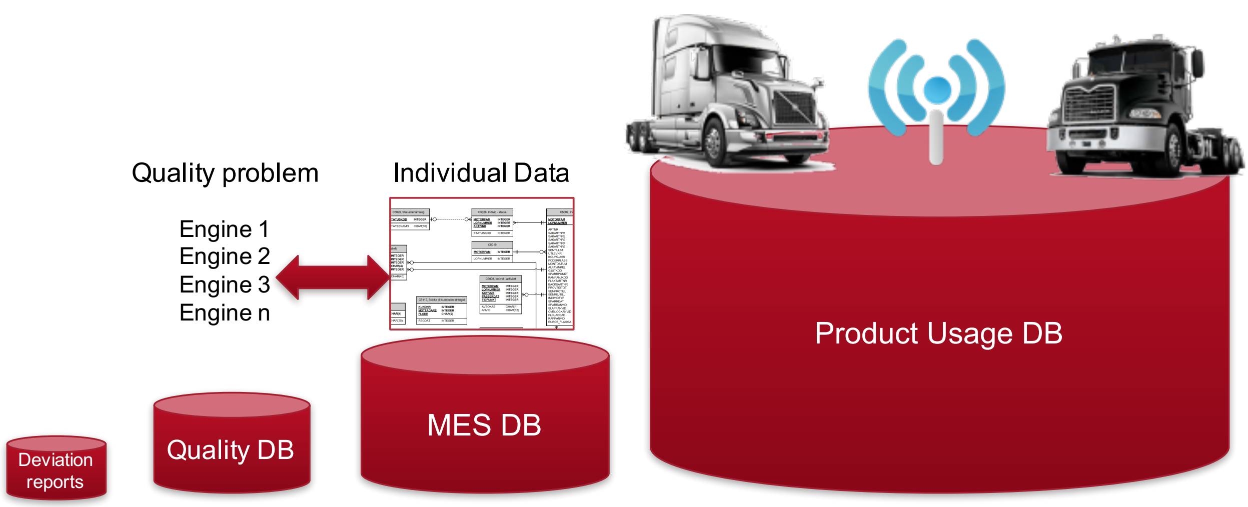
PROPID:

PROduction development using Product Individual Data



Författare: Dag Bergsjö

Datum: 2017-12-20
Projekt inom: BADA (Big Automotive Data Analytics)

FFI Fordonsstrategisk Forskning och Innovation

Contents

1 Summary ... 3

2 Sammanfattning på Svenska ... 3

3 Background ... 5

4 Approach ... 6

4.1 Objectives ... 7

5 Results ... 8

5.1 Case I: R&D ... 8

5.2 Case II: Operations ... 10

5.3 Case III: Aftermarket ... 11

6 Analysis ... 13

6.1 Analytics ... 13

6.2 Opportunities ... 14

6.3 Barriers ... 14

6.4 Concluding the Data Mining Experiment ... 15

7 Publications ... 16

7.1 Knowledge and Dissemination of Results ... 16

7.2 Publications ... 16

8 Conclusions and Future Research ... 16

9 Participants and Contacts ... 17

1 Summary

The aim of this research is to enable a paradigm shift in the automotive industry, to shift to an information centric product management system where product individual data can be correlated to big data, in order to predict future product behavior or problems. The focus of this research project is not how to improve big data algorithms or collection methods, but how to utilize knowledge derived from big-data downstream in the processes.

To enable this shift, the project will focus on a bottom up approach (component-product-engineer-management-society) to investigate how information is utilized to improve efficiency and to improve quality of design decisions and ultimately extend the product usability and lifetime. As a case study, different valuable components of heavy trucks will be used to identify difference in how the product usage can affect the durability, sustainability and lifecycle performance of the product, both as a part of an individual product system, or as integral parts of a fleet of products.

There is a long delay between the time when changes concerning product and process design are implemented and the time at which the consequences of those changes (in the market) arise. The consequence is that the company loses out on the most valuable source of feedback concerning its product and process improvement, namely its market and customers, because the large delay makes it hard to draw correct cause and effect conclusions. PROPID aims to increase the speed and accuracy of feedback from outcomes of product and process decisions by correlating data collected from product individuals with data collected from the production processes and aftermarket processes and using big data analytics to transform those data streams into useful knowledge by identifying patterns. The purpose is to help the quality functions at the OEMs to more quickly and accurately reveal cause and effect relations in order to support continuous improvement of products and processes.

The approaches used are interviews and observations presented as case studies, where people with connection to Big Data gives their view on the subject. It also includes a data mining experiment where a large set of data is examined in order to get a deeper understanding of requirements and barriers for working with Big Data analytics. The project concludes that there is a potential for knowledge-driven product development powered by Big Data. Actual data on product and user behaviour can be used to make decisions on product design, errors can be detected and feedback-loops can be shortened.

As this project is a Pre-Study, the project included not only practical activities to investigate and determine the potential for Big Data analytics within a product development setting, but it also focused on identifying potential future needs, perceived barriers and to broaden the network for a full scale, larger project.

