Bubble chart” (magnitude of words)
    - Change request diagram
    - Issues categorized into severity points

   The Python library Matplotlib (Hunter 2007) is used to create visualizations.
6. Perform text mining in Python using tools such as entity extraction (e.g., Maynard et al., 2001) and information retrieval methods (e.g., Nadeau and Sekine, 2008) to analyze and find patterns. Perform computational analysis on frequent parts and words using the programming language Python (Van Rossum and Drake 1995).

Implementation of the tool was performed in Python and the outcome of the demonstrator was evaluated with ten engineers ranging in roles within a typically PD project in order to understand what type of analyses delivered values to them and how to process with future work.

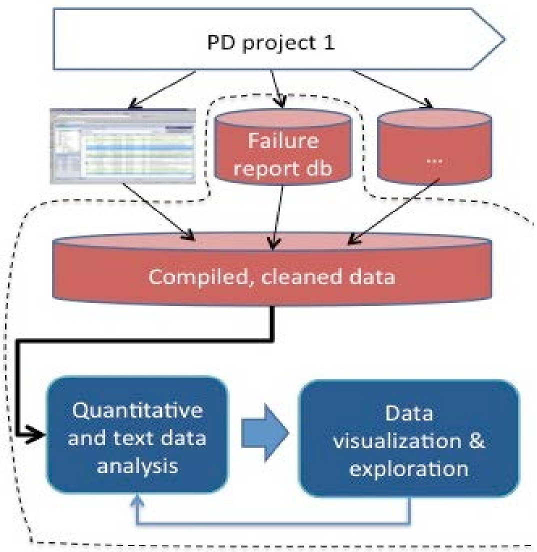

Figure 9. Flowchart of the methodology. Parts within dashed lines tested in case study (Paper A).

3.1.2 Paper B

The empirical part of Study B was based on semi-structured interviews with experienced product developers. The aim was to identify the needs of product developers towards ECR information. Interviews were selected to gain an in-depth understanding of their needs and evaluate proposals for data analyses. A heterogeneously sampled group was selected on purpose for the interviews (Figure 11). Twenty interviews were conducted with individuals of diverse experiences throughout the PD process ranging from engineering to testing and manufacturing. The interview guide comprised 16 questions split into four categories, demographics, behavior, values/improvements and wrap-up. The demographics section included background questions pertaining to their roles within the company and previous projects. Behavioral questions queried how and how often an individual came into contact with ECRs, asking them to describe the process for handling ECR reports, including their own involvement. Values/improvement questions were directed towards ECR data and processes, technical difficulties and recurring errors. Some examples of value/improvement questions are: “Is there something that can be improved in the product development process from which data can be used for learning purposes?” and “Is there a lack of analytics on ECRs from historical or on-going projects to support new decision-making?” The wrap-up section was used for open conversation during which interviewees could speak freely and add their own statements.

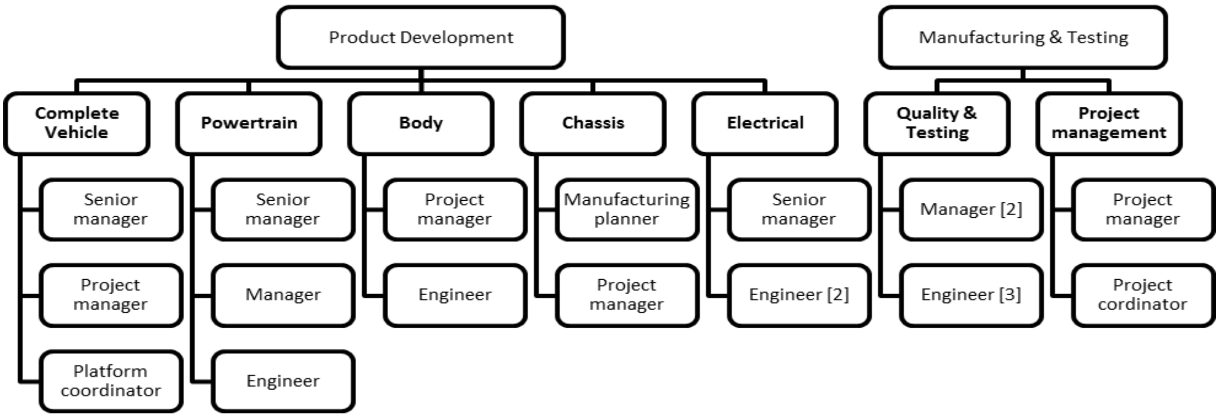

Figure 10. Graph describing the interviewee sample.

Interviews were performed in person or over Skype. The interview guide was handed out beforehand to help interviewees prepare. The interviews lasted approximately 60 minutes and were recorded and transcribed.

The interview data analyses followed Bryman's & Bell's (2011) recommendations and started with codifying and categorizing the topics. Each topic was phrased as a functional information need, such as "Identifying repeated ECRs". The topics were arranged into sections. Finally, two main categories of information needs emerged, needs related to data mining and analytics support and needs related to process and data quality. Results were validated with engineers and stakeholders in order to create a strategy for data analysis.

3.1.3 Paper C

Paper C contained empirical data analysis based on a product development database made up of ECRs from an industrial commercial vehicle development project. The aim was to model process pathways for ECRs. The Markov chain model was selected because the approach enabled a way of analyzing a large set of ECRs to provide new data driven insights. In this way, the manual analysis of a few ECRs at a time was avoided.

The data within the ECRs contained stage transitions that included timestamps (date and time) with the possibility of taking on more than 30 unique stages. Each ECR is assigned a stage under the resolution process starting with "ECR created" and ending with "ECR resolved". Each ECR contains data about the range of historical stage transitions.

The Markov chain is a stochastic process during which the transition probabilities between available stages fulfill the Markov property, i.e. the probability of evolving from one stage to another depends on the current stage. The implication is that the process is "memory less" and disregards the history of the process. When an ECR is considered as a discrete stochastic process, i.e.

when an ECR is at a specific stage at a specific point in time, the probability of being transferred from that stage to another only depends on the previous stage.

The process of using a Markov transition matrix that describes the probability of evolving from one stage to another represents the probability of an ECR evolving from stage x to stage y. By utilizing the available information, the Markov transition matrix for all ECRs can be estimated by counting the number of times that ECRs transition between the stages and by normalizing the matrix by rows.

Thereafter, the Markov transition matrix can be utilized for understanding the normal flow of ECRs, i.e., what seems to be the most common transition patterns. Product developers can also detect issues in the process by finding stages where many ECRs are reissued or prematurely closed. Another approach is to create Markov transition matrices for subgroups of ECRs or organizations within a company. By utilizing the different transition matrices, differences and similarities between the subgroups can be detected and conclusions drawn.

