

Jukka Keinänen

Assessment of Product Data Quality

Metropolia University of Applied Sciences

Bachelor of Engineering

Degree Programme of Electrical and Automation Engineering

Thesis

26 March 2018

Tekijä Otsikko	Jukka Keinänen Assessment of Product Data Quality
Sivumäärä Aika	35 sivua + 1 liite 26.3.2018
Tutkinto	Insinööri (AMK)
Koulutusohjelma	Sähkö- ja automaatiotekniikka
Ohjaajat	Jani Johansson Markku Inkinen
<p>Tuotetiedalla tarkoitetaan tuotteen määrittelevää informaatiota, useimmiten tuoterakenteita, tuotteen valmistukseen ja kehittämiseen liittyvää dokumentaatiota sekä muuta tuotetietoa. Master datalla tarkoitetaan yrityksessä jaettavaa, liiketoiminnan kannalta kriittistä dataa. Tämä on useimmiten informaatiota tuotteista, toimittajista, asiakkaista, resursseista ja pääomasta. Datan laadulla tarkoitetaan sen kykyä palvella sen omaa käyttötarkoitusta.</p> <p>Työn tarkoituksena on luoda yleiskuva tuotetiedan laadusta eräässä keskisuudessa elektroniikkateollisuuden yrityksessä. Yrityksen tuotteiden elinkaaret ovat pitkiä, ja niiden tuotetietoa täytyy ylläpitää vuosikymmeniä. Datan laadulla voi olla hyötyjä tai seurauksia pitkäksi ajaksi. Työn tarkoituksena on myös tutkia korkean datan laadun tuomia etuja ja esittää suositus kuinka hallita datan laatua paremmin tulevaisuudessa. Datan laatua pyritään samalla arvioimaan ja mittaamaan.</p> <p>Yrityksellä on käytössään raporttityökalu jolla voidaan tunnistaa poikkeamia datassa. Kriittisimmät tulokset löytyvät keskeneräisistä nimikkeistä, myytävistä nimikkeistä sekä tuottavuusmittareihin ja kustannuksiin vaikuttavista nimikeparametreista. Tulokset eivät ole aina yksiselitteisiä ja työssä on tulkittu niitä. Lisäksi tuotteen yhteensopivuutta RoHS- ja REACH-lainsäädännön kanssa on vaikea todeta ilman täydellistä valmistaja- ja komponenttietoa. Työn aikana tehtiin haastattelukierros jossa keskusteltiin tuotetiedan yleiskunnosta, käytettävyydestä, luotettavuudesta, tuotetiedan hallinnasta ja sen vaikutuksesta yrityksen liiketoimintaan.</p> <p>Yhteenvetona voidaan todeta, että tuotetiedan laatu on yrityksessä tyydyttävällä tasolla ja sitä voidaan käyttää yleisimpiin käyttötarkoituksiin. Yritys pitää yllä huomattavaa määrää tuotetietoa, ja tämä itsessään on yksi isoimmista haasteista. Eri sidosryhmät voivat osoittaa poikkeamia datassa. Tuotetiedan laatu ei ole täydellistä, mutta liiketoimintaa voidaan harjoittaa sujuvasti siitä huolimatta. Yritys on kehittänyt prosessejaan ja datakäytäntöjään vuosien varrella, ja etenkin uusissa tuotteissa datan laatu on melkein moitteetonta.</p>	
Keywords	master data, product data, data quality, Product Data Management, Product Lifecycle Management

Author Title	Jukka Keinänen Assessment of Product Data Quality
Number of Pages Date	35 pages + 1 appendix 26 March 2018
Degree	Bachelor of Engineering
Degree Programme	Electrical and Automation Engineering
Instructors	Jani Johansson Markku Inkinen
<p>Product data is defined as information describing and defining a product, typically being product structures, documentation related to manufacturing and developing the product and other product related information. Master data refers to the characteristics of critical business data within an organization, typically being information related to materials, product structures, suppliers, customers, resources and assets. Data quality is defined as its ability to serve its intended purpose.</p> <p>This study aims to create a summary of product data's quality in a certain middle-sized enterprise in electronic industry. The lifecycles of the company's products are long, and their product data has to be sustained for several decades. Therefore data quality can have benefits or consequences for a long time. The study also aims to investigate benefits of higher data quality and present a recommendation of how to improve data quality in the future. New ways of measuring and assessing data quality are also considered.</p> <p>The company utilizes a report tool that can be used to discover defects in data. The results include mainly defects in incomplete items, active sales items and parameters affecting productivity metrics and end product's expenses. Results of the report are not straightforward, and some interpretation is presented in this study. In addition to this, completely verifying product's compliancy with RoHS- and REACH-legislation is difficult without complete manufacturer and component data. A series of interviews was made to assess product data's general condition, usability, reliability, current management of product data and its impact to the company's business.</p> <p>To summarize, the company has data quality on an adequate level and it can be used for common business purposes. The company maintains a considerable amount of product data, and this itself is one of the biggest challenges. Different stakeholders can point out defects in data. Product data's quality is not perfect, but business can be considered to be fluent even so. The company has developed its processes and data policies over the years, and especially new products have almost blameless product data.</p>	
Keywords	master data, product data, data quality, Product Data Management, Product Lifecycle Management

Contents

1	Introduction	1
2	Definitions	2
2.1	Master Data	2
2.2	Transaction Data	3
2.3	Metadata	3
2.4	Data Quality	3
2.5	Data Owner	4
2.6	Data Cleansing	4
2.7	Product Data	4
2.8	PDM (Product Data Management)	5
2.9	PLM (Product Lifecycle Management)	6
2.10	ERP (Enterprise Resource Planning)	7
2.11	Item	7
2.12	Product Structure	8
3	Systems of Product Data	8
3.1	Scope of Product Data	8
3.2	Systems Using Data	9
3.2.1	PDM System	9
3.2.2	ERP System	10
3.2.3	Relationship Between PDM and ERP Systems	11
4	Current Condition of Product Data	12
4.1	Item Data in the PDM System	12
4.1.1	Service Items	12
4.1.2	Current Flag	13
4.1.3	Lifecycle Statuses	14
4.2	Item Data in the ERP System	15
4.2.1	Incomplete Items	16
4.2.2	Sales Item Parameters	17
4.2.3	Parameters Affecting Productivity Metrics	18
4.2.4	Parameters Affecting Product's Expenses	19
4.2.5	Results of the Report	20
4.3	RoHS Directive and Compliancy Data	21
4.4	Interviews	23

	Abstract	
4.4.1	Condition of Product Data	23
4.4.2	Business Impact	24
4.4.3	Root Causes	25
4.4.4	Solutions	26
4.5	Summary	27
5	Conclusions	28
5.1	Summary	28
5.2	Benefits of Data Quality	29
5.3	Solutions	31
	References	34
	Appendices	
	Appendix 1. Questionnaire	

1 Introduction

In PLM (Product Lifecycle Management) context, the term product data is widely understood to cover all product related information (Kropsu-Vehkaperä, Haapasalo 2011). Product data defines physical and functional attributes of a product, including detailed technical information, but also abstract and conceptual information related to product's nature (Sääksvuori, Immonen 2002). Typically, this consists of product structure and documentation related to manufacturing and developing the product. Product data defines and describes a product, being a major asset and a strategic resource in any company's business (Stark, 2015). Product Data Management assures consistency in data between all systems associated with it.

This study is made for a certain middle-sized enterprise in electronic industry. In this study, the company is not mentioned with its real name, but is referenced with 'the company', when necessary. The company has a considerable amount of products and product data associated with them. Management of this data is starting to prove challenging. Although company's business functions are fluent and product data supports it, problems with data quality sometimes hinder it. The company offers highly configurable products, and customizability of products requires high quality product data to support this.

New products and designs increase the amount of items and product data associated with it. Increasing amount of data calls for efficient methods to verify and manage data quality better than before. The question is how to verify so that item data is always complete, valid, accurate and up-to-date. In the company, data quality on average is in good condition and most often usable to conduct business. Inconsistencies in data are also discovered every now and then. Another challenge lies in measuring data quality, identifying incorrect data and distinguishing it from valid data.