2 Sammanfattning på Svenska

PROPID syftar till att öka hastigheten och noggrannheten av återkoppling från marknaden genom att korrelera data som samlats in från produktindivider med uppgifter som samlats in från produktionsprocesser och eftermarknadsprocesser.

Vi valde att rikta in projektet mot produktionsdata från en motorfabrik och undersökte historiska data både manuellt och med hjälp av kvantitativa

analysmetoder. Syftet var att identifiera korrelationer mellan kvalitetsproblem och tidigare sensorutfall i fabriken med syfte att stödja kvalitetsfunktionerna i produktutveckling samt att underbygga ständiga förbättringar.

Projektet hade en förstudekaraktär, därför har nedslag gjorts i den tilltänkta kompletta kedjan. Främst gjordes nedslag inom just motorfabriken där vi använde ett verkligt fall och sökte korrelationer mot två på marknaden existerande problem. Därutöver ville vi inom projektet undersöka möjligheten att använda samma typ av analys för att förutspå och stödja beslut kring underhåll, renovering och skrotning. Denna del undersöktes främst genom Lastbilsåterförsäljarens Rejmes medverkan. Idag sker en omfattande renovering av högvärdeskomponenter som t.ex. turbo, kompressor kylvätskepunp startmotor och generator. Det sker även en utveckling av fjärrövervakning av lastbilsindivider för förebyggande underhåll kring högvärdeskomponenter. Vi har valt att för ett uppföljningsprojekt lämna hypotesen med korrelationsanalys och gå vidare med att i utvärdera ”Machine learning”. VI har också valt att gå vidare att titta på avvikelserapporter inom produktutveckling snarare än produktionsdata. Under projektets gång har det blivit uppenbart att avvikelserapporter har en tydligare struktur som lämpar sig bättre för analys, i synnerhet för problem som löses tidigt i utvecklingsfasen.

Huvudmålet med projektet var att undersöka möjligheten att skapa kunskap utifrån existerande ”big data”, utan att erfarenhetsmässigt behöva genomgå proceduren att först lära sig av sina misstag. Dvs. att på något sätt kunna förutspå problem genom att analysera big data. Denna hypotes har inte i detta projekt kunnat bekräftats, utan projektet har snarare öppnat upp fler och nya frågeställningar inom ämnet. Vi har genom att genomföra praktiska tester på stora datamängder hos fordonsutvecklaren kunnat klargöra att produktionsdata inte kunnat korreleras, på ett verifierbart sätt, mot problem på marknaden. Vi fokuserade analysen på två olika problem, det första involverande 6 stycken motorindivider och i det andra fallet ett problem som drabbat över 200 individer. I första fallet saknades individerna i databasen, vilket gav den uppenbara indikationen att databasen inte var komplett. I andra fallet hittade vi de flesta individerna men det gick inte att identifiera några tydliga korrelationer mellan dessa individer och den övriga populationen. Ett tydligt resultat är att existerande data är i stort odokumenterad (eller att dokumentationen inte stämmer överens med verkligt insamlade data), samt att det saknas data. Utfallet betyder inte att det inte går att bygga kunskap på data utan att analysmetoden inte var den rätta för en så pass ostrukturerad datamängd som databasen visade sig vara. Framtida forskning ska därför riktas in antingen på datakällor där det finns en tydligare struktur och dokumentation eller på att först generera strukturerade data som kan användas. Ytterligare mål för projektet har varit att bygga ett konsortium för ett fortsättningsprojekt, om vi i våra initiala analyser fann intressanta data att gå vidare med. Vi har i synnerhet knutit nya kontakter inom både fordonstillverkaren men också med FCC inför ett uppföljningsprojekt.

Projektet har genomförts i samarbete mellan Chamers, AB Volvo och Rejmes Transportfordon. Konsortiet möjliggjorde ett fokus på problem som uppstår i utveckling eller produktion och som ger efterverkningar på marknaden. Upplägget som sådant var därmed värdefullt och gav många insikter, som inte hade kunnat fås utan en utökad värdekedja. Fokus på datainsamling kring utbyte/återtillverkning av högvärdeskomponenter som t.ex. turbo och kompressorer möjliggjordes tack vare denna ansats.

3 Background

This project focuses on utilizing data from various stages of the product lifecycle. The data that is collected, is to be analyzed and knowledge can be created as an effect of this analysis. As an academic research field, this project is connected to Knowledge Management (KM) and Product Lifecycle Management PLM. The academic area of PLM focuses on making sure that data and information created during one phase of the product lifecycle (e.g. development) can be used to support the work in another phase (e.g. manufacturing preparation or even manufacturing operation). In practice the concept of PLM has led mainly to focus on a feed forward flow of digital product representations (e.g. 3D models) from development to preparation of manufacturing or aftermarket systems for a specific product. There is however a vision in the academic visions of PLM to also feed data and information **back** to enable learning in the development organization by being able to connect cause-and-effect chains between product/process decisions and their outcomes in terms of product qualities or process qualities. In practice, however this remains a vision because the data flows which are generated in the different lifecycle phases (production, usage, sales, aftermarket) are completely disconnected and feedback concerning poor product or process decisions reaches those who were involved with them only when serious quality problems are detected through the quality functions.