Python was used to construct the Markov chain DSM and visualize the results in a matrix. The matrix was then evaluated by ten engineers examining the matrix to identify known and discover new patterns which could thereafter be validated with the help of manual data.

4 SUMMARY OF APPENDED PAPERS

The three appended papers are parts of a three-part project aimed at utilizing design analytical tools for analyzing ECR data to guide product development projects. Each paper builds on the findings from the previous paper and covers aspects necessary in order to perform data mining or data analytics.

Paper A focuses on the first “Data” part and looks into a database containing ECRs issued during a product development project to investigate how the data could be explored and visualized.

Paper B identifies the “Domain knowledge” or need of product developers for information about ECRs during development projects and what kind of analysis could support them in making data-driven decisions in future PD projects.

Paper C applies the final “Mathematical tools” component in order to perform data mining or data analytics based on the previous domain knowledge.

4.1 Paper A

The aim of the paper was to explore the process behind ECR data from data identification, data quality evaluation, cleaning, analysis and finally data visualization and exploration, see Figure 12.

Large complex system development projects, such as complete truck development projects, take several years to carry out. During a project, many design decisions often need to be altered due to the emergence of new information.

These changes are often well-documented in databases, but due to the complexity of the data, few companies analyze engineering change in a comprehensive and structured fashion. This paper argues that “big data” (specifically data mining) analysis tools can be applied to such analyses and proposes a process for carrying out the analysis and using the results for product and process improvements.

The paper further reports on experiences gained from testing the approach on a data-set consisting of 4,000 deviation and error reports created during a truck development project.

The results show visualizations of quantitative and text analyses and example of computational text analysis base on report titles, descriptions of fault/problem/cause/background and additional comments (Figure 13).

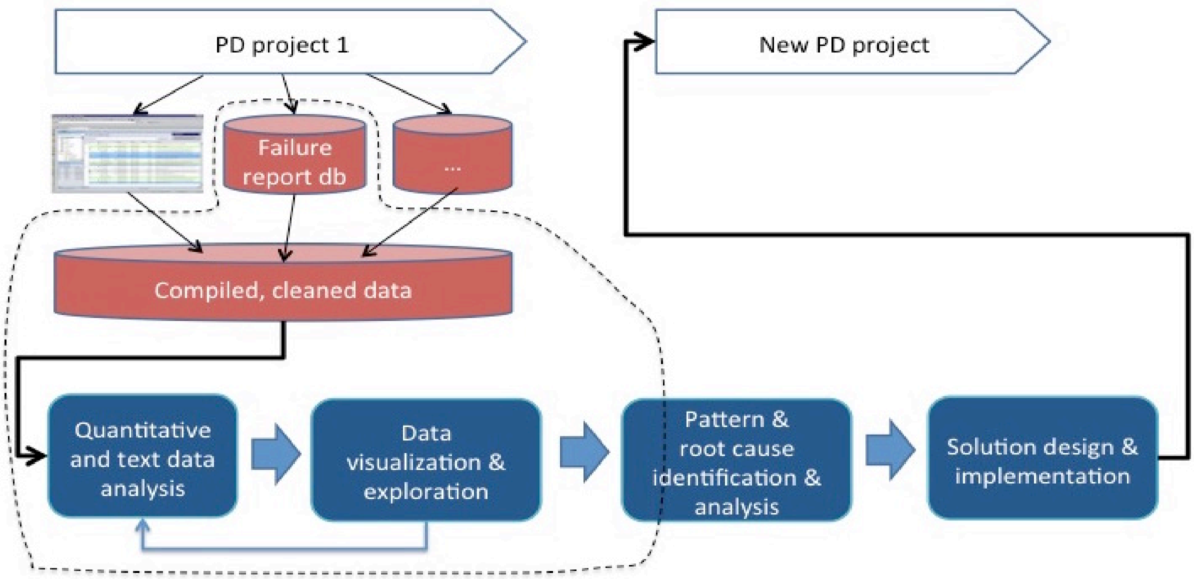

Figure 11. Flowchart of the methodology. Parts within dashed lines are tested in this case study.

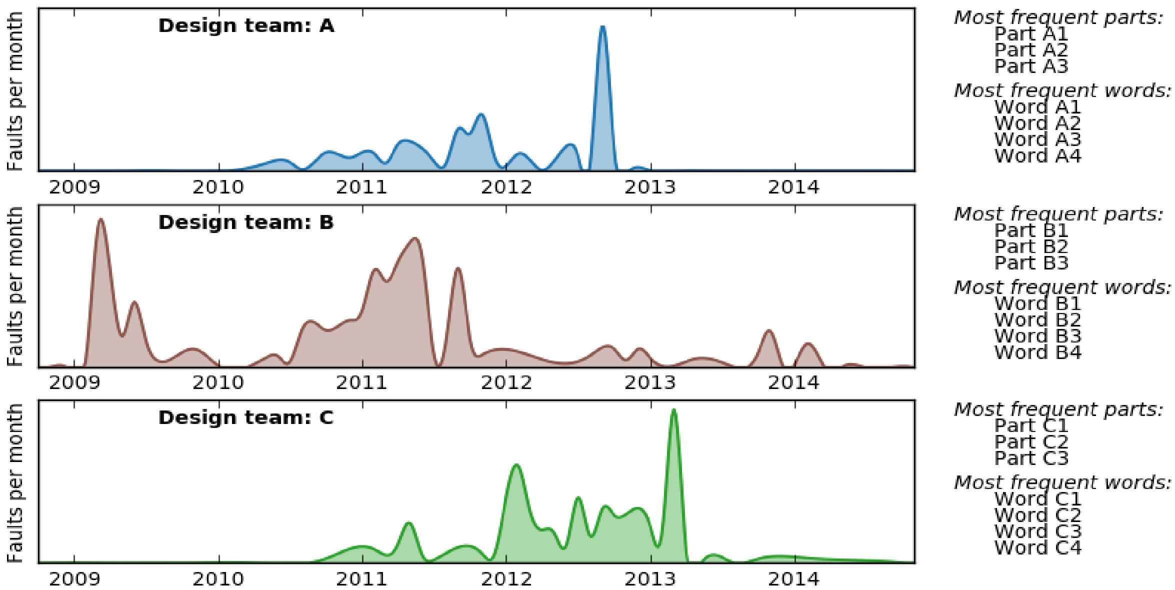

Figure 12. Example of the three design teams displaying a number of issues over time. To the right are the most frequent parts and words listed in a descending order.