Like in all manufacturing industry, raw materials, parts, components and sub-assemblies used to manufacture end products are managed with the concept of itemization. Data and documentation associated with the items needs precise supervision and management, or otherwise data becomes obsolete over time. Most often, end product endures a considerable amount of changes during its lifecycle, and product data has to conform

these changes as well and as fluently as possible. At the moment, the company maintains approximately 46,000 different items in its PDM (Product Data Management) system.

In order to improve item data quality and achieve some notable results, the whole organization has to commit and contribute to managing data better. This requires resources and therefore is one of the biggest challenges in improving data quality. Data quality is a common problem, because big enterprises often have many products and much product data. Naturally this creates a demand for good and efficient data management.

Improving product data quality can improve efficiency in operations and manufacturing. Well managed and reliable product data is a valuable asset. Data of a high quality can prove to be a remarkable competitive advantage for any company. Data is part of the product, and the quality of product data should be taken with the same level of seriousness as the quality of the product itself. Data management shouldn't be considered as just another technical liability or additional expense, but rather as a process that produces information products and creates additional business value. Improving the credibility and comprehension of data can also be a key factor in unlocking its full potential and discovering new ways to benefit from it.

Because measuring data quality is difficult at the moment, this study aims to create a summary of product data quality and its fitness for use in the company. Also, this study aims to investigate benefits of higher data quality and present a recommendation of how to improve data quality in the future. At the same time, new ways of measuring and assessing data quality are considered.

2 Definitions

2.1 Master Data

One common way to describe master data is that it refers to the characteristics of critical business data within an organization (Loshin 2010). This can refer to very different kind of data in different enterprises and environments. Typically master data is information related to materials, product structures, suppliers, customers, resources and assets. Ideally, master data is supposed to provide one true source of data and act as a common

point of reference when data is distributed to several systems (Ballard 2013). Master data describes the business-oriented aspects that are used in diverse applications across organizations (Kropsu-Vehkapera, Haapasalo 2011).

2.2 Transaction Data

Transaction data describes relevant events in a company, e.g. orders, invoices, payments, deliveries, storage records, etc. Transaction data uses master data, and therefore transaction data doesn't fulfill its purpose if master data is not correct. (Haug and Arlbjørn 2011).

2.3 Metadata

Metadata is data about data. In other words, information about the author, format or date of data itself (Sääksvuori, Immonen 2002).

2.4 Data Quality

Data quality is a concept with many definitions. Most often, data has a very distinct purpose in different environments and thus the definition of data quality varies.

Tozzi (2017) defines data quality as its ability to serve its intended purpose. This is a simple definition used in many contexts, and means if data can be used for its intended purpose, it has good quality. This isn't a very specific definition, but more of a general way to assess data quality that can be applied in varying circumstances.

Another way to define data quality is to present it as a multidimensional concept. These dimensions have many definitions, the most common being completeness, consistency, validity and credibility. However, the definitions of all data quality dimensions are not generally agreed upon, and they might have different names or meanings in different contexts. For example, Wang (1998) has defined 15 different data quality dimensions related to information products in his research. According to Silvola et al. (2016), the four most common ones presented in the literature seem to be accuracy, consistency, completeness and timeliness. When data quality is defined in specific dimensions, it allows

data quality to be evaluated more precisely. By observing the dimensions separately, data quality can be assessed more specifically or even measured. When measuring completeness of data for example, if one product is missing price data among 99 other products, then price data is 99% completed. Other dimensions are usually much harder to observe and measure, but completeness is one of the simplest data quality dimensions to assess.

In this study, one or both definitions of data quality are used according to circumstance.

2.5 Data Owner

Data owner is a person or a unit who has the responsibility over the data set (Lucas, 2010).

2.6 Data Cleansing

Data cleansing, also called data cleaning or data scrubbing, deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data (Rahm, Do 2000). According to Rahm and Do (2000), often a significant portion of the cleansing and transformation work has to be done manually, or by low-level programs difficult to maintain.

2.7 Product Data

In PLM context, the term product data is widely understood to cover all product related information (Kropsu-Vehkaperä, Haapasalo 2011). Product data can be divided into product's master data (such as product descriptions, item codes or price) and other general product data (Kropsu-Vehkaperä, Haapasalo 2011).

Product data defines physical and functional attributes of a product, including detailed technical information, but also abstract and conceptual information related to product's nature (Sääksvuori, Immonen 2002). According to Sääksvuori and Immonen (2002), product data can be roughly divided in three categories: specification data, lifecycle data and metadata.

Typically, product data includes product structure and documentation related to manufacturing and developing the product. Product data defines and describes the product, being a major asset and a strategic resource in any company's business (Stark, 2015).

2.8 PDM (Product Data Management)

PDM is a systematic, controlled methodology to manage and develop a product through its lifecycle (Sääksvuori, Immonen 2002).

PDM is also a system used to support and maintain product data. Figure 1 presents an example screenshot of an item card in a PDM system. The PDM application keeps product data as a strategic resource under control, making it available, whenever it's needed, wherever it's needed, by whoever needs it (Stark, 2015). Ideally, the system provides an unified way for the entire enterprise to manage its product data. Most of the PDM system's core functions are based on items and product structures. According to Sääksvuori and Immonen (2002), the most important features of the PDM system are:

- management of items and product structures
- documentation management
- product change management
- search and report functions of data.

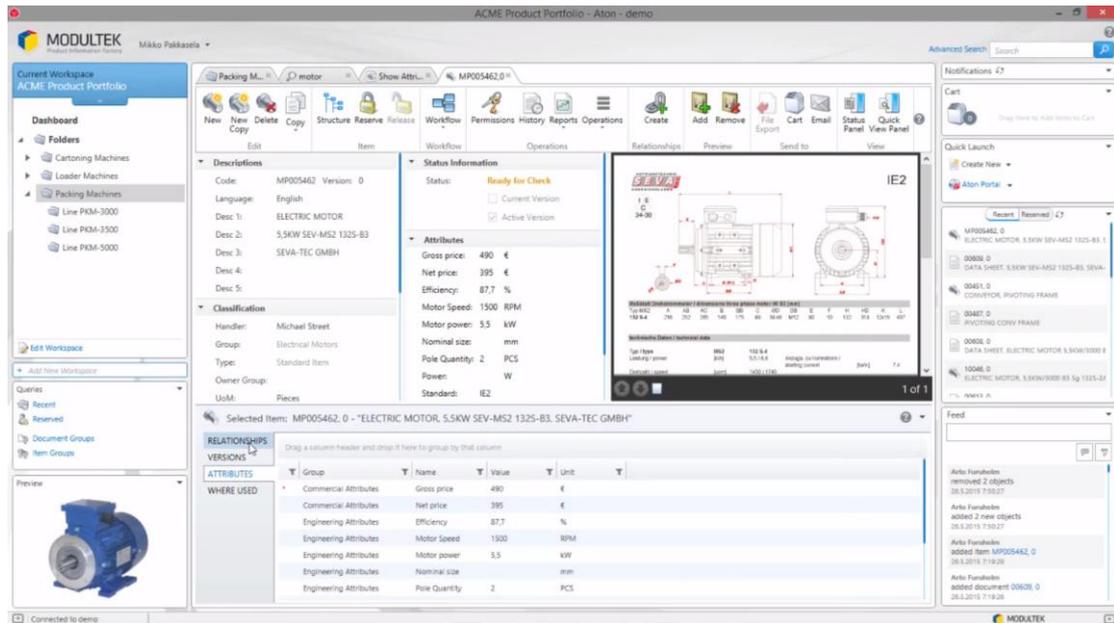


Figure 1. Example of an item card in a PDM system.

2.9 PLM (Product Lifecycle Management)

PLM refers to a larger framework of product data management, and especially the lifecycle perspective of data management (Sääksvuori, Immonen 2002). It is a group of systems and methods that enables development, manufacturing and management of a product through its entire lifecycle. Figure 2 presents 5 phases of product's lifecycle in PLM context (Stark, 2015):

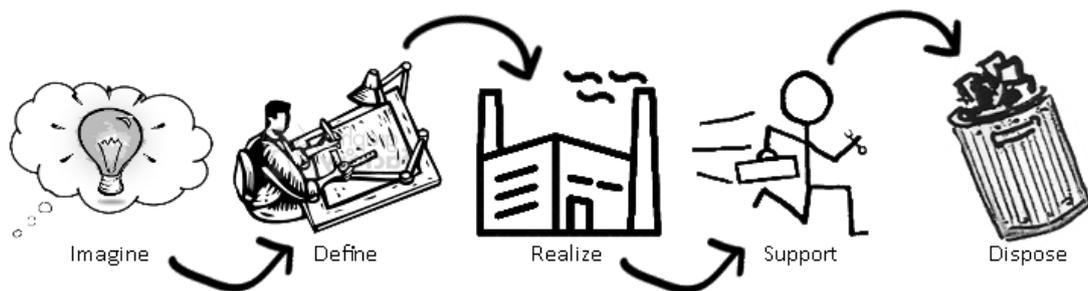


Figure 2. Product's lifecycle (Stark, 2015).