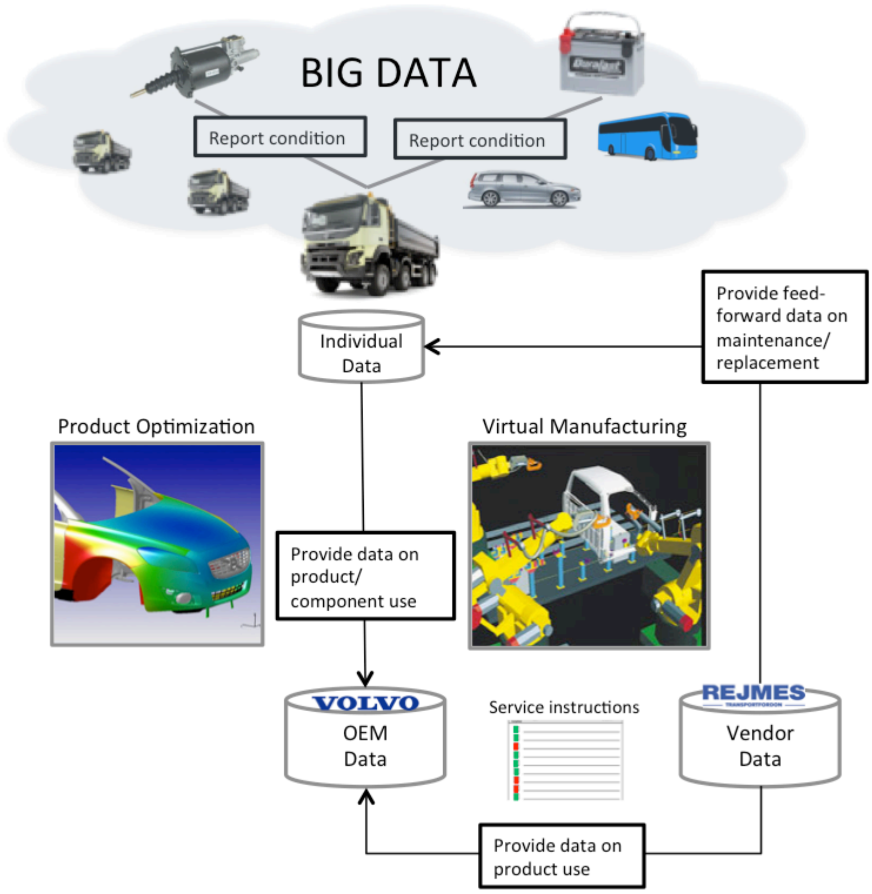


Figure 1: The individual data is used in relation to (Big Data) to optimize product utilization. Focusing on the relationship and business models between OEM and supplier.

Another source of feedback are product planning functions which scan the market feedback and benchmark competitors to produce new requirements. Yet another source of feedback (as seen from the perspective of development) is testing but the design of test protocols relies on data coming from reality to be able to represent relevant test cases.

From the perspective of a knowledge-oriented company that strives for continuous learning and improvement this is a serious problem. The problem exists in all traditional product developing and producing companies which develop and sell their products on a global scale and where the product production, usage and maintenance is spread across large geographic distances and dispersed across many different actors.

One of the major root-causes to the problem of disconnected data flows lies in the existing database oriented technologies that underpin the systems used to manage the data during development (PLM), supply chain (ERP) and manufacturing (MES) and later in aftermarket. The data and information models employed in each of these sources are usually adapted to the specific needs of the processes they and now also each product individual is generating data (according its own logic that is adapted to the process of operating the product itself). This means that there are challenges in merging the data flows by merging the databases because the underlying logics cannot be merged.

Big Data have the potential to have radical impact in PLM, and will improve many areas e.g. manufacturing, R&D and supply chain (Manyika et al., 2011). It will optimize assembly processes, reduce cycle time, and elicit customer needs so the future of Big Data in PLM is promising (Li, Tao, Cheng & Zhao, 2015). Thus far, much of the PLM research has focused on specific areas in the lifecycle, not as much has focused on the chain as a whole (Li et al., 2015).

The use of Big Data is expected to increase three trends may drive this change in coming years:

- Data will not only reflect performance, but also drive the business operations
- Organizations will aim for a more holistic approach to data and analytics, connecting the different departments of the business. This will allow for data management and analytics from the core to the edge of the business and possibly between different lifecycles.
- Executives will use data and analytics to shape business strategy, allowing new roles for analytics professionals.

With this in mind, studying the change and development of Big Data management and analytics in such a business as automotive is of high interest and relevance.

4 Approach

The purpose of this project was to investigate the possibility of creating knowledge based on existing big data, without having to experience the process of learning from mistakes. Hence, the focus has been to predict problems by analysing big data. Basically, this was both a feasibility study and a high-risk project and the difficulty of identifying potential knowledge gaps was evident. The approach was to identify useful by correlating the data from the different data flows and that this data can used to create new knowledge concerning product and process design and the outcomes. The research questions were:

RQ1: How can data collected from the different lifecycle phases of product individuals be analyzed to identify patterns that indicate correlation or causation between events across the different phases.

