21ST INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED17
21-25 AUGUST 2017, THE UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, CANADA

DESIGN ANALYTICS IS THE ANSWER, BUT WHAT QUESTIONS WOULD PRODUCT DEVELOPERS LIKE TO HAVE ANSWERED?

Arnarsson, Ívar Örn (1); Gustavsson, Emil (2); Malmqvist, Johan (1); Jirstrand, Mats (2)
1: Chalmers University of Technology, Sweden; 2: Fraunhofer Chalmers Centre, Sweden

Abstract

There is a growing need for data expertise and data analysis. Companies are looking more towards analytics for improvement opportunities within the business and products. Data collection is growing at a fast pace and we need capabilities to be able to analyze it. The data volume that companies are sitting on makes this task even more important. The paper presents interviews performed with product developers who have worked on a large complex system development project. The findings explain questions and needs developers are facing and what answers they are looking for with data mining. By Identifying beneficial and meaningful outputs from data mining and data analytics, developers can be supported in making better decisions for a new designs/re-designs and ultimately make a superior robust product. The paper further accounts for 20 heterogeneous purposefully sample interviews, ranging in project roles from product development to manufacturing and testing.

Keywords: Design informatics, Digital / Digitised engineering value chains, Information management, Knowledge management, Product Lifecycle Management (PLM)

Contact:
Ívar Örn Arnarsson
Chalmers University of Technology
Product Development
Sweden
varo@chalmers.se

Please cite this paper as:
Surnames, Initials: *Title of paper*. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 7: Design Theory and Research Methodology, Vancouver, Canada, 21.-25.08.2017.

1 INTRODUCTION

Complex vehicle development projects involve hundreds of product developers developing tens of thousands of parts and millions of lines of codes. The projects have durations of several years. During the course of a project, many design decisions are made. Many decisions need to be changed during a project due to, e.g., the emergence of new information, simulation or test results. Change needs tend to be discovered late in the process (Giffin *et al*., 2009) and it is well known that changes made late in a development project are very costly and may cause delays (Clark and Fujimoto 1991). Causes of late changes in product development projects can result in failure to meet objectives for budget, schedule, or technical performance.

Information about product changes in product development projects are often logged and stored in databases where structured data (i.e., numerical data and timestamps) are mixed with unstructured data (i.e., text inserted by engineers). With the fast development of data collection and storage, companies are now faced with large volumes of complex data from multiple sources (Wu et al., 2014). Roughly 90 % of the world's data has been created during the last two years: 2.5 Exabyte of data created every day (Data, 2015). We can assume the same growth of data for high tech industrial companies who are storing historical projects in databases. Many companies are waking up to the fact that they are sitting on large amounts of data that could be used for supporting decisions in new projects (Wu et al., 2014; Davenport *et al*., 2010). Companies are however also faced with the challenge of getting an overview of what data is available in their databases and what data that they are producing.

With the recent development within machine learning and data mining, new techniques for retrieving insights from complex data sets have emerged, and have initiated and interest for mining the data from on-going projects as well as previous projects in order to identify performance improvements and cost reductions within product development.

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, data can be identified and collected from the databases, and data analytics can be employed in order to extract insights from the historical data that companies are sitting on (Zheng *et al*., 2014). The traditional way of analysis has been to try to find answers in data with manual exploration, for example by users who have exported data to a spread sheet 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 that can help shed a light on a variety of complex issues that would not be obtainable through a "manual" inspection and analysis.

In order to perform data analytics one needs three important components: *i)* data, *ii)* domain knowledge, and *iii)* mathematical tools such as algorithms, optimization, and statistical models (Fayyad et al. 1996). This paper is the second part in a three-part project aimed at utilizing design analytics tools such as machine learning (methods where computers learn and find patterns and correlations in data without users explicitly defining the correlations) and big data analytics (statistical methods for large amounts of data) for analyzing and guiding product development projects. The three components of data mining and data analytics are focused in each part of this project; the first part (*Data*) investigated how a database containing the Engineering Change Requests (ECRs) that were issued during a product development project could be explored and visualized (Arnarsson *et al*., 2016).