This enables us to see the parts and texts were the most problematic for each design team. Through this, we discovered families of similar parts that did reoccur most often in the ECR. The main conclusions are that text data mining can provide new insights into ECRs by displaying frequent problematic words and parts and where changes appears on the project timeline.

4.2 Paper B

The purpose of Paper B is to identify information needs of product developers that can be derived with the help of design analytics. Companies are looking for improvement opportunities within their business or products through data analysis that can support engineers in their daily work.

The paper is based on interviews performed with product developers who have worked on a large complex system development project. The findings explain questions and needs faced by developers and the answers they are looking for with the help of data mining. Domain knowledge is important before starting on an analytics journey to identify beneficial and meaningful outputs that support developers in making better decisions for new product designs/re-designs.

The paper further accounts for 20 heterogeneous sample interviews, ranging from roles within product development to testing and manufacturing.

The findings were arranged into two main categories of information needs that emerged from the interviews. The categories are needs related to data mining and analytics support, and needs related to process and data quality, see Figure 14. Main conclusions are the engineers would like to identify related ECRs within projects, perform ECR process analysis and perform searches across multiple databases.

The paper further lists the top four information needs for both categories in Figure 15.

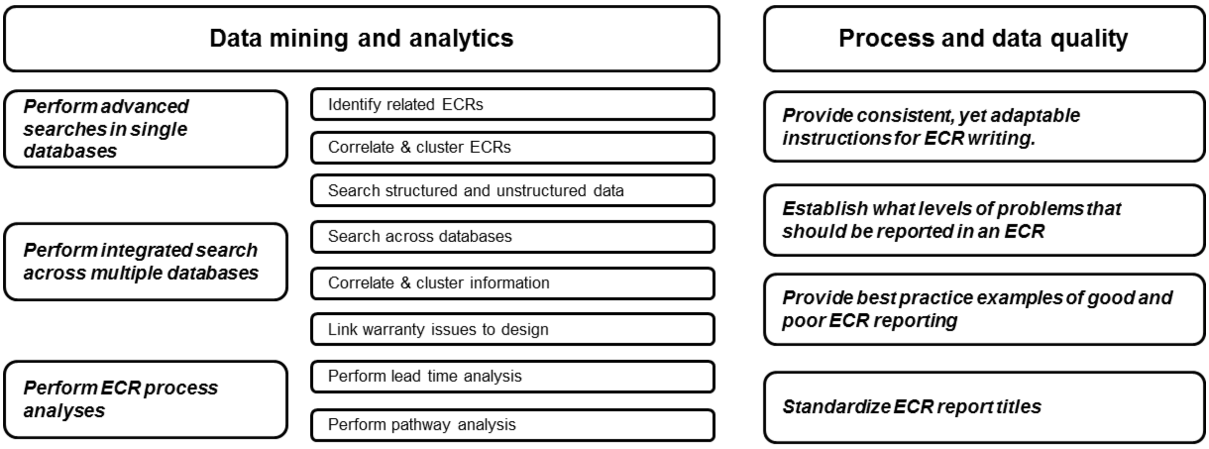

Figure 13. ECR-related information needs and issues as expressed by interviewees.

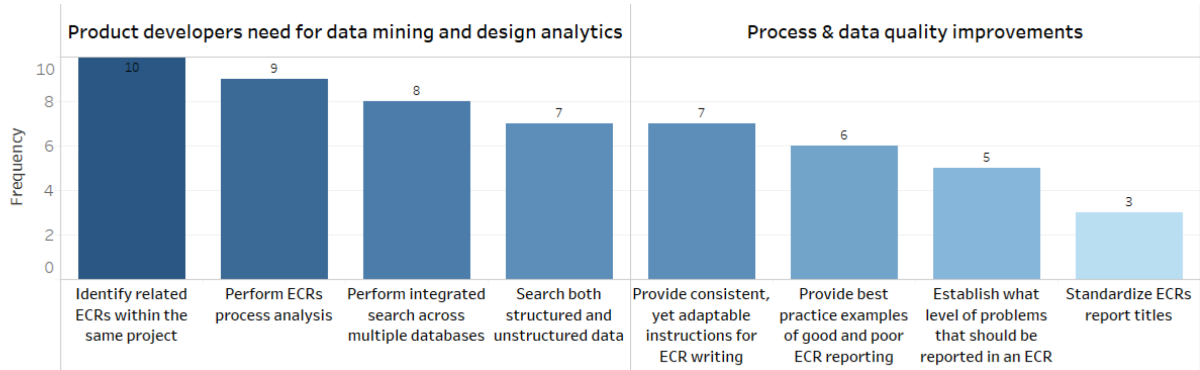

Figure 14. Frequency of information needs topics, as stated in the interviews.

4.3 Paper C

The motive for Paper C was to perform ECR process analysis based on information need of product developers. We utilize a Markov chain model on ECR data from a large PD project for process analysis and display the results in a Markov chain matrix. The matrix shows statistical probability of a transition pathway for an industrial design process and together with engineering domain knowledge, we identify patterns and improvement opportunities for the process. The Markov chain model was visualized in a matrix (Figure 16) that was evaluated together with engineers who had domain knowledge about the process and could identify areas of interest in the process.