RQ2: How can the resulting knowledge be incorporated into existing frameworks, processes, information systems and tools?

On a similar note, the feed forward of data and information from production, usage and aftermarket phases to support development of processes for remanufacturing/reuse of components as well as to support the execution of such activities can benefit from correlation of data collected at the level of product individuals instead of general classes of products (as represented in the PLM system).

The project consisted of four work packages. One that aimed to collect data from the different sources and fine tune the analysis. One that aimed to demonstrate how the output from the data analysis was used in a specific case to produce useful knowledge. One that aimed to demonstrate how the output from the data analysis could be used to optimize remanufacturing/reuse decisions for specific product individuals, and finally there was a project management and dissemination work package that was used to coordinate the other work packages and communication concerning results.

A master thesis project was also set up to support the researchers, and was conducted during spring 2017. The project was managed through regular meetings in the project group primarily involving representatives from the project participants.

4.1 Objectives

PROPID had two main goals:

1. To increase the speed and accuracy by which product developing companies can create new knowledge about their own products and markets based on actual data generated during each products individual lifecycle.
2. Maximize the value of each product individual by using each individual's history to extend the lifetime of high-value components.

These two project goals contribute mainly to the FFI program goal to increase the innovation capacity and competitiveness of the automotive industry in Sweden. In addition, the project setup contributes to the goals to promote collaboration with small and medium size enterprises as the consortium includes Rejmes Personbilar, and it also contributes to the goal to promote cooperation between industry and universities as the algorithm developed at Chalmers is used in a new industrial setting to achieve new goals.

More specifically the two project goals contribute to the goals of the BADA program to use big data analytics to develop businesses as well as to use big data analytics to develop new services. In addition, the second project goal, of maximizing the value of each product individual through reuse or remanufacturing of existing components has a direct impact on sustainability as this will also reduce scrap and energy consumption concerned with manufacturing new components.

5 Results

In this chapter, the results from the four different case studies performed are briefly presented. For a complete picture of the results and the analysis the master thesis is referenced.

5.1 Case I: R&D

In this section, R&D professionals with relations to the case company are interviewed and asked about Big Data related topics. Three of the interviewees have special knowledge about batteries, and three have special knowledge about data collection and analytics.

The main topics that have been covered is the following headlines, data, analytics, opportunities and barriers.

Data

Much focus from the R&D perspective is on the data itself. Big Data poses many interesting opportunities and those will be elaborated on later, but the characteristics of such data also leads to many requirements. The data trend has lead to problems with data storage and collection argues one interviewee, and the ability to collect data is vital. R&D professionals brings up many different aspects of this. Some of the interviewees focus much on storage itself and how this could be done best. A view is that more data is collected via cloud solutions, and that the trends goes towards that. On the other hand, it is still common to collect data in batch instead of streaming it live. Examples on these occasions are when trucks are at service vendors, or when products are collected for quality checks before assembly. According to some interviewees a big challenge will be how to get clean quality data from large data-sets, and uncertainties in organizations regarding how to handle this might be a problem. One opinion often mentioned by the R&D professionals is the need to standardize the data recording in the future, just to make sure that the data is of enough quality. High veracity data is not enough, a large part of the interviewees talks about the difficulties with monitoring the data once it is collected. R&D professionals in this study says that it is important to know what data is really necessary before collecting it. By knowing this beforehand there is no need to waste storage on data that is not useful, and the analysis becomes faster and more manageable. One with knowledge in Big Data analytics says that it is more important to have a flexible Big Data solution that rapidly can change what data is collected, instead of just collecting everything. Furthermore, the person argues that monitoring all data is very difficult. An interviewee says that breaking down data is necessary, and that it has to be filtered before one starts to work with it and analyse it.

Analytics

The second topic discussed to a larger extent is analytics. Monitoring data is hard so analysis needs to be done automatically says one with experience in Big Data analytics. "The ability to collect data, and conduct analytics that are being sent back to the customer is vital", he argues further. The old way of doing it is to send back data and then conduct the analysis. But a common view is that it is hard to monitor all the data that comes in, and that the analysis has to be done instantly and automatically. This is so important that it is considered a bigger challenge than data warehousing, especially in automotive. Better to design data collection and analysis in the consecutive steps:

1. Aim
2. Design collection/infrastructure
3. Statistical analysis

Interviewees in this study are mentioning some key points regarding Big Data analytics. When talking about monitoring data from batteries, one interviewee raised concerns about monitoring a large fleet, just because each battery is unique in terms of previous usage. This might make it hard to draw conclusions from mean values according to the interviewee, just because it might not be valid for the battery sample. With that said, it is still common to look at large samples and compare it to individual batteries. Another brings up that with large data sets irrelevant correlations are found in data mining, and it is important to be critical to correlations. One interviewee says that it is still many test vehicles they work on in terms of Big Data, and that simulations are more common than testing vehicles in the field. There is however clear that they think measuring on roads and customer behaviours poses opportunities. In terms of tools, there needs to be filtering and sorting tools, along with easily used analytic tools in order to facilitate data driven decision-making.

