THE INFLUENCE OF THE TIME FACTOR ON PRODUCT DATA QUALITY
IN THE FAST-MOVING CONSUMER GOODS INDUSTRY

Enrique Batani Oseguera

Company Supervisor: Yvonne Hoeting - Process Effectiveness Manager
University 1st Supervisor and Examiner: Dr. Sc. (Tech) Ville Ojanen
University 2nd Supervisor: Dr. Sc. (Tech) Janne Huiskonen
# ABSTRACT

**Author:** Enrique Batani Oseguera

**Title of thesis:** The influence of the time factor on product data quality in the Fast-Moving Consumer Goods Industry.

**Year:** 2016  |  **Place:** Verden, Germany.

**Master’s Thesis**
Lappeenranta University of Technology.
School of Business and Management.
Master’s Degree Program in Global Management of Innovation and Technology
124 pages, 30 figures and 15 tables and 4 appendices.

**Examiner from Mars GmbH, Germany:**
Yvonne Hoeting - Process Effectiveness Manager, Market Supply Chain.

**Examiners and Supervisors from LUT:**
Supervisor and Examiner - Dr. Sc. (Tech) Ville Ojanen
Supervisor - Dr. Sc. (Tech) Janne Huiskonen

**Keywords:** Product master data, Data quality timeliness, Data quality metrics, Master data management, Data quality management, Data availability, Supply Chain Management

As technology and digitalization expand to all areas of life and economy new applications and platforms need to be fueled with data of quality. Similarly, increasing legal and market requirements have led retailers from the Fast-Moving Consumer Goods industry to diversify their presence and services to end-consumers in an omni-channel context, thereby raising their product data requirements regarding content and data availability. However, a clear point in time for the delivery of product data has not been specified for product manufacturers causing uncertainty and impacting negatively on the quality of product data.

This thesis studies the influence of the time factor on the quality of the product data in the context of cross-company data communication. The research focuses on how a manufacturer can meet the required quality for its product master data on time to enable omni-channel commerce for its trading partners in the Fast-Moving Consumer Goods industry. The approach to answer the research question uses literature review and the multiple-case study methodology through face-to-face interviews with representatives of four leading companies in the retail industry in Germany. Additionally, this study analyses the feasibility of a solution proposed by industry specialists called “Preliminary Trade Item”.

The results show a way to harmonize the manufacturers’ data capabilities with the retailers’ product data requirements and to reduce uncertainty by delivering the product data when it is final and when it is needed by the retail company considering an omni channel perspective.
ACKNOWLEDGEMENTS

I would first like to thank the thesis supervisors and mentors from the company and university for their continuous support during the development of this work, Yvonne Hoeting and Mareike van Leeuwen from Mars GmbH, and Ville Ojanen from the Lappeenranta University of Technology in the School of Business and Management. They were always reachable whenever I had a question about my project or about the company. They motivated me to achieve the best results in the project whenever new challenges were found and helped me to steer in the right direction.

I would like to give my special thanks to the people that facilitated this study with active support in the organization GS1 Germany and its subsidiary companies Smart Data One and 1WorldSync.

I would like to thank the case company representatives who were involved in the conduction of the interviews for their engagement, cooperation and extensive support with the topic beyond the interviews for the best outcome of this thesis.

Finally, I must express my very profound gratitude to my family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and conducting this project. This accomplishment would not have been possible without them. Thank you.

Verden, Germany, November 2016.

Enrique Batani Oseguera
TABLE OF CONTENTS

TABLE OF CONTENTS ......................................................................................................................... I
LIST OF TABLES ............................................................................................................................... III
LIST OF FIGURES ............................................................................................................................... IV
LIST OF ABBREVIATIONS .................................................................................................................. V

1. INTRODUCTION ............................................................................................................................ 1
   1.1 Supporting Organization and Research Cooperation ................................................................. 2
   1.2 Background of the Study ........................................................................................................... 7
   1.3 Problem Definition and Research Objectives ....................................................................... 9
   1.4 Research Scope and Delimitations ....................................................................................... 16
   1.5 Structure of the Study ........................................................................................................... 18

2. THEORETICAL BACKGROUND .................................................................................................. 20
   2.1 Supply Chain Management .................................................................................................... 22
   2.2 Management of Product Cycles ............................................................................................ 28
   2.3 Management of Master Data ................................................................................................ 35

3. DATA COLLECTION AND METHODOLOGY ........................................................................... 48
   3.1 Research Design .................................................................................................................... 50
   3.2 Description of Study Cases .................................................................................................. 51
   3.3 Data Collection ..................................................................................................................... 52
   3.4 Data Reliability ..................................................................................................................... 59

4. CASE STUDIES ............................................................................................................................. 60
   4.1 Company A .............................................................................................................................. 61
   4.2 Company B .............................................................................................................................. 65
   4.3 Company C .............................................................................................................................. 69
   4.4 Company D .............................................................................................................................. 73
5. ANALYSIS OF RESULTS................................................................. 78
   5.1 The Listing Process................................................................. 79
   5.2 Preliminary Trade Item: How soon is soon enough? .................... 88
6. CONCLUSIONS.................................................................................. 95
   6.1 Theoretical Contribution............................................................ 98
   6.2 Managerial Implications and Recommendations ......................... 99
   6.3 Limitations................................................................................. 100
   6.4 Further Research......................................................................... 101
REFERENCES......................................................................................... 102
APPENDICES.......................................................................................... 111
   Appendix - 1: GTIN - GPC - Attributes - Values .............................. 111
   Appendix - 2: Data and Data Quality Conceptual Model ................... 112
   Appendix - 3: Questionnaire for Interviews....................................... 113
   Appendix - 4: PTI: Challenges - Requirements - Benefits ................ 115
LIST OF TABLES