	1	10	11	13	15	19	2	21	22	3	31	35	36	37	39	4	41	42	43	44	49	5	51	52	53	54	55	56	59	6	65	7	71	8
1: ECR created		0.98																																
10: ECR distributed			0.04		0.03	0.41				0.19	0.05	0.03																						0.08
11: ECR on hold		0.04					0.01			0.46	0.04				0.11	0.06		0.17		0.03														0.07
13: External ECR solved in this project		0.29			0.03	0.28		0.06			0.04	0.21			0.06																			0.17
15: ECR without solving responsible		0.21				0.31				0.18	0.01	0.11	0.01																					0.14
19: ECR with solving responsible		0.06			0.07			0.08		0.48	0.21	0.10			0.05																			0.02
2: ECR re-issued		0.06			0.02	0.11		0.04		0.23	0.07	0.03			0.02			0.06				0.03				0.04			0.01			0.05		0.20
21: ECR incomplete		0.07			0.05	0.11			0.06	0.33	0.11	0.17			0.07												0.03		0.01					0.08
22: ECR not approved								0.03		0.45	0.12	0.04	0.05	0.03	0.07									0.03		0.05	0.02		0.02					0.02
3: Identification of possible solution					0.04			0.13			0.30	0.14	0.06		0.17			0.02				0.03					0.03		0.02					0.10
31: Assessment of solution										0.10		0.19	0.19	0.07	0.33							0.02				0.04	0.01		0.05			0.06		0.08
35: Decision of assessed solution						0.04		0.04		0.08	0.05		0.09		0.07										0.04		0.03					0.02		0.53
36: Solution approved										0.02	0.08	0.05		0.08	0.58							0.02		0.09		0.08	0.04		0.04					0.16
37: Testing of solution										0.04	0.11	0.13	0.09		0.36							0.03				0.01	0.03		0.06			0.02		0.14
39: Solution ready										0.04		0.06				0.04						0.04		0.07		0.07	0.52							0.16
4: Verification by factory (senario 1)									0.04		0.02				0.03						0.05											0.03		0.79
41: Verification by factory (senario 2)		0.05																	0.03	0.03														0.95
42: Verification by factory (senario 3)																			0.02															0.98
43: Verification by factory (senario 4)		0.04																			0.01													0.96
44: Verification by factory (senario 4)		0.03																0.13	0.02															0.87
49: Verification completed											0.04				0.04	0.07																		0.83
5: Verification phase										0.03	0.02	0.05		0.03	0.15											0.09			0.07					0.42
51: No verification needed		0.03							0.02		0.02															0.06	0.04		0.03					0.64
52: Verification prepared									0.12																	0.04	0.02		0.04					0.76
53: Verification purchasing									0.02	0.08	0.01	0.04														0.05			0.30	0.02				0.35
54: Verification production									0.08															0.03			0.03		0.10					0.62
55: Verification testing									0.14	0.06	0.05	0.03			0.02														0.49					0.22
56: Verification aftermarket						0.02				0.04					0.04												0.10		0.23					0.55
59: Verification completed																											0.06							0.94
6: External ECR										0.11	0.06	0.06																				0.06		0.63
65: External ECR verified																											0.30			0.52		0.28		
7: ECR outside of project scope		0.05			0.03	0.04		0.05	0.04	0.08	0.12	0.02			0.04											0.06	0.09		0.09					0.42
71: ECR on hold									0.03	0.03			0.07										0.05						0.03			0.23		0.57
8: ECR solved		0.03				0.03	0.15		0.11	0.25	0.02	0.01	0.04					0.10								0.04	0.02		0.13					

Figure 15. The transition probability matrix (Markov chain matrix) for all ECRs in a single large product development project.

The matrix reveals the most common pathway for ECRs that can be read out by starting at Step 1 on the y-axis and reading the highest probability for a transition on the x-axis. If we repeat the previous process but start from the next step on the y-axis with the highest previous probability, we can then trace the most likely path ending at step 8. The pathway is therefore: 1 → 10 → 19 → 3 → 31 → 39 → 55 → 59 → 8.

Together with engineers with domain knowledge, conclusions were drawn that the most common process pathway for ECRs in not always followed, iterations can been seen in the process and that 8% of ECR reports are closed without any action.

5 DISCUSSION

This chapter answers the research questions one at a time. Research quality, validation and contribution to both science and industry are discussed.

5.1 Exploration of ECR data

RQ1. Is it possible to use historical ECRs by applying data mining to gain new insights into data for product developers?

The study performed in Paper A finds that ECR data lends itself well to data mining. Data mining tools provide a multitude of ways of filtering, visualizing and exploring the data. We can find out which parts and departments tend to produce most ECR reports, as well as those that are delayed from a project scope (Figure 17). We can find patterns, such as repeated ECR reports on the same part. We can find out which failure modes (e.g., “wear”) are most frequent. The **purpose** of these observations is to provide a good starting point to do the following:

- Identify the most frequent ECR issues (parts and words) on which product developers are working.
- Visualize where on the project timeline that those ECR issues occur and understand their timing and whether or not they are late.
- Explore ideas that contribute to unscrambling the root causes of ECRs and late ECR changes.

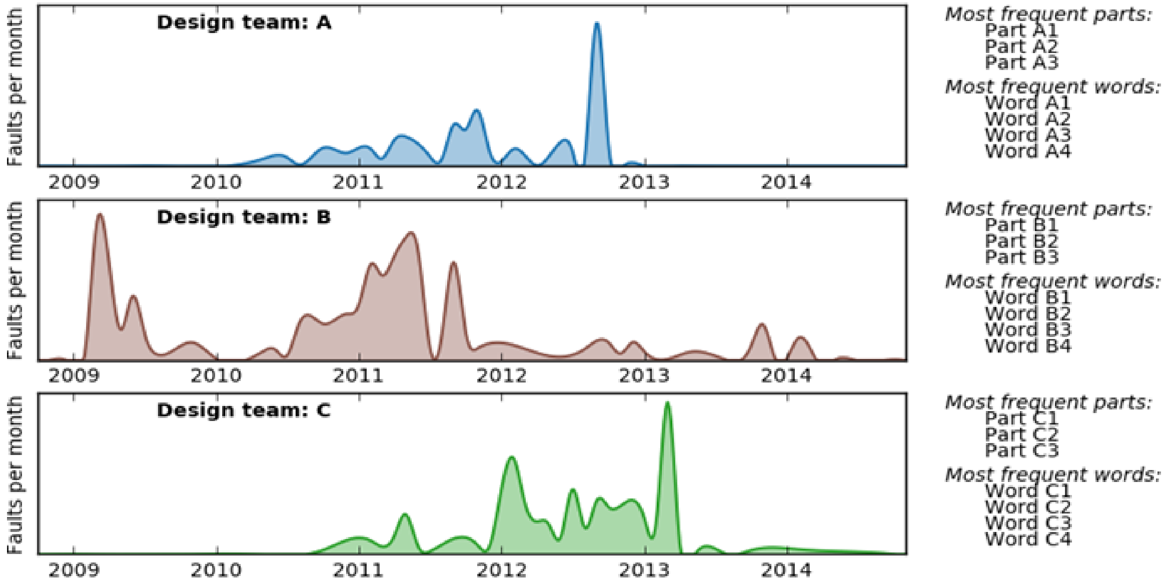

Figure 16. The three design teams displaying a number of ECRs over time. To the right are the most frequent parts and words listed in descending order.