In this paper, we build on the previous work by empirically investigating product developers' need for information about ECRs during development projects (*Domain knowledge*), and the how design analytics techniques can meet these needs. It is crucial when preparing for (automated) data mining to understand how the end user or the domain expert will (and wants to) use the extractable information. In this setting this means understanding the product developer's information needs, and how they can use the results as decision support in future product development projects. The product developers' input into the data mining process will provide important information for us when developing the statistical models and machine learning methods that will be developed during the final part of the project (*Statistics, Optimization, Machine learning*).

The aim of the work outlined in this paper is thus to identify product developer's needs for analysis of ECRs in the product development data and to understand the type of design analytics techniques that would best support them in making better decisions in future projects. The study was carried out as an interview study. Specifically, we addressed two research questions:

1. *What answers are product developers looking for from design analytics on ECRs in product development databases?*
2. *What design analytics methods and tools can be used to make better decisions in new product development projects, in particular relation to ECRs?*

2 EARLIER WORK

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 on data, combined with explanatory and predictive modeling. The models and analysis provide a fundament 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 and Frank (2005).

Previous information need-focused studies on analytics include Bichsel (2012), who interviewed four focus groups on how they related to analytics. Bischel's interviews covered data, analysis, strategic decisions, decision-making, and the culture and politics around analytics. He highlighted the balance between the benefits and challenges that people are faced with when working with analytics. Bischel argued that analytics should start with a strategic question and a plan to address that question with data, that analytics should be viewed as an investment and not as expenses, and that analytics does not require perfect data or culture but should be initiated when readiness and commitment is achieved. 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 them in their decision-making. They concluded that the use of analytics techniques is growing, but also that people relate to them in different ways. Similar to Bichsel, LaValle *et al.* argued that when starting an analytics implementation project, there needs to be a clear question backed up with 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 design engineers' needs for more design information than only standard design documents when developing complex products, and Reich (1997) who 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. Hicks (2013) 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. Arnarsson *et a*l. (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 quantitative and text data analysis, for data visualization and exploration, and for pattern identification and analysis.

Recently, the term "Design analytics" has been proposed by Horn *et al*. (2012), to identify the area of research that focuses on processes and tools that 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 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's needs in the late stages of the development process.

In conclusion, data mining and analytics ("design analytics") can provide an insight on a wide variety of uncertainties that many companies are facing. There exists some previous research within data design analytics for product development based on project and customer data but the 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 is not trivial as the data is so abundant and not necessarily produced with the intent of ultimately answering a specific question. This paper addresses this research gap, and thus aims to identify the strategic questions (hypotheses, information needs) that product developers have towards ECR data so design analytics can be applied.