Opportunities

Big Data in automotive has large potential and there are many opportunities. Both R&D personnel at the Case Company and at external companies working with Big Data within the automotive industry acknowledge preventive maintenance as a great opportunity. General product intelligence from customer behaviour can generate value to R&D says an interviewee. Another interviewee brings up benefits for both products and services, and adds that text based analysis in automotive is an area of interest in the future. The interviewees explain that much can be accomplished with search and text analysis in organizations. The company’s core is highly applicable on the deviation reports recorded by R&D, which is engineering reports recorded during development. Much of the data recorded in manufacturing is also text based. Insights derived from text based data are often just intuition from the individual. That is because there is a lack of tools for analysing text, which is therefore a business opportunity for third parties. This was also observed during a root-cause analysis session.

Barriers

There are of course obstacles and capability issues that needs to be solved to go forward with Big Data in automotive. The interviewees says that the right competencies are imperative for analytics, and that some roles in companies today might have to be redefined in the future. Today, it is good if you know what to look for, otherwise it is not easy, but there is great potential. Sensible filters and sorting is everything, then it is required in the end, someone who understands the data and find the root cause of the problem. The task of traceability (much focused on in this project) is considered to be a very interesting challenge, and some R&D professionals in this study requests more traceability. However, one barrier for traceability is the use of many different systems. According to Big Data experts, the key is connecting many smaller systems, rather than having one enormous system. Other capability issues that might arise is the fact that many of these projects includes many stakeholders, and that prioritizing between resources is not always easy.

A key issue is the personal integrity connected to this, and this must be closely monitored. There are different perspective on this, one R&D engineer says that changed laws connected to data collection might make the work much harder. At the same time the potential danger with logging more and more behavioural data are apparent and not taken lightly by any interviewee. There are different rules depending on where the market is geographically. At the Chinese market the government demands that data are collected from the vehiclee.

5.2 Case II: Operations

This perspective comes from stakeholders that very often look at and work with quantitative data. Decision-making is both driven by data and tacit knowledge. Data in the form of text tends to introduce a small element of subjective bias which was observed at study visits. For example, if the root-cause of a problem eludes workshop personnel then frustration might be expressed in the documenting process, or data recording because that personnel does the documenting.

Data

The general perception of data is "the more the better". That is concerning the amount of inputs for specific cases rather than a bigger populace to work with. This is due the root-cause analysis before taking action regarding a quality issue from the field. This a time-consuming process where the common denominators is the large amount of information about the products. Traceability thereby becomes an imperative requirement for much of the analysis. It is expressed that "a data-trail across the lifecycle would be great". There are organizational barriers to this but also plain access problems to data. For instance, accessing production data requires getting an administrative middle man that can assist. "Some cases require the extraction of CC’s, critical characteristics that is". Critical characteristics are features of the product where the company guarantees quality for the customer. That type of data origins from manufacturing.

The biggest input to the work comes from the market through the workshops/service providers. The data recorded at the workshops varies greatly in quality and quantity. Some of the data is recorded in a way that gives room for much variation. Text fields where the workshop personnel enter the fault and action taken varies in language, quantity of the description and foremost precision of the description. Standardizing the recording procedure is an obstacle from many perspectives but would minimize the investigative work needed as many hours are spent decrypting what the individual in each case tries to convey.

Analytics

Quality issues are mainly detected via input from the market, where different indexes are monitored. However, prioritization is still an issue and which in itself needs investigative work, specifically digging through vast amounts of data. There is much room for interpretation error in the data-sets, due to bias but also because circumstantial effects on specific cases. One academic example mentioned was that for example if a higher fault frequency would be detected on a high capacity battery , the where the quick conclusion would be that it is something wrong with the battery and we should not use it. However, circumstances are that the customers having that configuration are heavy users of the battery, much more than the ordinary customer. Therefore should the conclusion rather be to offer customers even higher capacity batteries. This is an example of the tricy discussion between failing component (the component that breaks) and the causing component (the root.cause) that is the actual faulty design or component. This is very difficult to identify in both reports and data analytics.

The data trend within production is moving towards utilizing advanced algorithms and models for optimizing and monitor production flows. "It is of interest to analyze deviations in real time in the flow". There are barriers in terms of old systems recording data and competence within the relevant fields such as machine learning etc. A learning deducted from the experiment is that much production data is recorded in text strings entered by the operator which adds to the competence requirement. However, that issue could be exclusive for the case.

The tools used range from products from SAS (Statistical Analysis Software) to more simple products like Qlikview and Excel. "Different data visualizations and sources require different tools", there is also a constraining factor of knowledge regarding some tools.

Opportunities and Barriers

The experiment conducted in this project is considered interesting by the interviewees. Training algorithms with data from later in the lifecycle and correlating with production data could enable prediction of quality issues in the field. However, there are so many parameters and dimensions to consider that it will be very difficult to realize.

An expectation on future capabilities of data analytics is what analyzing more streamed data will bring. Moving toward finding the quality issue before the customer brings it to the workshop is the main foresighted capability. "Things are happening fast, but there still a way to go and barriers along the way". Barriers refers to a notion of the technological but also legal compliance.