The **benefits** for product developers, is to have new ways of data visualization and exploration that have not currently been performed with ECR data at Volvo. As the data patterns are similar to other reports (e.g., Giffin et al., 2009), we argue that the test confirms the feasibility and potential of the approach and that it might be applicable to the analysis of complex system development projects in other industries. Such visualizations can assist engineers during projects to get an overview of products with the most design changes. Another benefit is to review historical projects to analyze is product changes in projects are front loaded or not. The **limitations** are that further testing is desirable using multiple data sources (e.g., PDM systems, software change request databases and manufacturing error databases). It would also be interesting to compare ECRs across multiple product development projects to find out whether patterns repeat themselves. Figure 18 [Paper A] shows a visualization of three design teams and the most frequent parts and words in ECRs with which they have been dealing during a project. Such visualization also shows where on the timeline they took place so estimates can be made by product developers whether ECRs are early or late. These research results fulfill our research purpose and pave the way for new data mining and exploration of ECR data.

5.2 The ECR information needs of product developers

RQ2. What are the ECR information needs of product developers and what methods are available to support these needs?

The **purpose** of Paper B was to investigate the need of product developers for ECR information during development projects and how design analytical techniques can meet these needs. When preparing for data mining, it is crucial to understand how end user or domain expert will use the extractable information. The main **benefits** for product developers is to look for beneficial patterns and meaningful outputs in product development data to support developers in making better decisions for new designs/re-designs and ultimately produce superior and robust products.

We discovered in Paper B that the most frequent request from product developers working with complex PD projects generating ECRs is their desire to make a connection between ECRs and other documents and models related to product development.

Product developers are looking for easier ways in which to find relevant historical information as part of a pre-study before producing new designs. They would like to identify related ECRs within the same project and to be able to compare ECRs from previous projects. This information exists in various databases for which different departments are responsible for but comprehensive searchability across databases is missing. This searchability should include structured and unstructured data as there is hidden information in the systems about previous similar designs and service experiences that current searches do not cover since they are limited to structured data. This searchability would break down knowledge barriers and enable product developers to conduct their own mining of the data. Ultimately, they want to answer what to test for, how much testing would be needed, what are the potential problems and risks that need to be considered before dealing with similar or new designs. They would also like to link current warranty issues back to the design. They perform this work manually today but are calling for a function that can identify this automatically for them.

Product developers want to perform a detailed ECR process analysis in which ECR lead times and paths are broken down and analyzed on design group levels, groups of parts or on projects with short and long lead times. Comparisons can be made on success factors and outliers and by asking why some departments move faster than others through specific project sections. A breakdown of the sections could help identify communication issues between departments and identify department cultural differences regarding decision lead-times.

The product developers interviewed also pointed out a number of process and data quality issues in the current IT environment.

There are many potential design analytical methods that can promote a more systematic and easily accessible use of the data stored in ECR databases, thus improving decision-making in new product development projects. One method with which to identify related ECRs is to use *information clustering* machine-learning algorithms. Clustering methods could also help address the issue of related information residing in multiple databases.

Text search and classification tools could help access the rich knowledge stored in the text body of reports. Using *pattern identification tools* in a way similar to plagiarism detection tools would also enable comparison of completed ECR reports.

Design analytics could utilize available ECR data, such as responsible departments, timestamps and statuses in order to understand variation in lead times and pathways when carrying out ECRs. Analyzes could be performed with a design structure matrix (DSM) using data regarding departments and ECR process steps. DSMs have been used previously to illustrate complexities in processes where no single model can fit all [Steward (1981), Eppinger & Browning (2012) and Browning (2016)] and have successfully been used to construct task-based models of design processes, including stochastic factors. DSMs offer support for both qualitative and quantitative analyses of processes, such as the visualization of processes, as well as computation of process lead times.

Path analysis could also be carried out with data mining the benefits of which are to identify slow and fast periods for ECRs and compare these periods between departments to find best practices. This could also identify waste in the system, such as how many ECRs are written and directly closed there after without any action.

As noted above, we do see **limitations** in relation to the number of processes and data quality issues in the current IT environment. This situation may cast doubt on the possibilities for making an accurate assessment of process performance, for example if one department were to avoid reporting quality issues, thus giving the impression that quality issues do not exist. Although it has been argued that effective use of design analytics does not require *perfect* data quality (LaValle *et al*., 2011; Bischel, 2012), a prerequisite of efficient design analytics is *sufficiently* good data quality.

5.3 Application of Markov chain on ECR data

RQ3. How can Markov chains apply to ECR data and what patterns and improvements can be gained from a Markov chain matrix?

The **purpose** of using Markov chain and Markov DSM was to understand the flow of ECRs through different ECR steps. By creating the transition matrix, one can use it directly for drawing conclusions by noting common patterns or by finding anomalies. An even more powerful utilization of the Markov chain is to compare transition matrices for different departments in the organization and realize differences in the work flow of the various departments.

The **benefits** of applying DSM are to help identifying the most common transition patterns for the steps taken by ECR. That insight can be beneficial and can open up a discussion whether ECRs are allowed to deviate from common pathways in the process or are allowed to take up any chosen step. Early on in the process before the approval of a solution, ECRs should follow a fairly similar pathway of creation, making sure that they are responsible for identifying a solution. We found out that the Markov chain DSM has the capacity of identifying these deviations. The conclusions are best drawn with the help of engineers who have domain knowledge about the different steps in the process and who can point out deviations and areas of interest to them.

Ways of potential improvements. We identified that there is room for improvements in the system of generating ECRs by observing the probability of an ECR being closed as a following step after it had been created. Initiating an ECR takes time and it is estimated that around 8% of all ECRs generated are prematurely closed which could prove significant for larger development projects that might generate more than tens of thousands of ECRs.

In the early stages of ECRs, the process of distributing and assigning individuals responsible for managing this task seems to be handled in an inconsistent way. It is not mandatory to take the step of identifying a responsible individual and the DSM confirms that the most common pathway for ECRs is not followed in the early steps for all ECRs, which can be seen as a deviation in the process and could be made mandatory.

The same goes for the step after a responsible individual has been appointed and the process of identifying a potential solution has begun. In that case, most ECRs enter the "investigation phase" by noting that an identification of a possible solution is underway, but ECRs often skip that step and instead take the next probable step of going straight into an assessment of a solution. Our research findings agree with the earlier work on Markov chain models (Norris (1998) or Gilks et al. (1995)), in which they are said to have many applications to real life situations where one wants to investigate and understand processes evolving between various discrete states.

Limiting factors: It is usually not enough to create the model but to apply domain knowledge to the process of analysis. Engineers with domain knowledge are needed to evaluate patterns and point to pathways interesting to their work. Another limitation of using the Markov chain is that the Markov property is memoryless but is often assumed to be the other way around. The path to a specific ECR status can of course affect the probability of transitioning to new states. However, an initial analysis indicated that the Markov memoryless property is in most cases a valid assumption.