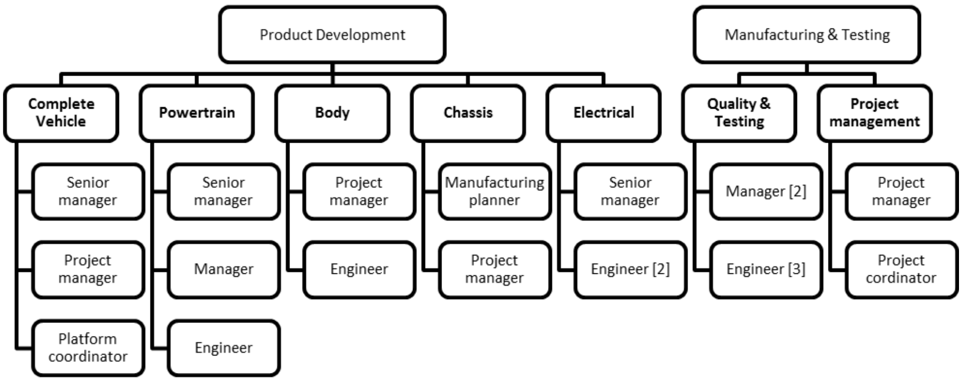

Figure 1. Graph describing the interviewee sample

3 RESEARCH APPROACH

The empirical part of this study is based on semi-structured interviews with experienced product developers. A heterogeneous purposefully sampled group was selected for the interviews (Figure 1). Twenty interviews were conducted with people with diverse experiences throughout the product development process, ranging from engineering, 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 role within company and previous projects. The behavior questions queried how and how often person came in contact with ECRs, and asked them to describe the process of how the ECR reports are handled, including their own involvement. The values/improvement questions were directed towards the ECR data and process, technical difficulties and reoccurring errors. Some examples of value/improvements questions are: *"Is there something that we can improve on in the product development process, where we need data to back up and learn from?"* and *"Is there a lack of analytics on ECRs from historical projects or ongoing projects that can support new decision making?"* The wrap up section was used for more open conversation were interviewees could speak freely and also provided the respondents with an opportunity to add to their statements so far.
The interviews were performed in person or over Skype. The interview guide was handed out before the interviews to help the interviewee to prepare. The interviews were recorded and transcribed. The duration of the interviews was approximately 60 minutes.
The interview data analysis followed the recommendation of (Bryman & Bell, 2011) and started by codification of the as topics and then grouped. Each topic was then phrased as a functional information need, e.g., "*Identify repeated ECRs*". The topics were arranged into sections. Finally, two main categories of information needs emerged, i.e., needs related to data mining and analytics support, and needs related to process and data quality. See Figure 2.

4 FINDINGS

In this section, we present the findings from the interviews. The section thus summarizes the information needs of product developers. The focus is on needs that are not fulfilled by the current IT support, and thus opportunities for application of data mining and design analytics. The findings are described in two main sections, one needs for data mining and design analytics, and one on process and data quality improvement. They are further decomposed into statistics in Figure 3.

4.1 Product developers needs for data mining and design analytics

Three main categories of data mining and design analytics needs were identified: a need for more advanced searches within single ECR databases, a need for integrated searches across multiple databases and a need for performing analyses of ECR process lead times and paths.

4.1.1 Perform advanced searches within a single ECR database

It was stated in the interviews that repeatable ECRs occur frequently and information and data about them resides in the ECR database. There are areas that have repeated quality problems, but which areas are they and why? These are typical questions a product developer would like to answer before

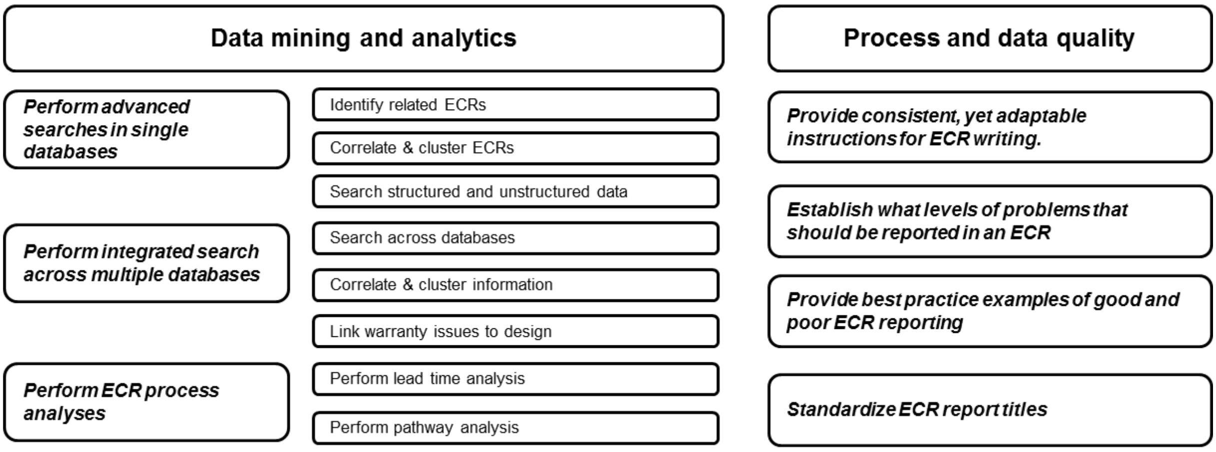

Figure 2. ECR-related information needs and issues as expressed by the interviewees

initializing a new project. There can be quality problems that were solved, and then another problem on the same part occurs later. This could be a result of changes in other areas.