According to Stark, Imagine phase defines product requirements and its main functional aspects. Define is detailed design and development of the product, prototype testing and eventually, release of product. Realize refers to selling, manufacturing and delivering

products. Support refers to maintenance of installed base products. Dispose refers to product's ramp down, for example disposal of material objects or archiving product information.

According to Stark, PLM is a business activity of managing, in the most effective way, company's products all the way across their lifecycles; from the very first idea of a product all the way through until it is retired and disposed of. PLM is carried out to meet business objectives of increasing product revenues, reducing product-related costs, maximizing the value of the product portfolio, and maximizing the value of current and future products for both customers and shareholders.

CIMdata (2018) defines PLM as a strategic business approach, applying a consistent set of business solutions that support the collaborative creation, management, dissemination, and use of product definition information. PLM supports the extended enterprise (customers, design and supply partners, etc.) and integrates people, processes, business systems, and information.

2.10 ERP (Enterprise Resource Planning)

ERP system is a large, integrated information system with multiple functions and acts as a core system for company's business (Sääksvuori, Immonen 2002). It combines engineering, production, sales, marketing, financing and human resources to a single entity. Typically, the system directs the company's flow of resources, materials, money and information.

2.11 Item

Item is a standard way to identify, code and name a physical product, part of a product, component, material or service (Sääksvuori, Immonen 2002). Items can also be various other things, like documents, tools or software.

2.12 Product Structure

Product structure is a model that presents product as a hierarchy of items (Sääksvuori, Immonen 2002). It is also known as item structure or BOM (Bill of Materials).

Product structure represents the product, information linked to it, and the relationships between its components (Kropsu-Vehkaperä, Haapasalo 2011). Product structure can be used to standardize the understanding over a product for the entire company (Kropsu-Vehkaperä, Haapasalo 2011). Figure 3 presents an example of a product structure.

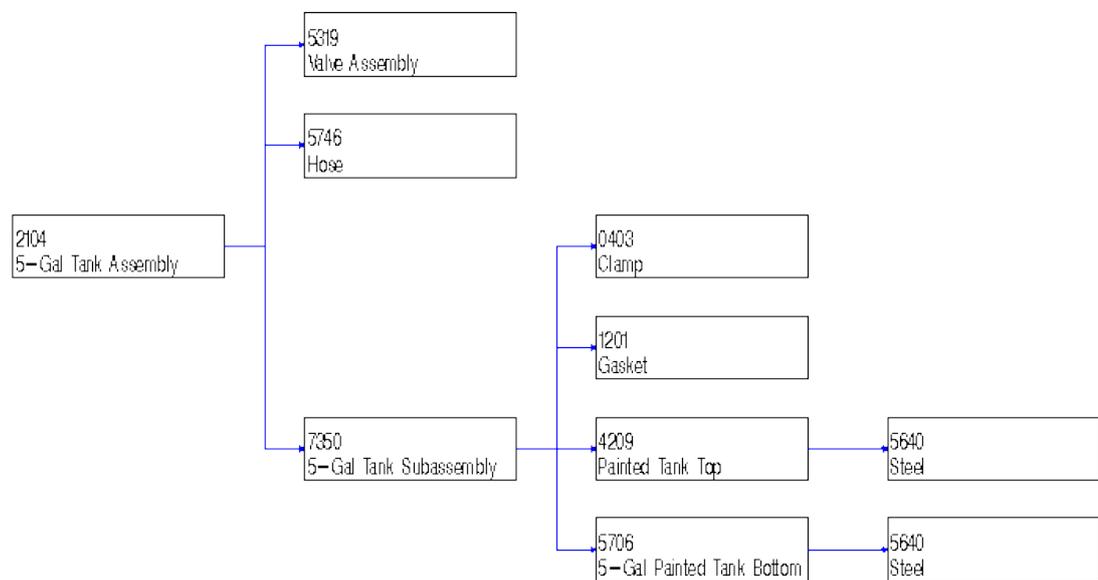


Figure 3. Example of a product structure.

3 Systems of Product Data

3.1 Scope of Product Data

Product data is a large concept, and it is imperative to define data that is in the scope of this study. Focus is in the master data that defines a product in the PDM and ERP systems. This includes items, products structures and data attached to them. The PDM system was chosen because it is the main system used to manage the company's master

data related to products. Some data is partly transferred to the ERP system. This system is critical for business activities and it is worth inspecting data quality in this system also.

3.2 Systems Using Data

3.2.1 PDM System

The PDM system used in the company is the master system for item and product structure information. When new items are created, the basic characteristics and master data is first defined and then maintained in the PDM system. If item data is required in some other system, a link between systems is created for the item and data is then transferred from this system. This is important to remember when item data is changed. When data is altered, it might be necessary to transfer the items to other systems again. Otherwise there will be inconsistencies between systems.

The PDM system is also the master system for product and manufacturing documentation, document codes, revisions and statuses. In fact, documents, drawings and instructions are items on their own, with relationships to actual document files. In the company, the PDM system is used to maintain these basic characteristics of an item:

- item code
- structure
- lifecycle status
- description
- revisions
- grouping
- product family
- technical attributes

- all documentation, drawings and instructions required to manufacture the item.

Table 1 presents an example item in the PDM system:

Table 1. Example of an capacitor item in the PDM system.

Item code	111128
Description	Cap, Ceramic 22uF 6.3V
Structure	Consists of items: 222256 ABC100
Parent items	Used in items: 333512 EFG200
PDM status	Accepted
Product Family	Electrical components
Group	1050010 Capacitors, Ceramic
Revision	A
Documentation	Data sheet

3.2.2 ERP System

The ERP system is used for many purposes in the company's operative activities, including inventory management, handling of purchase and sales orders and production planning. This can be comprehended as transaction data that uses master data as a foundation. In the company, the ERP system is critical for production and data quality problems can even cause production to halt temporarily. Items are first created in the PDM system and then transferred to the ERP system. After that, item is reinforced with setups and data related to transaction data and inventory data. Table 2 presents previously mentioned capacitor item in ERP view:

Table 2. Example of an capacitor item in the ERP system.

Item code	111128
ERP status	Active Component
Owning organization	Helsinki
Serial controlled	No
Make/Buy	Buy
Buyer	John H.
Warehouse locator	INS/A6-3F
On Hand Quantity	1024 pcs in Helsinki 64 pcs in London
Product Manager	Steve D.
Delivery Time	5 days
Routing	None

3.2.3 Relationship Between PDM and ERP Systems

New products and structures are always first created in the PDM system. If item is used in production or it is required in sales configurator, it is transferred to ERP system. Not all items are required to exist in ERP system. PDM system contains many immaterial services and documents that are items in a same way as physical materials used in production. Product structures in PDM system contain also component information from items purchased from a supplier, but these items are not required in the client's ERP system.