5.4 Validity and transferability of results

Research validity is partly based on the close involvement with industry over the past two and half years that the main research author has had with Volvo. By spending most of the research time at Volvo the researcher has gained insights into the industry and developed a deep understanding of the research subject, something that gives credibility to the findings. Creswell (2013) states that the more connected the research is to its environment, the more accurate and valid the research findings will be.

All three papers have undergone peer review and have been accepted by recognized European or international conferences where a public presentation and defense took place and where subject experts had the opportunity to comment on research results. The results were also shared with a broad range of experts residing in academia and industry. Additionally, the preliminary findings have been presented before academic and industrial participants at the Wingquist Laboratory.

The verification of design research can be performed by acceptance of experts (Buur, 1990) when new results are presented and in industry when results mirror reality. In the context of transferability, this study was conducted in a large multinational firm that develops and manufactures commercial vehicles. The way in which ECRs are managed is typical for other large firms that develop complex systems and products, for example in the aerospace and defense industries. ECR data structures and processes are similar; we therefore argue that the findings related to needs and potential solutions can be transferred to these contexts. However, the product developers interviewed had a mechanical or electrical backgrounds. It is possible or even likely that software developers have somewhat varying information needs. It is likely that the same results will be generated by using similar research methods. The only uncertainty regarding reliability is the industrial focus at the time as interviewees might answer differently due to a shifting focus within the industry.

5.5 Scientific contribution

The general aim of the licentiate thesis is to *explore how the quickly growing field of data mining and design analytics can be applied in the context of ECRs to understand the information needs of product developers for such analysis.*

Specifically, the contributions listed below have been made:

- Presented a methodology for the process of ECR data extraction from a database and the steps needed to gain insights from the data (Paper A).
- Demonstrated text data analysis of ECR data (Paper A).
- Mapped out product developer needs for data mining and design analytics (Paper B).
- Demonstrated process modeling for ECRs using Markov chains (Paper C).

An important new scientific contribution is the “needs of product developer’s for data mining and design analytics” as this has never been identified before towards ECRs (Paper B).

Previous research has utilized Markov chain models for analyzing product development processes, e.g., Ahmadi et al. (2001), in which the authors employ Markov chains to minimize iterations during the development process that adversely affect development time and costs. Cho and Eppinger (2001) use Markov chains to simulate a product development process for the aim of providing better project planning and control and Dong (2002) tries to employ ideas from Markov chain models to understand organizational interactions during product development processes. However, despite the fact that the data is available, statistical ECR DSM analyses have not been performed so far. Engineers would like to identify how the resolution process of an ECR looks like to improve the process (Paper B). There exists no previous research that specifically analyzes ECR transition data in PD projects (Paper C). We have therefore addressed this specific research gap. The results of this research enable us to guide our research efforts to mirror the needs in reality. Paper A proved that ECR text data mining is possible and beneficial visualizations can be created that support engineering in their projects.

5.6 Industrial contribution

The industrial contribution has focused on identifying meaningful data mining and design analytical tools capable of extracting new information from ECRs.

The industrial contributions are listed below:

- Demonstrated how text data analysis can be visualized (Paper A).
- Identified the most frequent parts and words over the project lifespan (Paper A).
- Discovered areas of interest to product developers towards ECRs (Paper B):
    - Data mining and analytics
    - Process and data quality.
- Highlighted the most important focus areas for product developers (Paper B).
- Visualized the probability of transitional steps for ECR (Paper C).
- Identified process improvement areas based on the Markov chain DSM (Paper C).

6 CONCLUSION & FUTURE WORK

This final chapter concludes with a summary of the results and outlines the direction of future work.

6.1 Conclusions

This licentiate thesis has been focusing on ECR data in the product development process in order to extract new information that will lead to product and process improvements for product developers.

The First Research Question was concerned with large amounts of ECRs containing both quantitative and qualitative data that pose challenges to analyze then in a systematic manner. Data mining tools provide the means for analyzing large volumes of complex data. Such analyses include combining data sources and types, cleaning up the data, charting, analyzing large text data-sets, as well as visualizing and exploring multi-faceted and multi-level data. A potential benefit of such analyses includes finding patterns in ECR report data, such as repeated ECR reports on the same part and occurrence of similar failure modes (wear, corrosion, and misfits). This work has so far developed a basic data mining approach for analyses and testing of ECR data on a limited data-set. The approach has been tested on a data-set comprising 4,000 ECRs and seems to work well.

The Second Research Question chartered out information needs and interests of product developers in a large multinational firm focusing on ECRs. The aim was to identify questions that are difficult to answer but that may be addressed by applying data mining and design analytical tools and methods using current IT systems.

The interviewees confirmed that the amount of ECR data that is collected today is not used for proactive purposes and that there was a potential for data mining and design analytical tools to assist in this effort.

Two main categories of information needs were identified. The first category was directly related to data mining and design analytical capabilities and included requests for functionality conducting comprehensive and flexible database searches. Such databases contain structured and unstructured data for functions enabling integrated searches across multiple databases and support for analyses of lead times and pathways in ECR processes. The second category of needs was related to process improvements and data quality of ECRs. The developers were concerned about the quality and consistency of logging ECR data and wish to communicate best practices to all groups within the organization. An overall impression of the interviews is that developers wanted to acquire more easily accessible tools for analyzing previous projects and gain knowledge of the pitfalls and risks before choosing specific product designs. If they are to be used by a majority of product developers, ease of use and quick learning are essential for data mining and design analytical tools.

The Third and Final Research Question was concerned with modelling ECR data using the Markov chain DSM. Results clearly show that this can be accomplished and has been proven useful in analyzing transitions of ECRs. We displayed four cases as key results of the study. First, the model identified the percentage of ECRs which had been closed directly after creation. Second, the model confirmed that there is a tendency to skip the transition to identify an individual responsible for the solution, the most common way of working and instead transition into identifying a possible solution. Third, there is another tendency to skip the transition to identify a possible solution and transition straight into the assessment of a solution. Finally, there is a possibility of identifying frequently occurring iterations in the transition process, i.e. we see loops in the model where ECRs transition back to an earlier point.

The results show that the Markov chain model has proven beneficial in analyzing ECR process. It is possible to learn from these transitions to improve the creations of ECRs so that they are not put to rest after creation, make certain transitions mandatory if they should follow the most common paths and show where iterations in the process can be found.

6.2 Future work

As a continuation of the study, we plan to use the knowledge gained from interviews with product developers to evaluate and further develop existing and novel data mining and design analytical tools to support developers. The ECR data from previous product development projects is available and will be used to answer other identified gaps.