Identify related ECRs. The interviewed identified two types of related ECRs: structurally related or contextually related. Structurally related ECRs happen when there are several ECRs on the same part or system, or when there is a chain of ECRs whose resolution might affect each other.
Contextually related ECRs are ECRs that deal very similar issues or problems, e.g. corroded parts. The product developers suggested that search functions that identify contextually related ECRs within the same project could also be used to identify similar ECRs issued in earlier projects. Finding more relevant and detailed information on what quality deviations have been seen in previous development project and would enable making a validation plan with an emphasis on early testing before starting with the new design process. It would also ensure better risk analysis and therefore enhance pre-studies on earlier projects. The interviewees claimed that this would save time and money on testing and contribute to a more robust design.
Correlate and cluster ECRs. The interviewed product developers further wished to be able to do a correlation (i.e., clustering) analysis of ECRs in previous projects, to know how to prepare for an issue, and to learn to locate repeated problems to eliminate their root causes. Searchability on this could also enable the less experienced developers to find such relationships by their own, and not only by relying on seasoned developers who in the end might not be able to identify this relationship. *"This can be useful because there is never the same project, but there is always a similar project that has been done." (Senior manager powertrain).*
Search both structured and unstructured data. ECRs are a mix of structured data (i.e., numerical data and time stamps) and unstructured data (i.e., text inserted by product developers). The interviewees acknowledged that the firm's IT systems are very mature when it comes to searching for structured data. However, the interviewees also noticed lacking for support for searching within unstructured data. Most of the knowledge documented in ECRs is written into text; report title, description of the challenge, action taken, comments, and technical solution. When searching within an ECR database based on structured data such as part number or function, there is therefore a risk that relevant knowledge is not found. The product developers expressed a need for the ability to search numerical and text data at the same time. That would expand the search and give more relevant results back. *"There is "hidden" information in the text body of a report." (Senior manager powertrain).*

4.1.2 Perform integrated search across multiple databases

In product development, multiple systems are often built to serve different needs and the result is that data is stored in different databases. Databases accessed during product development include the CAD vault, product structure, ECR, and warranty databases (see Figure 4). For efficient problem investigation and resolution, it is important to be able to, e.g., connect the current warranty data from the field back to the product development data. Further, document management systems contain lessons learned from old projects that can provide relevant information if a correlation is made to the CAD or product structure database. However, support for this kind of search is lacking, according to the interviewees. *"There exists a tremendous amount of knowledge from old projects that we do not use in new projects." (Senior manager powertrain).* The interviewees explained that there is knowledge in the form of "lessons learned" written text in the document system, but to be able to find the knowledge you are looking for in relation to a new design can sometimes only be possible for more experienced developers who know where to look.