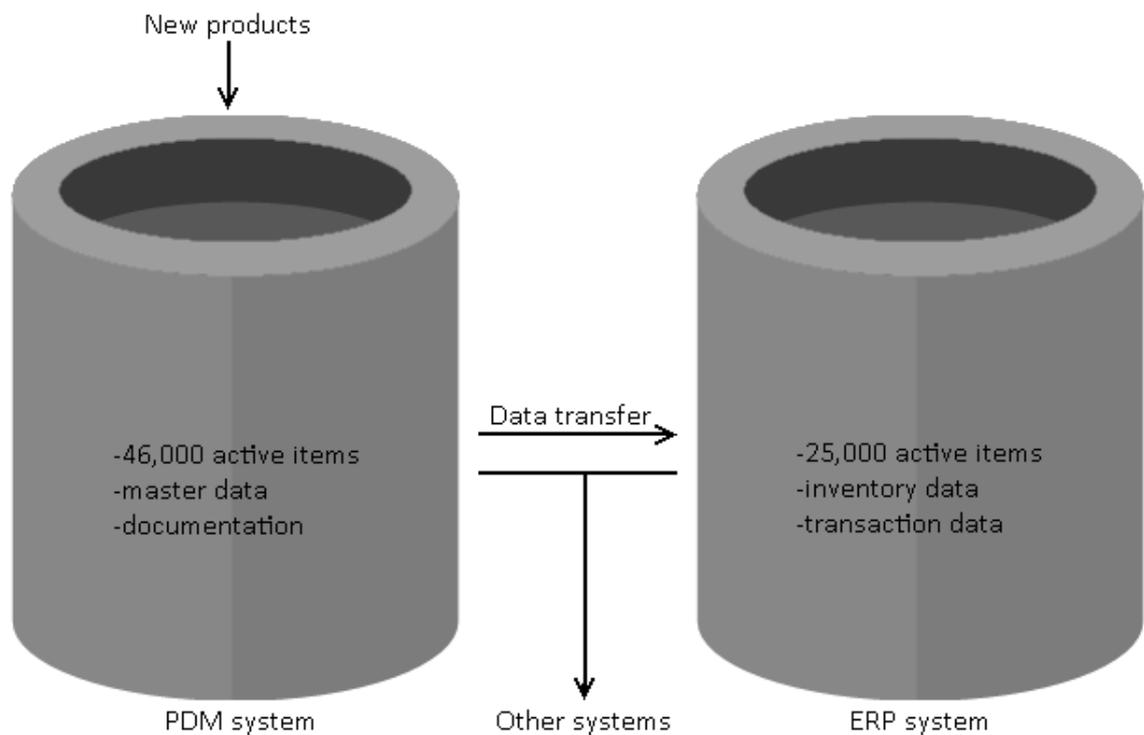


Figure 4. Data flow between systems.

Although the two systems are used for completely different purposes, they use the same item data as a foundation to function, as presented in figure 4. The two systems use a different database with similar product data. Naturally, this creates a possibility that there are inconsistencies between systems. Roughly speaking, the PDM system has more producers of data, while the ERP system has more consumers of data.

Sometimes data is also transferred to other systems, including a compliancy system for environmental legislations, Business Intelligence tools and the company's web store.

Data is not erased from the PDM or ERP systems, even when a product reaches its end of life activities. It is still contained in the systems, but its status becomes inactive.

4 Current Condition of Product Data

At the moment, product data quality can't be evaluated thoroughly in the company. Data quality is difficult to present with numeric values. There are some ways to measure data quality, but they are limited and one has to know how to interpret the results. This study aims to provide an overview of product data's quality and usability with observations, interviews and report tools at hand.

4.1 Item Data in the PDM System

As defined in chapter 2, item is a standard way to identify, code and name a physical product, part of a product, component, material or service (Sääksvuori, Immonen 2002). Conventional user inspects and modifies items one by one with item cards. Item card presents data attached to the item, like description, item code, lifecycle status and relationships to files, structures and other items. In the company, items are mainly used to represent components, assemblies, features of a product, end products, services, software, documentation and instructions. Main observations are related to lifecycle statuses, descriptions and acceptances of items.

4.1.1 Service Items

In the company, services and spare parts are sometimes detached from the main product and don't have any relationships to the them. These kind of items also have weak descriptions like "travel expenses" or "support plan", and usually there's no documentation. One has to just know and remember rest of the item's purpose and content. This is an alarming phenomenon from many reasons. It is difficult to move item's ownership to another person. There is a risk that no one owns the data, or that it is never deactivated. This can lead to increasing number of active items and more data to sustain. Floating items without connections to any product structure present a risk they can't be managed properly.

This could be solved by connecting all separate service items and spare parts to product's main structure. Another way would be to create a product structure beside the main product structure to contain all service items related to it. Data ownership would then be easier to transfer as a single entity. Product's ramp down would also be more straightforward, if it would be enough to deactivate one or two item structures. These kind of structures have already been made for spare parts, but not all spare parts in the system are connected to them.

4.1.2 Current Flag

In a single item, one of the revisions can be activated as "current" revision in the PDM system. Ideally, the latest revision used in production is meant to be marked as current revision. The system has a feature that automatically decides the current revision, based on the revision that has been accepted latest. When modifying old revisions, user has to change old revision to "Modification" status and then change it to "Accepted" because of status rotation. This process always changes the modified revision to current, and should be corrected by accepting the latest revision again. Because of this feature in the PDM system, it is very easy to accidentally change current revision when making modifications in old revisions. As a result, current revision has become an unreliable attribute in the system and revision marked as current revision might not be the current. At the moment, it is not advisable to use the information. This doesn't have damaging effects for the company's activities, but this data has to be interpreted. Table 3 presents an example of an item ABC100 which should have revision C marked as current. In most occasions, revision B shouldn't be in "Accepted" status. When user finds items from the system, first step is to choose the desired revision from the search results. Choosing correctly the current revision might be confusing.

Table 3. Example of an item with incorrect current revision.

Item code	Revision	Current revision	Status
ABC100	C	No	Accepted
ABC100	B	Yes	Accepted
ABC100	A	No	History

4.1.3 Lifecycle Statuses

The PDM system has many items left to “Main Information Created” or “Checked” statuses. These statuses are only used for items with work in progress, and they shouldn’t be left for this status for a long time. A simple query from the PDM system returned 5,862 item revisions with “Checked” status. 4,427 of these (75.5%) have been created over a year ago. Same search for item revisions with “Main Information Created” status returned 16,608 results, with 13,899 (83.7%) of these created over a year ago. There’s probably many old component items staying active in the system for no reason. If these items aren’t used, they should be moved to “History” status in order to make the sustainment of product data more efficient.

When inspecting documents, the results are somewhat similar. The company uses 67 different groups for documents, the most common ones being drawing groups, 3D-models, datasheets, engineering changes, instructions and other technical documentation.

Table 4. The most common document groups used in the company and their statuses.

Group	Total	Accepted	Main Information Created	History	Other statuses
SolidWorks models	44,402	38.0%	45.3%	11.5%	5.2%
SolidWorks drawings	36,627	40.3%	36.1%	15.0%	8.6%
Mechanical drawings	29,029	76.8%	8.3%	11.3%	3.6%
System drawings	24,641	43.6%	49.3%	4.6%	2.5%
Instructions	20,478	44.8%	22.6%	27.6%	5.0%
Engineering changes	19,151	82.3%	10.9%	0.8%	6.0%
Data sheets	17,085	51.4%	31.7%	2.4%	14.5%
MDFs	13,959	38.4%	55.2%	2.7%	3.7%
Electronic drawings	9,083	62.9%	19.4%	12.6%	5.1%
User manuals	6,618	50.1%	18.2%	28.1%	3.6%
Technical documentation	5,157	15.0%	14.8%	68.1%	2.1%
Other drawings	4,941	52.1%	18.5%	20.2%	9.2%
PCB documents	4,828	56.9%	26.4%	10.2%	6.5%
Other groups	32,834	40.9%	33.9%	5.5%	19.7%
All groups	268,833	49.2%	31.5%	11.7%	7.6%

Table 4 presents the most common document groups used in the company and their statuses. At the moment, there’s a total of 84,766 documents with “Main Information Created” status. This is an alarming number when compared to “Accepted” documents, which is 132,291. The purpose of status is to describe item’s or document’s stage of lifecycle, and at the moment this data doesn’t do that. There’s no way to reliably separate

finished documents from unfinished. This causes confusion and additional work, for example when data is shared with suppliers and one has to investigate if the document is finished. This could be avoided, if document's producers would take responsibility in bringing all finished documents to "Accepted" status.

When new revision is created for an item, the status of old revisions is usually changed to "History". New item revision has to be updated to all product structures using the item, or otherwise product structure contains old items with possibly expired documentation. In the company, some product structures haven't been updated and still contain old item revisions with "History" status. This makes it difficult to share up-to-date data with suppliers, because the system picks expired documents from the product structure. This could be avoided by updating the revision in all product structures when the revision change occurs.

The PDM system also contains "Accepted" items that are not in active use any more. It is difficult to present statistics of these, because item's activity is difficult to state automatically. The activity of items is constantly questioned, and the group maintaining product data has to respond to these inquiries now and then. This produces additional work that could be avoided if items were moved to "History" status properly.