5.3 Case III: Aftermarket

The aftermarket has knowledge about the market and are in close contact with customers. Therefore their view on how Big Data can improve the situation for themselves and the customers are of vital importance. To spark an interest for this question, the project focuses on a product that is in the center of attention right now due the massive increase of electronics in trucks. This product are batteries, and it is a product that obviously is vital to further electrical functions including the development of sensors and potential data streaming.

Data

Today, there are service packages available for truck owners provided by the vendors where data are streamed from vehicles and monitored to inform driver when the components are in need for service. The only package where the drivers data are logged and analyzed are the so called "gold" service level. This however is likely to change and include more service levels. When not streaming data directly from the vehicle, the data is collected manually in the workshops. When only collected at workshops, the vendors lose the continuity in data collection. Much focus when talking to Aftermarket professionals have been on batteries, and information such as oil levels, topography and "State of Health" for batteries are collected to determine the life span for the batteries. Normally they can only check 4-5 vehicles at the time. The service vendor has contact with the OEM, and they can make suggestions regarding what logged data they want access to in the vehicle. However, this process takes time and requires interaction with both the OEM and customer.

Analytics
Even with today's data levels, there is no way to monitor all data that comes in, all the time. This would be preferable of course, but it would demand manual supervision of the data, which is not practical. What is needed is an automated analysis done instantly, especially since the levels of data collected are likely to increase. Today there is a signal when something is wrong in a truck, but there is no information regarding what the cause of the problem might be. Someone have to log in and manually check this on a computer (and it only applies to gold service clients). The value today is that the driver can call and ask what the problem is, but as stated the check-up process is manual.

Opportunities
In the business, there is emphasis on developing predictive maintenance, and being able to inform customers bout potential problems before they will notice. This is where there is a potential competitive edge has been identified, in being able to predict what action the driver needs to take and in what time-span. In the end, it is all about uptime and efficiency, standing still is expensive for the customers. Instead of the driver noticing something wrong, often too late, then determining the root-cause and the wait for the parts to arrive can be costly. With predictive maintenance, based on knowledge derived from streaming information it could only be one stop, the fault can be identified earlier and parts can be delivered in time without having to be in stock for longer than necessary.

Barriers
There are multiple barriers to successfully implement Big Data analytics in the aftermarket/service market according to the interviewees. Better algorithms that can determine appropriate actions, as well as doing it automatically are currently missing. It should be possible to log more data, and doing so in more detail. There needs to be ID's on not only each vehicle, but at all important components in order to really make sense of root-causes. Furthermore, there needs to be better and more robust sensors, and the uptime of the sensor needs to be monitored as well to be able to detect false data recording. Today, as stated the drivers are the most critical sensor when it comes to avoid and detect faults.

Another barrier could be personal integrity, but it does seem to be a manageable problem as rules and regulations are becoming clear. According to one interviewee truck owners signs an umbrella agreement in which the owner of the vehicle has signed the right to collect data. It is the owners responsibility to inform the driver about these agreements, and the driver seldom has a choice but to accept since the agreement applies to many vehicle owners. This is central in the industry and there are no personal data concerning the driver is normally collected. There are laws that some data needs to be collected, and the vehicle owner obviously needs to follow this.

Data
The data itself and collection methods are central for R&D, Operations and Aftermarket, however to different degrees. Aftermarket is interested in data, but do not want to get too specific regarding how to collect it. For them it is more about application, and how they want the data presented.

For Operations the data is important, and much emphasis is put on traceability and volume of data. It is also of importance where it is gathered and what is needed to make sense of it in terms of connecting data throughout the lifecycle. Operations at the case company are often links between departments and since being a large corporation it is often hard to find existing information. There is also value for companies to collect information in the middle of the lifecycle so Operations plays a big part in terms of linking the company together (Shin et al., 2009, 1-3). Quality of data are important however Operations also want as much data as possible to make their own analyses. This is quite different from how R&D works. One reason for this might be that.

R&D is also the division that focuses the most on the collection itself, how to store and analyse data. It is R&D's task to make all things works, whereas Operations and Aftermarket is more application inclined. R&D stress that the ability to collect data is vital and that the data trend demands more from the data divisions, a fact seen in many businesses (Manyika et al., 2011). R&D are focusing not only on the quantity, but also on the quality of the data.

6 Analysis

It is of upmost importance to knowing what data is collected, the consistency and quality of the data. This is in line with the learning's from Wozniak et al. (2015)'s study, where investigation of the data is considered vital. Furthermore (Wozniak et al., 2015) says that it is important to know how the data has been collected and maintained to guarantee its quality and relevance. This view is also strengthened after conducted the data mining experiment in this project, where parameters and variables were unknown. This makes it hard to analyze results and find validity. Even if correlations is found it does not mean that they are relevant (Wozniak et al., 2015). This view is strongly agreed upon among R&D professionals and will be discussed further in Analysis.

In terms of storage, R&D argues that the development goes more towards a cloud solution, where data is vaulted and accessible in a virtual layer, but it is clear that much data logs are being stored locally on physical servers, e.g. in service workshops or when component quality is logged in production. Different types of data storage are expected to develop and R&D argues that the key will be to get several systems to work together. (Johanson et al., 2014) has a model where data first are stored based on geographic location, and then aggregated into a cloud. Solutions like this might be appropriate for the case companies as well.