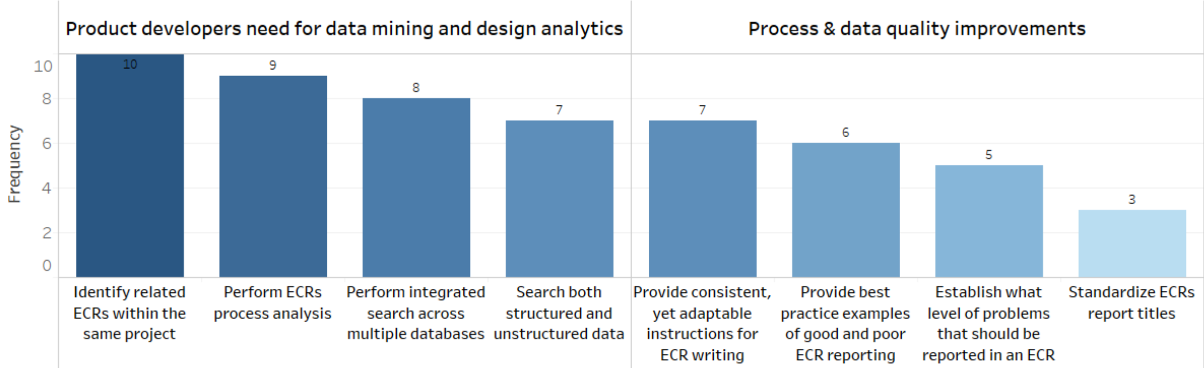

Figure 3. Frequency of information needs topics, as stated in the interviews

Another integrated but manual search and analysis of this kind is performed today when a warranty issue arises and it is manually distributed to the relevant department for corrective action. The interviewed product developers requested a functionality that would enable them to search for related files across multiple databases, find the root cause for a problem and correct the possible fault that was e.g. created in an earlier development phase. The interviewees argued that such functionality would also enable proactive explorative investigations on the current warranty problems the company is faced with and take them into consideration before the next design. They said that it was important to ask the question why the problem occurred in the first place, why a problem was not foreseen in the design phase and what would be the appropriate testing methods to detect them for upcoming projects.

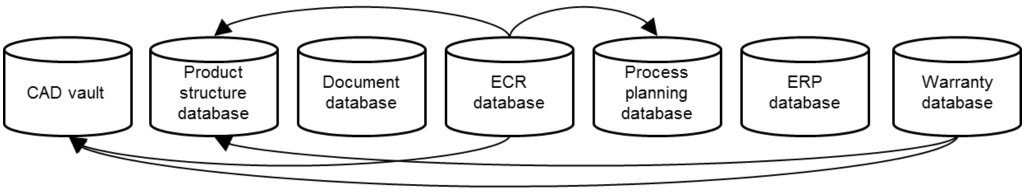

Figure 4. Illustration of the multiple databases that product development companies often have and some of their links

4.1.3 Perform ECR process analyses

Perform ECRs lead time analysis. The firm logs ECRs according to project numbers and a single project number can contain from hundreds up to tens of thousands of ECRs. The process starts with "report written", and is structured into stages such as investigation, testing and verification, and finally ending with a status "report closed". Each ECR contains a timestamp for status changes, having the possibility to take on 40 unique statuses. ECRs can have different owners during the process. There are intersections between engineering departments where the ownership of the ECR is not clear or changing to a new owner. An ECR can even be in between owners during some periods.

Many interviewees mentioned that ECR lead time was of recognized importance and for a long time the total lead time for ECRs had been tracked. However, the interviewees also identified an opportunity to break these lead times down into sections related to particular stages or tasks and compare on different levels to identify slow and fast tasks. *"By looking at sectioned lead times we can start to identify waste in the system and understand who is sleeping? It is very obvious that there is sleeping time here." (Senior manager complete vehicle).*

Perform ECR pathway analysis. The multitude of statuses means that ECRs take many pathways from report written to report closed. Some interviewees proposed that the pathways should be analyzed, e.g. with regards to how they are distributed from one department to another.