4.2 Item Data in the ERP System

Measuring data quality is a difficult task. It is hard to unambiguously distinguish correct data from incorrect, or expired from up-to-date. If data quality can't be measured, it is difficult to evaluate data's usability or properly intervene in data quality.

One of the few ways to measure data quality directly is to examine results of a report tool used in the company. The report is meant to be used as an inspection tool when creating setups for new items in the ERP system, but it can also be used to run the entire collection of items in the system. This tool evaluates item's nature and completeness of data associated with it. Results can be exported to Excel-file, where identified defects are highlighted. The results are based on simple rules, like that certain type of items need to have certain parameters. This provides some kind of overview of the completeness of item data in the entire system. In a way, the results of this report can be considered as a completeness metric of the company's ERP data.

Purpose of data can be very complicated, and therefore it is not possible to build a perfect metric. Naturally, the results need some interpreting, because rules and use cases have exceptions. While interpreting results, one has to consider carefully what is truly a defect in data quality and what isn't case-by-case. For example, the results might contain a considerable quantity of items that have been archived and are not maintained any longer.

In the company's ERP system, a single item can contain over 140 active parameters. All of these can't be monitored, and the report inspects 39 different parameters critical to business. The data quality of these parameters can't be inspected perfectly, but most rules check that the parameter has at least been defined for necessary type of items. Most rules don't validate the accuracy of data, because verifying this automatically is very difficult. The report was run from the entire ERP system, containing 32,950 items. This also includes items with inactive status. 17,818 items had at least one defect, with a total of 49,199 defects. A summary of the results is presented in some detail in following subsections.

4.2.1 Incomplete Items

ERP system has some items with incomplete setups. When items are transferred from PDM system to ERP system, item always starts with an empty base type. This can be identified by inspecting item's type. In practice, empty type means that item data is incomplete and missing all item setups.

Most of these are items that have been accidentally transferred to ERP system, and shouldn't be there in the first place. Since items can't be erased from the system, these items are just hanging with empty item setups. A new item type could be created for these kind of items to indicate they are not used. Some of these items are brand new, and item setup is still work in progress. Items are transferred to ERP on a daily basis, and therefore it is normal that there is a constant number of new items with this type. In this defect, it is best to not inspect items that have been transferred recently. This check returned 46 items, which is not an alarming number and doesn't damage the company's business.

4.2.2 Sales Item Parameters

Active sales items are specified to have certain parameters. Table 5 presents identified defects related to active sales items.

Table 5. Identified defects in active sales items.

Parameter	Report tool's defects	Defects in sales items
Country of origin	4,514	369
Customs tariff code	12,828	180
EU ECCN	10	10
US ECCN	49	49
Product manager	1,016	1,016
US tax classification	2	2
Weight	14,592	2,259

Tariff codes, country of origin and weight are reported to customs when items are shipped over country's borders. During this study, the report's scope was modified to check items with all statuses. Naturally, the existing data doesn't immediately conform to this kind of sudden change in specifications. Even though there's plenty of errors in the report, only active sales items need this data in practice. For this reason the measurement might give a wrong impression. The company doesn't really have thousands of problem causing items. If everything else than active sales items are filtered out of the results, only 369 items are missing country of origin information and 180 items are missing tariff codes. A small number of these are intangible items with incorrect item type, and shouldn't be considered to have this data. Rest of the defects are small in numbers, and product manager information is not business critical.

However, this data is very business critical. Sales items missing tariff codes can halt the delivery when the goods are handed over for a forwarding company, or in customs. This can affect the product's on-time-delivery. If a product has to be delivered right away, the dispatch department can add the tariff code manually for the bill of lading. This is additional work, that could be avoided if item data would include the tariff code. Tariff code also has effect on how much tax is payed when item is shipped over country's borders. This is a good example of master data that has a direct impact in the company's expenses. Improvements in data quality could possibly yield some savings.

Export Control Classification Number (ECCN) is a five character alpha-numeric designations used to identify dual-use items for export control purposes. Defects are missing values in sales items.

All tangible items also need to have a weight and a unit of measure. Tangible items can be identified with item's type, so certain service items, contract items and virtual options do not need this parameter. Weight rule is clearly the most common error of the report. This check returned 14,592 items violating this rule. The scope of the report for this parameter was also changed during this study to contain items with all ERP statuses. If this report would include only items sellable to the customer, the number of defects would be 2,259. At the moment, weight data is of most use to these items only. This is an example of the need of interpretation when reviewing the results. Unit of measure check returned 0 defects.

Some kind of weight could be automatically calculated by adding its substructure's item's weights together. This requires that the weight has been determined in all items in the substructure, or otherwise the result remains partial. The same method could be applied to verify accuracy of weight data, by comparing substructure's added weight to its parent item's weight.

Although weight is specified as a needed parameter, this data is utilized minimally in daily activities. At the moment, weight data isn't complete enough to be utilized and weight has to be measured case-by-case for all deliveries before they are shipped. This data is also proved to contain accuracy problems, because values in the system are not always corresponding to the reality. Higher data quality could be used to reduce time used in the company's logistics, if shipments wouldn't have to be measured on every delivery.

4.2.3 Parameters Affecting Productivity Metrics

The company uses productivity metrics to evaluate how efficient its factories have been. The results affect yearly bonuses of the factory's employees. Productivity is calculated with product data, as presented in equation 1.

$$productivity = \frac{\textit{employee routing time of completed products}}{\textit{registered working hours}} \quad (1)$$

Item's routing defines how much time product's manufacturing should take. This time is compared to the registered working hours in the factory. Productivity is calculated as

division of these values. Table 6 presents some identified defects in parameters affecting productivity metrics.

Table 6. Identified defects affecting productivity metrics.

Parameter	Report tool's defects
Routing is missing in a manufactured item	32
Routing exists, but completion locator is missing	1
Routing exists, but routing lead time is missing	19
Routing exists, but preprocessing lead time is missing	23

These results have been added together from the company's all manufacturing organizations. The productivity metrics are not distorted, because items can't be manufactured without a routing. These items probably have a wrong item type, and they shouldn't be of manufactured type. According to the report, 32 manufactured items are missing a routing. In addition to missing routings, some of them are incomplete and missing lead times. To put the problem to the right scale, the company has 3511 manufactured items when all manufacturing organizations are added together.

The report doesn't inspect the accuracy of routings, because it isn't possible to verify this automatically. In order to achieve true results in productivity metrics, lead times defined in routings have to be as accurate as possible. This is especially important in high volume products, because even the slightest imprecisions affect the entire mainstream.

4.2.4 Parameters Affecting Product's Expenses

Lead times defined in routings also have significant impact in product's calculated expenses. Routing defines how much expenses should be added from the working hours used in manufacturing. Accuracy of these cannot be verified with this report. Also, configuration of products isn't considered in routings. Some variations of product consume more time to manufacture than others. However, this has been taken into account in product's pricing, because different features have separate costs.

Item also has a parameter used to transfer its expenses to the upper level in the item structure. In purchased items this flag should be off, or otherwise item's expenses are passed on to its parent item's expenses. In manufactured items this flag should be on, or otherwise item structure doesn't generate expenses for the end product. This parameter is checked for following errors:

- parameter is “off” in items where it is needed (all manufactured items and models)
- parameter is set “on” in items where it shouldn't be (purchased items)

Item's type is used to determine whether it is manufactured or purchased item. Defects in this data lead to distortion in end product's expenses. This check returned 65 items violating this rule. Some of these defects were corrected during this study, and the results are quite good anyway.

4.2.5 Results of the Report

The report also inspects other errors related to routings, distribution of items, serial control rules, installed base data, planning methods and consistency of item types. The results are usually small numbers, most of them being less than 50. Some of these defects were completely cleansed during this study. However, more of these defects have also emerged. Some errors are generated from items with empty setups and not used in practice.

The measurement might give a wrong impression and the results need interpreting. Even though there's defects in over 17,000 items, there's often a sensible explanation for them. Typical explanations for the defects are minor omissions of item specifications, exceptions in specifications or just brand new items that still have work in progress. The results would be more truthful, if the report would scope out items with incomplete setups. On the other hand, the system shouldn't contain these kind of items at all. Even better solution would be to clean incomplete items out of the system. Some checks are made for history items, even if the data has no use anymore. If inactive data would have been filtered out of scope, the results would also be somewhat smaller.