6.1 Analytics

As the volume of data increases, so does the requirements on the tools and methods for analysis (McAfee & Brynjolfsson, 2012). There seem to be consensus between the interviewees regarding the need for automating processes for data monitoring and analysis. Specifically for Operations and Aftermarket, who's daily tasks consists of searching through product individual data. Aftermarket see the need arise for new processes with the increase of customers connected to their remote maintenance service. Similar needs reside in Operations where the interviewees see the investigative work as time-consuming and an ineffective way of working. For Operations, much efficiency could be accomplished by standardizing the data recording process. It would require less knowledge about the circumstances under which the data was recorded. Standardizing the recording would further increase the veracity of the data and decrease the risk of false correlations and interpretation (Demchenko et al., 2013; White, 2012, 4).

Interviewees from R&D also typically express concern regarding the risk of deducting false correlations and interpretations. Quality over quantity is preferred and considered more important. Meta data, regarding the data-sets, of interest for all interviewees are product configuration, climate where the product is used or supposed to be used and lastly, how the product is used. There is no need for bigger populaces but rather more dimensions that describes the circumstances and the life-cycle of the product.

Facilitating data-driven decision-making in the automotive business is a challenge due to the many dimensions required for obtaining high veracity data (White, 2012, 4). There is a need to fit the statistical model to the format of the data-set (Shin et al., 2009, 1-3), which was realized during the experiment and adopted for instance with the logistic regression. Furthermore, it is considered a harder task to interpret the results and elicit value from models with many dimensions which has also been observed by Demchenko et al. (2013). Some of the interviewees from R&D stress the need for the proper technical competence for interpreting and analysing results to elicit value. Knowledge-driven product development certainly require a high level of competence of the analyst as reported in related research (White, 2012, 4; Wozniak et al., 2015). However, there might be value in challenging pre-existing assumptions which was explored in the experiment presented in chapter 6.4.

6.2 Opportunities

All three division see a large potential with Big Data in automotive, as have experts within the industry before them (Johanson et al., 2014; Wozniak et al., 2015). R&D interviewees mention that the value of each product can increase enormously, since they could provide insight ito how products are used and what their status are in terms of quality. This development has been identified as positive and advantage in e.g. Johansson et al. (2014) and the prospect of knowledge-driven product development where the actual product guides decisions are exciting among the interviewees. Bringing data into the product development process is of focus for R&D, and the potential for shortening the feedback-loop for this shows great potential according to interviewees. Another opportunity is predictive maintenance, and this could bring insight and value to all three divisions.

The potential for aftermarket is imperative from a competitive point of view. It can lead to less stops for drivers, less components in storage and better knowledge about product and driver behaviour. In operations product quality can be predicted based on test-values in manufacturing as tested in this projects experiment, and the value and interest in this kind of tests are validated by interviewees. These kinds of services and solutions have previously been identified in Big Data and automotive (Johanson et al., 2014; Luckow et al., 2015) and there is consensus among the interviewees that this will with certainty increase. With more of the right data, the production and quality processes could be enhanced and improved (Zhang et al., 2016; Luckow et al., 2015). There will be flows of data with the launch of autonomous vehicles (Luckow et al., 2015). There are also large amounts of reports created during product development that may hold some opportunity in terms of data analytics (Arnarsson et al., 2016).

6.3 Barriers

The majority of the interviewees are concerned about future legislation regarding personal integrity. There is an upcoming EU-directive regarding data collection and personal integrity that will take effect in May of 2018 (Datainspektionen, 2017). There is general consensus that security with data collection is an aspect that definitely needs to be prioritized. It is mentioned that the legislation requires changes from a technological perspective as well. According to an interviewee in R&D, the new legislation will demand that users should be able to delete personal related datasets. This means that all data needs to be fully traceable, purposeful, structured and valid.

Shifting towards data-driven decision-making requires new and added competence in the organization according to (LaValle et al., 2011, 2; McAfee & Brynjolfsson, 2012). The interviewees are aware of this, Operations interviewees suggest new tools for facilitating the change. Some from R&D suggest automation of the analytics and make it centralized and cloud-based. This presents a major challenge but it is also an opportunity for third parties.

6.4 Concluding the Data Mining Experiment

As a part of this research several experiments where undertaken based on the available data from the production facility. This experiment was undertaken to fulfil the hypothesis stated by the research project an the potential has also been corroborated by previous researchers (Luckow et al., 2015; Zhang et al., 2016). The aim to find improvement potential in product development was aimed in quality case 1 of the experiment. In the experiment, it was realized that traceability would play a large role for the analytics as it was the main failing component, resulting in a partial analysis. However, the state of the data and other parameters might be the cause of not finding the product individuals from quality case 1. The quality case that generated result is however not irrelevant. Operations stated that it is of interest to pursue predictive analysis and machine learning methods for improving quality. Such pursuits should however be executed with a different approach, over a longer period of time than was possible in this pre-study. The recommended process to follow, if would have been to firstly, define the aim with the analysis, then ensure the data collection and recording method (and make sure that all that data is properly documented and understood), and lastly perform the statistical analysis (Shin et al., 2009, 1-3).