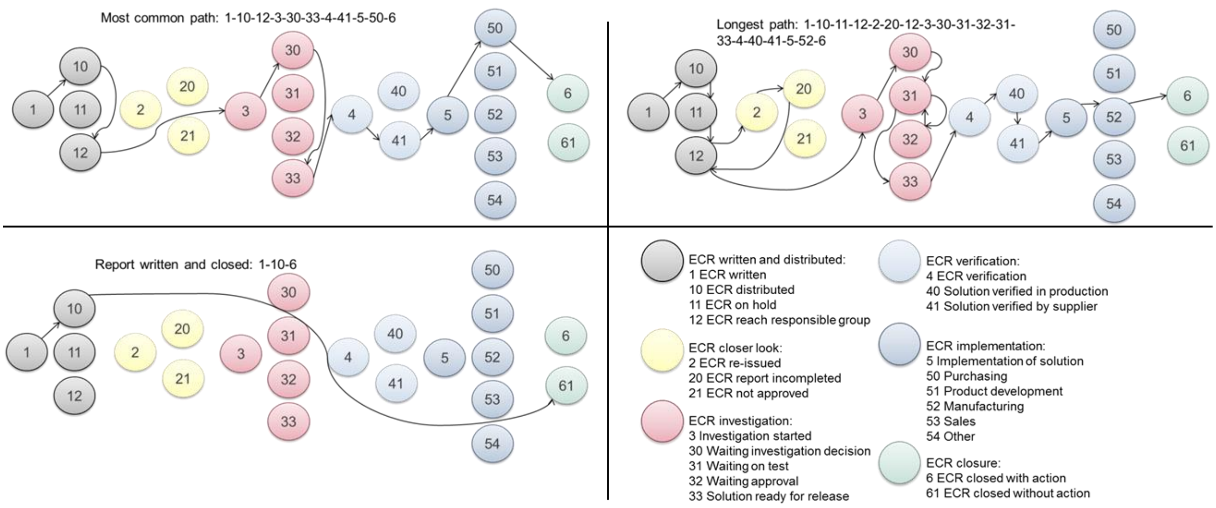

Figure 5. Illustration of different ECR pathways.

Rather than addressing a current shortcoming, this was claimed to open up opportunities for deeper understanding and improvement of the ECR process: Does it take longer time to distribute reports to some departments? Can we improve the time it takes for reports to be assigned to a responsible department? Can we identify process issues? *"If a problem is not distributed for two weeks then there is possibly something wrong." (Project manager chassis).* ECRs can be incorrectly distributed and the receiving department will not accept it and does not redistribute it, this should ring an "alarm clock" after some time has passed. Another ECR path analysis that the interviewees mentioned as interesting would be to look at what percentage of ECRs reports go from being written and take on a closing status directly after (see Figure 5).

4.2 Process & data quality improvements

The interviewed product developers further reported a number of problems with the current environment, as outlined below.

Provide consistent, yet adaptable instructions for ECR writing. Many interviewees stated that following the ECR process instructions was a complicated and involved too much of administrative work for minor ECRs. If product developers follow the instructions to the fullest it will slow things down, they feel like it is a burden and sometimes shortcuts are taken to solve the issues faster. The interviewees also emphasized the importance of consistency when writing ECRs so that similar issues reported from different locations are not described differently.

Establish what levels of problems that should be reported in an ECR. The interviewed product developers argued that there is often uncertainty on what the severity or level of problems should be reported. Some expressed doubts for why an ECR report needs to be written and what purpose it has. For some product developers ECR reporting is associated with a substantial amount of administrative work to log minor issues that could be solved with a face-to-face discussion. This makes reporting less consistent due to that some development groups report all quality issues and others only major ones.

Provide best practice examples of good and poor ECR reporting. The interviewees claimed that well written ECR reports reduce lead time and confusion of issues. However, they also claimed that a poorly written ECR could set the change request up for failure. ECRs should contain the relevant information, without being too much hustle to write the report. *"The knowledge of the report issuer is probably the biggest factor of poor quality of reports." (Engineer quality & testing).* A measure that had already been taken was to reduce the number of people who are allowed to write the reports, because they said it started to lead to higher quality of reports and lower inflow of duplicated/poorly written reports. In order to communicate a clear understanding of how a good ECR is written the interviewees requested that a best practice ECR should be provided, along with good compact visual examples like an A4 page of instructions that highlight the most important aspects.