Critical defects in item setups cause immediate problems in manufacturing or planning. Usually when item setup is incorrect, feedback is immediate and the problem is solved right away. It is difficult to develop a perfect metric for data quality, but this report tool gives a basic overview of item data's completeness in the ERP system. Most of the parameters have less than 50 discovered defects in a database of over 30,000 items. This is a quite good result. However, it is likely that there are also some problems in data quality that have not been identified and measured yet.

It is also worth noting that the report mainly inspects only completeness of data, but not its accuracy. Data's accuracy refers to how accurately data corresponds to the real thing it represents (Batini, Scannapieca 2006). Measuring accuracy of data could produce useful results, but doing this automatically is challenging.

During this study was discovered, that two instructions were maintained for setting up new items in the ERP system. Other version was officially accepted instruction, and unofficial one was developed and updated from that. Comparing them revealed that the instructions were mostly identical, but not completely. In practice, this instruction defines how item parameters measured in the report tool should be set up in the ERP system. Ideally, there should be only one universal instruction for producing this data.

4.3 RoHS Directive and Compliancy Data

RoHS 2 Directive, abbreviated from Restriction of Hazardous Substances and officially titled Directive 2011/65/EU, restricts the use of certain hazardous substances in electrical and electronic equipment (EEE). EEE is defined as "equipment which is dependent on electrical currents or electromagnetic fields in order to work properly and equipment for the generation, transfer and measurement of such currents and fields and designed for use with a voltage not exceeding 1000 VAC and 1500 VDC" (Directive 2011/65/EU). RoHS-legislation requires certain hazardous substances (such as lead, mercury, cadmium, hexavalent chromium, polybrominated biphenyls (PBB) and polybrominated diphenyl ethers (PBDE)) in EEE to be substituted by safer alternatives. These substances are not completely forbidden, but their concentration is strictly limited.

The directive is a legal document that has been adopted in national laws in all EU countries. However, not all EEE has to comply, because the directive has defined exemptions. Unless a product is specifically listed as one of the exemptions, it will apply to every manufacturer of EEE in EU. Compliance with the RoHS Directive is required before manufacturer can place the CE mark on the product. The directive places an obligation on manufacturers to ensure that any EEE they place on the market has been designed and produced in line with the requirements set out in the legislation.

From product data's perspective, this means that all product structures defined as EEE have to be checked for RoHS compliancy. The company uses a separate "compliancy

system” for verifying and collecting compliance data. This compliancy system provides a storage for compliance reports of single components and product structures. Items that should be verified for compliancy are transferred from PDM to compliancy system, where external company investigates and then verifies component’s compliancy.

However, product data in compliancy system isn’t complete, and the compliancy cannot be fully stated in most products. All items from PDM system shouldn’t be in the compliancy system, but it should include all tangible items that remain in the end product. At the moment, the compliancy system has a total of 27,146 items maintained in it. 3,740 of these items have been dropped intentionally out of scope, leaving 23,406 items to be processed. 7,721 items (33.0%) of these have still unknown or incomplete status with current RoHS compliancy. EU commission has also issued another directive (officially titled Directive 2015/863/EU) that adds 4 phthalates to the list of restricted substances. This directive applies from 22.6.2021. Currently 13,527 items (57.8%) have unknown or incomplete status in this new directive.

This presents a risk that the company has a product containing a RoHS incompatible substance. In the worst case this might conflict with the company’s code of conduct, which clearly states that the company’s products are to comply with international environmental standards and legal requirements. The company’s brand might suffer a negative impact if an external party identifies restricted substances above the threshold limits stated in the Directive.

When defining the end product’s compliancy, each item in the structure is inspected individually. This requires that the item has a reference to the supplier and the supplier’s item code in the PDM system. This data is mandatory when defining the compliancy. Especially old products are lacking this data, making it difficult to verify compliancy for them. This has been taken into account when developing new products in the company, and new product data is in a good level.

Customers are inquiring more frequently RoHS and REACH compliancy of the company’s products. The company has the compliancy system that would respond to this demand, if imported product data and its supplier references would be complete. At the moment, the compliancy system is a large investment that can’t yet be used at its full potential. If product data in this system would be complete, data might also be usable in marketing products, possibly giving the company a competitive advantage.

4.4 Interviews

One aim of this study was to create an overview of master data's quality in the company. Series of interviews were made to support this goal, in order to collect views and opinions of data's quality and usability. Target group included 27 interviewees involved with producing, sustaining or consuming product data.

During interviews was discovered, that product data can be comprehended in many ways. On some occasions, this concept was associated with installed base data, like serial number generation and production documentation. This has nothing to do with product's master data, which is the point of focus in this study.

Appendix 1 presents the questions that were asked from the interviewees. A summary of the answers is presented in the following subsections.

4.4.1 Condition of Product Data

Product data's quality achieves a basic level and it can be used well for general purposes. Product data's quality is not perfect and it is worth questioning. Minor defects can be discovered on a daily-basis. Business can be considered to be fluent even if data quality isn't perfect. Data quality is at sufficient level to make business.

Product data's quality is managed much better than before. For example, the documentation of the company's new products is in better condition than old products, on average. However, lifecycles of the company's products are very long. Product data of bad quality can be a burden for a decade or two. Data receives new requirements over time, and existing data doesn't fulfill all new requirements. Product data's quality is already going in better direction, especially when the oldest products and their data reaches end of life activities. Importance of data quality is also well understood and recognized. Product data has been considered in the company's current processes. When compared to other companies, this company has good processes and data quality is on a satisfying level.

The interviewees use product data for different purposes, and therefore the replies also varied greatly. Active users use the systems on a daily-basis, while infrequent users need to use them very occasionally. Compared to active users, infrequent users might spend more time when searching for information. Nevertheless, information is available even if

infrequent users can't find it immediately. Systems have versatile features for data management. Technical basic attributes, item structures, quantities and descriptions are rated very reliable. Product data overall is trustworthy, but it might contain exceptions that need interpretation. For example, when inspecting validity dates of documents or transitions from one revision to another, it might be difficult to find out when change exactly occurred.

Item descriptions are considered to be sometimes short or unclear. In addition to this, items in sales configurators are wished to contain descriptions in multiple languages. Some items contain translated descriptions, but it varies greatly between products. Sales managers need descriptions of active sales items when preparing quotes for customers. From their point of view, descriptions in multiple languages would increase customer satisfaction, and save working hours from sales managers when preparing quotes in local language. Translations for item descriptions has been taken into account in New Product Development –process, but it might be worthwhile to inspect how it works in practice.

4.4.2 Business Impact

Product data of low quality leads to unnecessary investigations or additional work. Frustration and wasted time caused by this is difficult to estimate, but it certainly exists. Working hours and expenses can be manifold when compared to the effort of creating complete product data right from the start. One interviewee mentioned, that sometimes document has to be requested from a subcontractor, if it doesn't already exist in the company's PDM system. This causes unnecessary waiting periods. Product changes are implemented with quality results, but it could be done more efficiently.

Product data has a great impact on the company's business and its fluency. Quality of product data reflects to customers and suppliers, and it can affect their will to work with the company. It might be even worth investigating how the company's subcontractors evaluate the company's documentation. Data quality and product quality are connected, and data quality problems can lead to quality problems in the real product. This might lead to lower customer satisfaction or product reclamations. In the worst case scenario, the company might sell a ramp downed product if it has active data. This is a quite awkward situation for the company, because the product has already been promised to the customer, but in reality it can't be manufactured anymore. One interviewee noted, that

he has received complaints from a customer related to the documentation of a product and its accuracy.

If product's design process produces low quality data, it leads to ineffectiveness when utilizing data. This can be considered to be a quality problem inside the company. Verifying data's validity consumes time. Data quality has sometimes a direct impact in the product's OTD. It is essential that the company delivers right products and they are delivered on time, and this basic level is achieved quite well. However, this might be done more efficiently with higher data quality.

A couple of interviewees noted, that unreliable product documentation presents challenges in supplier changes. In some occasions, the supplier might have a different version of a document than in the company's PDM system. This has happened for example due to uncontrolled changes made in the supplier's end. Consistency in documentation is important in ensuring that the product has the same properties after the supplier change.