As a result of observations together with engineers when the current DSM was reviewed, we see three opportunities to continue the Markov chain DSM.

- *First, include the quantity of ECRs together with the probability of a transition to* enable engineers to identify the volume of ECR traffic transitioning through the matrix and the size of process deviations.
- *Second, to introduce the time element into the model and evaluate lead time between transitions,* break down the analysis of lead time into sections to identify slow and fast project periods. Comparisons can be

made on success factors and outliers by asking why some departments move faster through specific project sections than others. A breakdown into sections could help identify communication issues between departments, improve the speed by which quality issues are solved and identify department differences regarding decision periods.

- *Third, make a risk assessment of ECRs based on predicted time of solving the ECR issue from creation.* Use historical information to predict the efforts needed for the visual management of ECRs, including resource planning and lead time for solving issues.

We were recently awarded a grant from the Swedish innovation agency Vinnova for machine-learning in the automotive industry. The project is called Machine-Learning-for-Engineering-Knowledge (MALEK) and this *research proposal* is on how to create knowledge from existing data sources (ECRs and check-sheet data) and utilize this efficiently in the development process. The potential solution is to utilize machine learning algorithms and smart assistants to identify the right knowledge at the right time for the right individual, which in the context of this project is an engineer or service technician making an "uninformed" decision. Chalmers with the Wingquist Laboratory (WQ) is the main applicant, Fraunhofer Chalmers Research Center (FCC) is the driving partner of Machine Learning knowledge and Rejmes Transportfordon AB and AB Volvo are the main industrial participants and primary users of the research result. *The aim* of this research is to enable a transition from an experienced based development process to a proactive product management system in which machine learning is used, predict decision points in development process and provide underlying situation adapted knowledge customized for those engineering decision points. The developed tools will be introduced in industrial projects for testing and validation. We see further opportunities by expanding into manufacturing and warranty data sources.

REFERENCES

Acar, B.S., Benedetto-Neto, H. and Wright, I.C., 1998. Design change: problem or opportunity. In *Engineering Design Conference, Brunel University, UK: Professional Engineering Publishing*.

Ahmadi, R. Roemer, T.A., Wang, R.H. (2001), "Structuring product development processes", *European Journal of Operations Research*, 130(3), pp.539-558. doi.org: 10.1016/S0377-2217(99)00412-9

Andreasen, M. and Hein, L. 1987. *Integrated product development.* Bedford, UK: IFS Publication Ltd.

Arnarsson, Í.Ö., Gustavsson, E., Malmqvist, J. and Jirstrand, M., 2018. Applying Design Analytics to Understand Engineering Change Request Information. Proceedings of the 15th International Design Conference on Excellence in Design (Design 2018), Dubrovnik, Croatia.

Arnarsson, Í.Ö., Gustavsson, E., Malmqvist, J. and Jirstrand, M., 2017. Design analytics is the answer, but what questions would product developers like to have answered?. In *DS 87-7 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 7: Design Theory and Research Methodology, Vancouver, Canada, 21-25.08. 2017*.

Arnarsson, I.Ö., Malmqvist, J., Gustavsson, E. and Jirstrand, M., 2016. Towards big-data analysis of deviation and error reports in product development projects. In *Proceedings of NordDesign 2016, Trondheim, Norway* (pp. 83-92).

Bae, J.K., and Kim, J. (2011), "Product development with data mining techniques: A case on design of digital camera", *Expert Systems with Applications*, Vol. 38 No. 8, pp. 9274-9280

Basili, V.R. and Weiss, D.M., 1984. A methodology for collecting valid software engineering data. *IEEE Transactions on software engineering*, (6), pp.728-738.

Berkhin, P. (2006). A Survey of Clustering Data mining Techniques. In *Grouping Multidimensional Data*, 25-71. Springer Berlin Heidelberg.

Bichsel, J. (2012), Analytics in higher education: benefits, barriers, progress, and recommendations, Research Report, CO:EDUCAUSE Center for Applied Research, Louisville, USA.

Blessing, L.T. and Chakrabarti, A., 2009. *DRM: A Design Reseach Methodology* (pp. 13-42). Springer London.

Browning, T.R. (2016) "Design structure matrix extensions and innovations: A survey and new opportunities", IEEE Transactions on engineering management, Vol. 63, No. 1, pp 27-52

Buur, J. (1990), *A theoretical approach to mechatronics design*.

Cho, S.-H., Eppinger, S. (2001), “Product development process modelling using advanced simulation”, *Proceedings of ASME Design Engineering Technical Conference*, Pittsburgh, Pennsylvania, USA.
Clark, K. B., & Fujimoto, T. (1991). *Product Development Performance: Strategy, Organization, and Management in the World Auto Industry*. Cambridge: Harvard Business School Press.
Creswell, J. W. (2013) *Research design: Qualitative, quantitative, and mixed methods approaches*, Sage publications.
Cross, N. and Roy, R., 1989. *Engineering design methods* (Vol. 4). New York: Wiley.
D3JS.org. (2013). *Data-Driven Documents.* Retrieved from http://d3js.org/
Dale, B.G., 1982. The management of engineering change procedure. *Engineering management international*, *1*(3), pp.201-208.
Davenport, T. H. and Jeanne, G. H. (2007), *Competing on Analytics: the New Science of Winning*, Harvard Business Press, Cambridge, MA, USA
Dong, Q. (2002), Predicting and managing system interactions at early phase of the product development process, Ph.D. Thesis, Massachusetts Institute of Technology, USA.
Eckert, C., Clarkson, P.J. and Stacey, M., 2003, August. The spiral of applied research: a methodological view on integrated design research. Design Society.
Eger, T., Eckert, C. M., & Clarkson, P. J. (2007). Engineering Change Analysis during Product Development, *Proceedings of ICED’07*, 629-630, Paris, France.
Eppinger, S.D. and Browning, T.R (2012), *Design structure matrix methods and applications*, MIT Press, Cambridge, MA, USA
Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) “From data mining to knowledge discovery in databases”, AI Magazine Vol. 17 No. 3, pp. 37-54
Fernandes, J., Henriques, E., Silva, A., & Moss, M. A. (2015). Requirements Change in Complex Technical Systems: An Empirical Study of Root Causes. *Research in Engineering Design*, *26*(1), 37-55.
Giffin, M., de Weck, O., Bounova, G., Keller, R., Eckert, C., & Clarkson, P. J. (2009). Change propagation analysis in complex technical systems. *Journal of Mechanical Design*, *131*(8), doi:10.1115/1.3149847.
Gilks, W.R., Richardson, S. and Spiegelhalter, D. (1995), *Markov chain Monte Carlo in practice*, CRC press.
Hamraz, B., Caldwell, N.H.M, Wynn, D.C., and Clarkson, P.J. (2013), “Requirements-based development of an improved engineering change management method”, Journal of engineering design, Vol. 24, No. 11, pp 765-793, doi.org/10.1080/09544828.2013.834039
Han, J., Kamber, M. and Jian, P. (2011), *Data Mining: Concepts and Techniques*, 3rd ed, Morgan Kaufman, New York, USA.