Many interviewees, who argued that more pictures should be included in ECRs, mentioned the value of photographs for documenting ECR issues. It was pointed out that the photographic information should not be limited to the problematic part but also the surrounding environment (important when working

with prototypes). It was claimed that A really well worked through context would help the product developer to understand the issue and also save cost (hours needed to solve issue).
Standardize ECR report titles. It was argued that the title of an ECR report should be written so you understand the issue in 5 seconds. Consistency in report title was viewed as particularly important when overseeing larger projects with large numbers of reports written during the process. The interviewees suggested that there should be a standard way or even a tool for writing titles along with examples for people to quickly understand the best practice for it.

5 DISCUSSION

Let us now consider the stated research questions in relation to the findings and ideas for future work.

5.1 RQ1: *"What answers are product developers looking for from design analytics on ECRs in product development databases?"*

We found that the most frequent request from product developers working in complex product development projects that have generated ECRs is that they wish to be able to make a connection between ECRs and other documents and models related to product development. Product developers are looking for easier ways to find relevant historical information as a part of a pre-study before making a new design. They would like to identify related ECRs within the same project and to be able to compare ECRs from passed projects. This information exists in various databases that different departments are responsible for, but comprehensive searchability across the databases is missing. That searchability should include structured and unstructured data as it was stated that there is hidden information in the systems about similar previous designs and service experiences that current searches do not cover since they are limited to structured data. This would break down knowledge barriers and enable product developers to do their own mining of the data. Ultimately, they want the answers that could support answering questions like what to test for, how much testing is needed, what are the potential problems, and what are the risks they need to consider before dealing with a similar or new design. They would also like to link current warranty issues back to the design. They do this manually today but are calling for a function that can identify this automatically for them.
Another question product developers had was to do detailed ECR process analysis, where lead times and paths of ECRs are broken down. This can be analysed on design group levels, groups of parts, or on projects with short and long lead times. Comparison can be done on success factors and outliers, and by asking why e.g. some department move faster through specific project sections than others. A breakdown of the sections could help to identify communication issues between departments, and identify department cultural differences regarding e.g. awaiting decision periods.
As noted above, the interviewed product developers also pointed out a number of process 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 avoid to report quality issues and thus gives 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 for efficient design analytics is *sufficiently* good data quality.

5.2 RQ2: *What design analytics methods and tools can be used to make better decisions in new product development projects, in particular relation to ECRs?*

There are many potential design analytics methods that can promote a more systematic and easily accessible uses of the data stored in ECR databases and thus improve decision-making in new product development projects. A few examples are discussed below, in relation to the identified information needs.
As noted above, the most frequently mentioned information need was the ability to identify related ECRs. This could be with ECRs from the same project but also with ECRs from earlier projects. This would help manage risks with a new design and to anticipate impact of simultaneously on-going changes. This task could be supported by using *information clustering* machine learning algorithms. Clustering methods could also help address the issue with related information residing in multiple databases: *"The ECR database is the system we should use for knowledge management and we have so much data in the system. But what we are missing is to connect and combine the warranty system to the ECR database. That will be all the data we really need." (Senior manager quality & testing).*

Further, many interviewees noted the rich knowledge is stored in the text body of reports, and that searches today are based on numerical data and pre-categorized functions groups. There is a need for tools that enable search criteria that include numerical and text data. They expressed needs that could be addressed with *text search and classification tools.* Using (text) *Pattern identification tools* in a similar way to plagiarism detection tool to would also enable comparison of complete ECR reports, with the difference here that a high degree of resemblance would be positive – it would signify an ECR for which the process and likely the solution could be re-used.
The second most mentioned need related to understanding variation in lead times and pathways when carrying out ECRs. Understanding such variations would provide data on process bottlenecks etc. that could inform ECR process improvement effort. Design analytics could utilize available ECR data such as responsible department, timestamp and status in order to, e.g., *visualize* the differences between departments. This could be done with a matrix with departments and ECRs process steps on the axes and lead time in the cells (see fictitious example in Figure 6). Path analysis could also be done with data mining and the benefits of that are to identify slow and fast periods for ECRs and compare between departments to find a best practice process. This could also identify waste in the system such as how many ECRs are written and then directly closed without any action.