The company currently gets along with its product data, but there's also a chance to improve it and possibly gain more business this way. Current level might not suffice in the future. Digitalization requires that product data is adaptable with suppliers, company's online shop and other new systems. Data has to be reliable and accurate so that it is adaptable in new environments. New systems can't be utilized if data has a poor quality. This can prove to be a real problem in the future. For example, the company is currently commissioning a collaboration tool, used for sharing product data with suppliers. This tool could be utilized more effectively if lifecycle data and validity dates of documents would be more reliable.

4.4.3 Root Causes

Uncontrolled product changes and intentional omissions of the company's processes lead to lower data quality. Data is made by people and it inevitably contains mistakes. Employees that are not aware of data's functionality sometimes produce it. This is sometimes necessary if one wants to complete an item. Technical execution is often done well, but complete understanding behind it is not. Producing high quality data requires some competence from its producer.

Company acquisitions and products acquired in the process import product data that isn't conforming with the company's existing data. Some products have sub-structures that have not been modeled in the item structure. A couple of interviewees mentioned about a product structure that has been documented only in an Excel file. Some software products are missing items completely.

Sometimes data without real use stays active in the systems, especially when people and responsibilities change. This indicates that data's ownership isn't shifted properly. There's almost 50,000 active items in the PDM system, and data management would be easier without inactive items with active data.

4.4.4 Solutions

Quality is achieved through good standards. The company has already improved its data policies over the years. A couple of interviewees emphasized, that the current processes and data policies are highly functional, and it is the work community that needs more discipline in following them. There's no value in developing new processes, if the previous ones are not yet adopted. Good and clear instructions are needed to support this. One responder emphasized, that probably no one has intentionally damaged data quality, but that the data policies are sometimes unclear.

There has to be a fine balance in data management. Too strict data management would make the process of producing data slow, and require an unnecessary amount of resources. On the other hand, too careless data management would produce incomplete product data, which results in additional work when data is used and refined for new purposes. More surveillance is needed when accepting new product data, to prevent production of weak product data. Ideally, the surveillance has a definite standard that isn't to be violated.

The company could use more metrics in data quality. Data quality doesn't have a systematic surveillance at the moment. Data quality is not in entire work community's interests. Not everyone perceives that data is in everyone's responsibilities, and therefore data ownership should be clarified. Data owners should commit to managing their data, and be able to inspect and analyze its quality effortlessly. This could be improved by creating metrics or reports for different stakeholders, measuring defects in data in their responsibilities.

Some responders think that data ownership should be shared more between the work community. Some responders had the opposite view: a small group of competent people could manage product data better. When responsibility is shared for a larger number of people, the group includes more infrequent users with too little competence to fulfill this responsibility. This might result in new data quality issues. There's also a compromise available. Data ownership could be shared between the work community, with only selected, competent people able to modify the data.

Product data's sustainment is currently centralized to a small group of people, with good competence and a high number of repetitions to do that. On the other hand, data production is more decentralized. It is produced by a larger number of people who do it more rarely. They have less repetitions and competence, which affects the quality of new data. Good single products also arise, but the best practices should be standardized and become a frequent way of conducting.

Perseverance and patience are needed. The subject of data quality should remain present continuously, year after year. Data quality has a meaning. Appreciation for data should be emphasized more. The entire work community needs to understand the value of high data quality.

4.5 Summary

One purpose of this study was to form a balanced view of product data's quality in the company. To summarize this: data is not perfect, but it works reliably for the basic purposes. Business can be executed, but sometimes data delays it. The challenge of data management is increasing, because there's more products and active items all the time. Data achieves a satisfying basic level, but it doesn't fulfill some "bonus" functions. Data meets new requirements in the future, and it would be recommended to invest in data quality as soon as possible. This is a good chance to improve the company's efficiency in daily activities. Data quality can be improved, there's no doubt about that.

5 Conclusions

The goals of this study were to create an overview of data quality in the company, investigate benefits of higher data quality, but also present a recommendation of how to control data quality better.

5.1 Summary

Main concerns in PDM system's data quality were disconnected service items without documentation, inactive items with active data and lifecycle statuses. The company utilizes a report tool that can be used to discover defects in data. The results include mainly defects in incomplete items, active sales items and parameters affecting productivity metrics and end product's expenses. In addition to this, completely verifying product's compliancy with RoHS- and REACH-legislation is impossible with incomplete manufacturer and component data.

A series of interviews was made for 27 people associated with product data. The questions related to product data's general condition, usability, reliability, current management of product data and its impact to the company's business. Appendix 1 presents the entire questionnaire. Most interviewees stated, that generally speaking product data's quality is on a good level in the company. Different stakeholders can point out defects in data quality, as mentioned in the previous chapters. Data quality varies between products and product areas. The company has developed its processes and data policies over the years, and especially new products have almost blameless product data. The company maintains a considerable amount of product data, and it is maintained by a moderately small group of people. This itself is a challenge, and the company is currently planning to decentralize responsibilities related to data management.

The lifecycles of the company's products are long, and their product data has to be sustained for several decades. Therefore data quality can have benefits or consequences for a long time. Some interviewees stated, that the current level might not be enough in the future, because digitalization will bring new systems that have to be able to utilize master data. To summarize, the company has data quality on an adequate level and it can be used for business purposes. Perfect level can never be reached. The company

maintains a large product portfolio, resulting in remarkable amount of data. When considering the amount of product data, its data management can be considered to be in good shape. Existing product data contains some quality issues, and from many different reasons. It is definitely worthwhile to examine data quality more, because surely not all problems were covered in this study.

5.2 Benefits of Data Quality

Stark states in his book (2015), that product data is important part of PLM. Product data contains the knowledge and know-how about the way product is designed, manufactured, supported, used and recycled. The quality of product data is a key element of product's success. Poor data management results in wasted time, rework costs, and slow time to market. Speed to react is often a strategic factor in the manufacturing industry (Sääksvuori, Immonen 2002). Data doesn't look after itself, and like anything that's not properly organized and maintained, won't perform as required. Product data is a strategic resource, and its management is a key issue.

Generally speaking, PLM activities help any company in introducing new products more efficiently, managing its product's lifecycle better and enabling offering of new services for existing products (Stark, 2015). According to Stark, PLM is a high-level business activity, uniting lower-level product-related activities together. Stark also categorizes business objectives of PLM into four categories, as presented in table 7.

Table 7. Business objectives of PLM (Stark, 2015).

Category	Business objective
Financial Performance	<ul style="list-style-type: none"> ▪ Earlier market intro and increased revenues this way ▪ Reduced product development costs ▪ Extended product life and increased revenues this way ▪ Reduced recall costs
Time Reduction	<ul style="list-style-type: none"> ▪ Reduced overrun project times ▪ Reduced engineering change time ▪ Reduced time to market ▪ Reduced time to profitability
Quality Improvement	<ul style="list-style-type: none"> ▪ Reduced manufacturing process defects ▪ Reduced product returns ▪ Reduced customer complaints ▪ Reduced scrap
Business Improvement	<ul style="list-style-type: none"> ▪ Increased new product release rate ▪ Increased the part reuse factor ▪ Increased product traceability ▪ 100% configuration conformity ensured

Product data is used and reused by many of the organizational entities within the company (Stark, 2015). The approach to its overall management has to be cross-functional. Stark states, that attempts made by individual departments to impose order might be frustrated by other departments. Product data belongs to the product, not to individual departments.

Integrating new systems to the existing data bases is not possible without reliable source data. The company has a topical example, because it is currently commissioning a new tool used to share product information with its suppliers. During product changes, this tool is used to assemble a data package containing engineering change order and updated documents and product structures to the supplier. Collecting correct documents is difficult, because not all active documents have been changed to "Accepted" status. Creating data package requires interpretation and manual work when picking the valid documents. In ideal situation, data quality would be at so high level that there would be no need to investigate which one is the latest document.

Productivity metrics used at the company's production could be advanced by defining routings to all manufactured items. Improved accuracy in routings would also define end product's manufacturing expenses more accurately. Generally speaking, all reports based on master data will become more precise with higher master data quality.

The PDM system is missing some documents especially in the oldest products. Sometimes a document has to be requested from a subcontractor, and engineering would be more fluent without waiting periods caused by this. Supplier changes would become more fluent with reliable product data and documentation. Quality of product data reflects to customers and suppliers, and it can affect their will to work with the company. Higher data quality could increase the company's reputation, especially among suppliers.