The logistic regression analysis was considered good and easy to understand. Studying the relationships between the variables in the plots, there seems to be a possibility to identify clusters. As the data-sets are classified after product family, perhaps the clusters are specific product types within the family. There has been no meta data available (at least to the researchers) in the database to make that distinction even though it seems likely. If there were a way to make the distinction between specific product types, it could lead to a different result.

The outlier analysis was considered even less fruitful. In hindsight, it seems likely that a nominal value has tolerance limits and is thereby already monitored. However, the analysis could point towards patterns in the data if a certain distance from the global mean was attained for a sequence of nominal values. This would mean that it is not a specific parameter that causes a function failing but rather a synergistic effect between a range of parameters. The sample sizes in the experiment showed however no such result.

7 Publications

In this section, the published reports and how the new knowledge is being used within the companies. The main achievement of this project is to have secured a continuation in a follow up project within the FFI program that started during Fall 2017 that is called Malekc (2017-03059). This was also one of the main considerations when starting a pre-study project: secure enough knowledge to determine whether or not future research may be fruitful.

7.1 Knowledge and Dissemination of Results

Planned dissemintation of Results	Mark With X	Comment
Increase knowledge within the area	X	Focus has been to increase knowledge about the potential.
Transfer to other advanced technical development projects.		
Transfer to product development projects.		
Market introduction		
Being used in policy development.		

7.2 Publications

We are currently working on a journal publication aimed towards the International Journal of Product Lifecycle Management. This publication aims towards being finalized during 2018. Insofar there is a master thesis published from the master program in Product Development at Chalmers.

Kristoffer Hertzman, Johannes Bladh, "Big Data and Product Lifecycle Management" Göteborg : Chalmers tekniska högskola, 2017.
http://studentarbeten.chalmers.se/publication/250443-big-data-and-product-lifecycle-management

8 Conclusions and Future Research

By carrying out practical tests on thin slices of Big Data, we have been able to clarify that existing production data could not be consistently be correlated, to market problems. The main issue is the availability of data that can be interpreted and correlated towards the problems. Meta Data is lacking (to us) and a new strategy involving the creation of more consistent data needs to be undertaken. This does not falsify the initial hope to build knowledge on data, but it may indicate that the analysis method used was not the right for an unstructured data set. The research will continue with the focus on building knowledge on more structured data such as written deviation reports. Interviews undertaken in the pre-study project further confirmed this potential and conclusion.

The project has (During Fall 2017) continued with a larger follow up project within FFI called MALEKC, where focus has been shifted to deviation reports in product development, but also including hads on engineering methods and validation techniques to validate the generating, and existing datasets.

9 Participants and Contacts

The project has been carried out in collaboration between Chamers, AB Volvo and Rejmes Transportfordon AB. The consortium made it possible to explore problems that origin in development or production, and which arise in the marketplace. The arrangement as such was valuable and gave many insights that could not be obtained without an extended value chain. Case studies regarding data collection of high value subsystems were made possible thanks to this approach.

References

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

Chen, H., Chiang, R. H. L. & Storey, V. C. (2012, December). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*, p.1165–1168.

Datainspektionen. (2017). Homepage. Retrieved May 3, 2017, from http://www.datainspektionen.se/lagar-och-regler/eus-dataskyddsreform/

Demchenko, Y., Grosso, P., de Laat, C. & Membrey, P. (2013, May). Addressing big data issues in scientific data infrastructure. *IEEE*.

Johanson, M., Belenki, S., Jalminger, J., Fant, M. & Gjertz, M. (2014, October). Big automotive data: Leveraging large volumes of data for knowledge-driven product development. *IEEE*, p.736–741.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S. & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management review*, *52*.

Luckow, A., Kennedy, K., Manhardt, F., Djerekarov, E., Vorster, B. & Apon, A. (2015, October). Automotive big data: Applications, workloads and infrastruc-tures. *IEEE*, p.1201–1210.

McAfee, A. & Brynjolfsson, E. (2012, October). Big data: The management revolu-tion. *Harvard Business Review*.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. (2011, May). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute, MGI*.

Shin, J.-H., Jun, H.-B., Kiritsis, D. & Xirouchakis, P. (2009, January). Function performance evaluation and its application for design modification based on product usage data. *International Journal of Product Lifecycle Management*, *4*, 84–113.

Terzi, S., Panetto, H., Morel, G. & Garetti, M. (2007). A holonic metamodel for product traceability in product lifecycle management. *International Journal of Product Lifecycle Management*, *2*.

White, M. (2012). Digital workspaces: Vision and reality. *Business information review*, *29*.

Wozniak, P., Valton, R. & Fjeld, M. (2015). Volvo single view of vehicle : Building a big data service from scratch in the automotive industry. *ACM*.

Zhang, D., Xu, B. & Wood, J. (2016). Predict failures in production lines: A two-stage approach with clustering and supervised learning. *IEEE*.