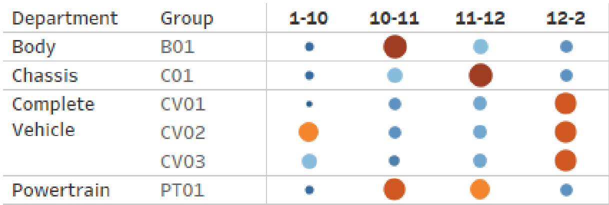

Figure 6. Fictitious ECR lead time breakdown analysis visualized in matrix form

5.3 Transferability of results

This study was conducted in a large multinational firm that develops and manufactures commercial vehicles. The firm’s way of managing ECRs is typical for other large firms that develop complex systems and products, for example in the aerospace and defense industries. The ECR data structures and processes are similar. We therefore argue that the findings related to needs and possible solutions can be transferred to these contexts. However, the interviewed product developers mainly had a mechanical or electrical background. It is possible; even likely, that software developers have somewhat different information issues and needs.

6 CONCLUSIONS AND FUTURE WORK

The study chartered out the information needs and interests of product developers in a large multinational firm, focusing on a central product development document type, i.e., engineering change requests (ECR). The aim was to identify questions that are difficult to answer using current IT systems, but that may be addressed by applying data mining and design analytics methods and tools.
The interviewees confirmed that the amount of ECR data that is collected today is not used for pro-active purposes, and that they saw the potential for data mining and design analytics tools to help with that. The fact that data is collected and stored throughout the product development process is a good start for an analytics implementation project.
In the study, we interviewed twenty product developers. Two main categories of information needs were identified. The first was directly related to data mining and design analytics capabilities, and included requests for functionality for comprehensive and flexible search in databases with 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 improvements of the process and data quality of ECRs and its related process: The developers were concerned about the quality and consistency of how ECR data is logged and want to communicate best practices about that for all groups within the organization.
One overall reflection of the interviews is that the developers wanted more easily accessible tools for analyzing previous projects and to gain knowledge of pitfalls and risks before they choose specific designs for the products. Ease of use and short learning times are essential for data mining and design analytics tools, if they are to be used by a majority of product developers.

As continuation to this study, we plan to use the knowledge gathered from the interviews to evaluate and further develop existing and novel data mining and design analytics tools to support developers. The ECR data from previous product development projects is available and the questions that the developers are asking are now identified. Another interesting avenue for future work is to investigate if and to what extent the findings can be transferred to software development.

REFERENCES

Arnarsson, I.Ö., Malmqvist, J., Gustavsson, E. and Jirstrand, M. (2016), “Towards big-data analysis of deviation and error reports in product development projects”, *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

Bergman, B. and Klefsjö, B. (2010). *Quality from Customer Needs to Customer Satisfaction*, Studentlitteratur, Lund, Sweden

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

Bryman, A. and Bell, E. (2011), *Business Research Methods,* 3rd ed, Oxford University Press. Oxford, UK

Clark, K. B. and Fujimoto, T. (1991), *Product Development Performance: Strategy, Organization, and Management in the World Auto Industry*, Harvard Business School Press, Cambridge, MA, USA

Data, I.B. (2015). *IBM - Bringing Big Data to the Enterprise*

Davenport, T. H. and Jeanne, G. H. (2007), *Competing on Analytics: the New Science of Winning*, Harvard Business Press, Cambridge, MA, USA

Davenport, T.H., Harris, J.G. and Morison, R. (2010), *Analytics at Work: Smarter Decisions, Better Results*, Harvard Business 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

Giffin, M., de Weck, O., Bounova, G., Keller, R., Eckert, C., and Clarkson, P. J. (2009), “Change propagation analysis in complex technical systems” *Journal of Mechanical Design*, Vol. 131 No. 8, doi:10.1115/1.3149847

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

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

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

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

Reich, Y. (1997), “Machine learning techniques for civil engineering problems”, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 12, No. 4, pp 295-310

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

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. and Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, USA

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

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