RoHS and REACH legislation creates new requirements for the company's products. A considerable amount of working hours could've been saved in the project designed to verify the compliancy of existing products, if product data would've supported this. Some products have item structures that have not been defined to the component level, and some components are entirely missing manufacturer information. The end result is not perfect, because verifying product's compliancy requires manufacturer information from all of its components. Customers are inquiring more frequently the compliancy of the company's products, and complete data would enable responding to this demand. If

product data in this system would be complete, data might also be usable in marketing products, possibly giving the company a competitive advantage.

5.3 Solutions

First priority in cleansing existing data should be to distinguish inactive structures and items from active. Data is proved to contain many items and revisions that are no longer used. Useless items and outdated revisions should be archived before anything else. This would diminish the amount of active data, and at the same time increase efficiency in the upcoming data cleansing. This way, future data cleansing would be directed to the most critical data. Remaining items could still prove to be a massive amount of data. Completely cleansing existing data would require too much resources, simply put. Therefore, first step should be to prioritize the defects. Cleansing should be focused in the most business-critical data, for example products with the most revenues, new products or items with active sales status.

Cleansing data is very time consuming manual work. Usually items have to be opened from the system one at a time to make modifications. Any new tool that would make it faster to modify data would be warmly welcome. At the moment, the company doesn't even have an employee who does data cleansing regularly. Therefore, a resource has to be determined for this task. Some data was cleansed during this study, especially defects discovered from the ERP system's data. It was easy to interfere with them by utilizing the report tool mentioned in chapter 4.2. Although some of the defects were completely corrected, more have also emerged during this study. It should be investigated where the defects are coming from, and interfere with the source of problems.

Cleansing old data would likely improve its usability, but it doesn't solve the core problem in data management. Ad hoc repairs for data corrects mistakes in data, but not the mistakes of producing low quality data in the first place. Data quality should be good to begin with, so that it doesn't have to be cleansed later. Ideally, data would have a clear, defined standard. This standard would determine a level of quality for product data, and all new data should first be verified to conform this level before it is accepted. This would make sure that new product data has a good quality, or at least the quality defined in the standard. According to Kropsu-Vehkaperä, Haapasalo and Harkonen (2009), there's always

need for a person to check the correctness of data, as the systems cannot be programmed to conduct self-checking.

Some interviewees mentioned that the company has well developed processes to produce high quality data, but data quality suffers more from people not following the processes. According to Stark (2015), it is only when all the resources of a company are organized and managed to achieve the objectives of PLM, that the objectives can be met. Data quality has to be built in company's procedures and working culture. Error creation and propagation must be prevented, because it's only too easy for a user to introduce an error into data. Once the error is in the data, it can be difficult to find, and it can be even more difficult to remove its effects. One instruction for creating new items in the ERP system was discovered to have official and unofficial version.

Reason for the lack of process discipline might be people not understanding data policies, or recognizing the value of data quality well enough. This should be examined more before making a final conclusion. A new instruction or training could be developed to clarify the company's data policies and processes. Over the years, the company has developed its processes to create high quality data, but a sufficient quality surveillance could take data quality to new level. Data quality monitoring needs more metrics and report tools to support this. Especially weak spots and business critical aspects of data should be monitored, so they can be interfered with. The point of focus in the metrics could be:

- incomplete items (in both PDM and ERP systems)
- not "Accepted" items and documents in active structures
- consistency between the company's documents and its subcontractor's documents
- expenses caused by data quality

Next step could be to compose an estimate of expenses caused by data quality. At the moment, there's no method in the company to do this. This could be approached by first identifying all stakeholders utilizing product data, or so called data consumers. Only they can describe what kind of data quality problems they have to deal with, how much time

they waste on them and what kind of data needs they have. The estimate of expenses should be done in co-operation with them. This is a valid subject for another study. The problems should be inspected in the finest detail, and the research done in this study is not enough for this.

At the same time, it should be estimated how much does it cost to repair existing data. Calculating these can prove to be difficult, but this would put the problems to the right extent. When there are facts about how much the problems are causing expenses to the company, it would be possible to decide if they are worth intervening or not. This would allow making a business decision: is it sensible to get along with present problems, or would it be more profitable to fix them? To achieve this, the problems of data quality should be examined more. Points of interest should be how much time data cleansing consumes in the entire company, how much time is wasted in additional work caused by low data quality, and what kind of effect data quality has in product's lead time. Data management always brings mandatory expenses, and these cannot be avoided. This should be taken into account in the estimate, and it should differentiate mandatory expenses from unnecessary expenses.

The company's products are increasing in number, and so is the amount of product data. In fact, the company has more NPI projects this year than ever before. Therefore, the business value of data quality will become more important than before. Current level is good, but it might not suffice in the future. Data quality and its potential should be examined now.

References

Ballard Chuck. 2013. "Align MDM and BPM for Master Data Governance, Stewardship, and Enterprise Processes". Retrieved 23.1.2018 from <http://www.redbooks.ibm.com/abstracts/tips1055.html?Open>.

Batini Carlo, Scannapieca Monica. 2006. "Data Quality: Concepts, Methodologies and Techniques". Springer-Verlag Berlin Heidelberg.

CIMdata. "All About PLM". Retrieved 24.1.2018 from <https://www.cimdata.com/zh/resources/about-plm>.

"Directive 2011/65/EU of the European Parliament and of the Council of 8 June 2011 on the restriction of the use of certain hazardous substances in electrical and electronic equipment Text with EEA relevance". Retrieved 20.2.2018 from <http://data.europa.eu/eli/dir/2011/65/oj>.

Haug Anders, Arlbjørn Jan. 2011. "Barriers to master data quality". *Journal of Enterprise Information Management* Vol. 24 No. 3 (2010). pp. 288-303.

Kropsu-Vehkaperä Hanna, Haapasalo Harri. 2011. "Defining Product Data Views For Different Stakeholders". *Journal of Computer Information Systems* (2011).

Kropsu-Vehkaperä Hanna, Haapasalo Harri, Harkonen Janne. 2009. "Product Data Management Practices In High-Tech Companies". *Industrial Management & Data Systems* Vol. 109 Issue: 6 (2009). pp. 758-774.

Loshin David. 2010. "Master Data Management". Elsevier Inc.

Lucas Ana. 2010. "Towards Corporate Data Quality Management". *Portuguese Journal of Management Studies* XV.2 (2010). pp. 173-95.

Rahm Erhard, Do Hong Hai. 2000. "Data Cleaning: Problems and Current Approaches". *IEEE Techn. Bulletin on Data Engineering* (2000).

Shtub Avraham, Karni Reuven. 2010. "ERP: The Dynamics of Supply Chain and Process Management, Second Edition". pp. 14.

Silvola Risto, Jääskeläinen Olli, Kropsu-Vehkaperä Hanna, Haapasalo Harri. 2010. "Managing one master data – challenges and preconditions". *Industrial Management & Data Systems* Vol. 111 Issue: 1 (2010). pp. 146-162.

Silvola Risto, Harkonen Janne, Vilppola Olli, Kropsu-Vehkaperä Hanna, Haapasalo Harri. 2016. "Data quality assessment and improvement". *International Journal of Business Information Systems* Vol. 22 No. 1 (2016). pp. 62-81.

Stark John. 2015. Product Lifecycle Management, 21st Century Paradigm for Product Realisation, Third Edition. Springer International Publishing Switzerland.

Sääksvuori Antti, Immonen Anselmi. 2002. "Tuotetiedonhallinta – PDM". Talentum Media Oy.

Tozzi Christopher. 2017. "What is Data Quality? Explaining What Data Quality Actually Means". Retrieved 24.1.2018 from <http://blog.syncsort.com/2017/07/big-data/what-is-data-quality-definition-examples/>.

Wang Richard. 1998. "A product perspective on Total Data Quality Management". Communications of the ACM 41.2 (1998). pp. 58-65.

Appendix 1. Questionnaire

- What is the condition of our product data from your point of view?
- Do you think product data can be used for our common business purposes?
- Does product data serve your purposes?
- Is product data reliable from your opinion?
- Does product data's quality cause problems for you?
- What other observations have you made from product data's quality?
- Is product data causing expenses, or could it be possible to get savings with better product data?
- What is product data's impact to the company's business?
- Is product data management in control from your opinion?
- What are the consequences of bad product data quality?
- What has to be done or what is needed, so that product data's quality stays in control?