Han, J., Kamber, M. and Jian, P. (2011), *Data mining: Concepts and Techniques*, 3rd ed, Morgan Kaufman, New York, USA

Hicks, B. (2013). The Language of Collaborative Engineering Projects. *Proceedings of ICED'13*, Seoul, South Korea.

Hubka, V. and Eder, W. E. 1996. *Design science: introduction to needs, scope and organization of engineering design knowledge*, Springer.

Inness, J.G., 1994. *Achieving successful product change: A handbook*. Financial Times/Pitman.

Jarratt, T.A.W., Eckert, C.M., Caldwell, N.H. and Clarkson, P.J., 2011. Engineering change: an overview and perspective on the literature. *Research in engineering design*, *22*(2), pp.103-124.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised Machine-learning: A Review of Classification Techniques. *Informatica*, 31: 249-268

Kuffner, T.A. and Ullman, D.G. (1991), "The information requests of mechanical design engineers", *Design Studies*, Vol. 12 No. 1, pp. 42-50.

LaValle, S., Lesser, E., Shockley, R., Hopins, M.S. and Kruschwitz, N. (2011), "Big data, analytics and the path from insights to value", *MIT Sloan Management Review*, Vol. 52 No. 2.

Leech, D.J. and Turner, B.T., 1985. *Engineering design for profit*. Halsted Press.

Lewis, K., and Van Horn, D. (2013), "Design analytics in consumer product design: A simulated study", In *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V03BT03A003-V03BT03A003). American Society of Mechanical Engineers

Ma, J. and Kim, H.M., 2014. Continuous preference trend mining for optimal product design with multiple profit cycles. *Journal of Mechanical Design*, *136*(6), p.061002.

Ma, J., Kwak, M. and Kim, H.M., 2014. Demand trend mining for predictive life cycle design. *Journal of Cleaner Production*, *68*, pp.189-199.

Maull, R., Hughes, D. and Bennett, J., 1992. Special feature. The role of the bill-of-materials as a CAD/CAPM interface and the key importance of engineering change control. *Computing & Control Engineering Journal*, *3*(2), pp.63-70.

Maxwell, J.A., 2012. *Qualitative research design: An interactive approach* (Vol. 41). Sage publications.

Menon, R., Tong, L.H. and Sathiyakeerthi, S. (2005), "Analyzing textual databases using data mining to enable fast product development processes", *Reliability Engineering & System Safety*, Vol. 88, No. 2, pp. 171-180.

Norris, J.R. (1998), *Markov Chains*, Cambridge University Press, Cambridge, UK.

Otto, K. and Wood, K., 2001. Product design: techniques in reverse engineering and new product design. *Prentice-Hall*.
Pahl, G. and Beitz, W. 1996. *Engineering Design: A Systematic Approach,* London, Springer-Verlag.
Pikosz, P. and Malmqvist, J., 1998, September. A comparative study of engineering change management in three Swedish engineering companies. In Proceedings of the DETC98 ASME design engineering technical conference (pp. 78-85).
Reich, Y. (1997), "Machine learning techniques for civil engineering problems", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 12, No. 4, pp 295-310.
Roozenburg, N. F. and Eekels, J. 1995. *Product design: fundamentals and methods,* Wiley Chichester.
Smith, R.P., Eppinger, S.D. (1997) "Identifying controlling features of engineering design iteration", *Management Science*, Vol. 43, No. 3, pp. 276-293.
Söderberg, R., Rosenqvist, M., Falck, A.-C., Wahlborg, P.-J., Silow, N. (2013). *Proaktiv monteringsergonomisk och geometrisk kvalitetssäkring för hållbar produktion*, Projektbeskrivning, Wingquist Laboratory, Chalmers tekniska högskola, Göteborg.
Steward, D.V. (1981), "The design structure system: A method for managing the design of complex systems", IEEE Transactions on engineering management, Vol. 28, No. 3, pp 71-74
The Standish Group. (2014). *Standish Group 2014 CHAOS Report.* https://www.projectsmart.co.uk/white-papers/chaos-report.pdf.
Thomke, S., & Reinertsen, D. (2012). Six Myths of Product Development. *Harvard Business Review*, 90(5), 84-94.
Tucker, C. and Kim, H. (2011). "Predicting emerging product design trends by mining publicly available customer review data". In *Proceedings of ICED'11,* Vol. 6: Design Information and Knowledge, Lyngby, Denmark
Ullman, D. G. 1992. *The mechanical design process,* McGraw-Hill New York.
Ulrich, K. and Eppinger, S. 2012. *Product design and development*. McGraw-Hill Higher Education.
Van Horn, D., Olewnik, A. and Lewis, K. (2012), "Design analytics: capturing, understanding, and meeting customer needs using big data". In *ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. 863-875). American Society of Mechanical Engineers
Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
Wright, I.C., Duckworth, A.P., Jebb, A. and Dickerson, D.B., 2005, February. Research into the process of engineering change within incremental product

design. In *Design for Excellence: Engineering Design Conference 2000* (p. 449). John Wiley & Sons.
Wu, X., Zhu, X., Wu, G.Q. and Ding, W. (2014), "Data mining with big data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 1, pp. 97-107
Wynn, D.C., Clarkson, P.J. (2017), Process models in design and development, *Research in Engineering Design*, Vol. 28 No. 1, pp.1-42. doi.org/10.1007/s00163-017-0262-7
Yin, R.K., 2017. *Case study research and applications: Design and methods*. Sage publications.
Zhang, X., Hao, Y. and Thomson, V., 2015. Taking ideas from paper to practice: a case study of improving design processes through detailed modeling and systematic analysis. *IFAC-PapersOnLine*, *48*(3), pp.1043-1048.
Zheng, J. and Dagnino, A. (2014) "An initial study of predictive machine-learning analytics on large volumes of historical data for power system applications". *2014 IEEE International Conference on Big Data*, pp. 952-959.

APPENDICES