Table 2.1 - Multi-channel versus Omni-channel Management ............................................. 26
Table 2.2 - Product Lifecycle Phases .................................................................................... 29
Table 2.3 - Master Data Elements ....................................................................................... 36
Table 2.4 - Examples of PLM Product Information ............................................................... 37
Table 2.5 - Top 20 Quality Dimensions ............................................................................... 45
Table 3.1 - Research Design ............................................................................................... 50
Table 3.2 - Summary of Questionnaire ................................................................................ 54
Table 3.3 - Attribute Clusters ............................................................................................. 56
Table 4.1 - Core Case Study Definitions ............................................................................. 60
Table 4.2 - Listing process of Company "A" ...................................................................... 63
Table 4.3 - Listing process of Company "B" ...................................................................... 67
Table 4.4 - Listing process of Company "C" ...................................................................... 71
Table 4.5 - Listing Process in Company "D" ..................................................................... 75
Table 5.1 - Proposed General Durations for Listing Process ............................................ 82
Table 5.2 - Number of Times Data Is Retrieved ................................................................. 90
LIST OF FIGURES

Figure 1.1 - Mars Inc. Total Global Sales per Segment 2016 .................................................. 3
Figure 1.2 - Mars Inc. Overview 2016 ....................................................................................... 4
Figure 1.3 - GS1 Germany Involvement 2016 ......................................................................... 6
Figure 1.4 - Problem Definition .............................................................................................. 11
Figure 1.5 - Deductive Approach to the Research Problem ...................................................... 13
Figure 1.6 - Structure of the Thesis ........................................................................................... 19
Figure 2.1 - Focus Area of Research ....................................................................................... 20
Figure 2.2 - Main Elements of Enterprise Management ........................................................... 27
Figure 2.3 - Delivery of Goods in the Retail Industry ................................................................. 32
Figure 2.4 - PLM, MDM and SCM Architecture Integration ...................................................... 34
Figure 2.5 - Examples of Product Master Data Content ............................................................. 38
Figure 2.6 - GPC Hierarchical Structure .................................................................................. 39
Figure 2.7 - Data Synchronization Solutions ............................................................................ 41
Figure 2.8 - GDSN Data Communication Process ................................................................... 42
Figure 2.9 - MDM and DQM ................................................................................................... 43
Figure 3.1 - Knowledge Sources .............................................................................................. 53
Figure 4.1 - Listing Process in Company "A" .......................................................................... 62
Figure 4.2 - Listing Process in Company "B" .......................................................................... 66
Figure 4.3 - Listing Process in Company "C" .......................................................................... 71
Figure 4.4 - Listing Process in Company "D" .......................................................................... 75
Figure 5.1 - Cross-company Product Data ............................................................................ 78
Figure 5.2 - Listing Process Compilation ............................................................................... 79
Figure 5.3 - Extrapolation from Company "A" to "D" ............................................................... 81
Figure 5.4 - Generalized Listing Process .................................................................................. 84
Figure 5.5 - Time Factor Offline Commerce ........................................................................... 86
Figure 5.6 - Time Factor Online Commerce ........................................................................... 86
Figure 5.7 - Attribute Usage in the Listing Process ................................................................. 87
Figure 5.8 - PTI: Requirements - Benefits - Challenges ............................................................ 89
Figure 5.9 - Vicious Circle of Defective Data ......................................................................... 92
Figure 5.10 - PTI Proposal ...................................................................................................... 93
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2B</td>
<td>Business to Business</td>
</tr>
<tr>
<td>B2C</td>
<td>Business to consumer</td>
</tr>
<tr>
<td>BOL</td>
<td>Beginning of Life</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer Aided Designs</td>
</tr>
<tr>
<td>CPD</td>
<td>Collaborative Product Definition</td>
</tr>
<tr>
<td>CPG</td>
<td>Consumer Packaged Goods</td>
</tr>
<tr>
<td>DQM</td>
<td>Data Quality Management</td>
</tr>
<tr>
<td>EANCOM</td>
<td>European Article Number Communication</td>
</tr>
<tr>
<td>ECR</td>
<td>Efficient Consumer Response</td>
</tr>
<tr>
<td>EDI</td>
<td>Electronic Data Interchange</td>
</tr>
<tr>
<td>EPC</td>
<td>Electronic Product Code</td>
</tr>
<tr>
<td>FIR</td>
<td>Food Information Regulation</td>
</tr>
<tr>
<td>FMCG</td>
<td>Fast-Moving Consumer Goods</td>
</tr>
<tr>
<td>GDSN</td>
<td>Global Data Synchronization Network</td>
</tr>
<tr>
<td>GLN</td>
<td>Global Location Number</td>
</tr>
<tr>
<td>GPC</td>
<td>Global Product Classification</td>
</tr>
<tr>
<td>GS1</td>
<td>Global Standards One</td>
</tr>
<tr>
<td>GTIN</td>
<td>Global Trade Item Number</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>LUT</td>
<td>Lappeenranta University of Technology</td>
</tr>
<tr>
<td>MDM</td>
<td>Master Data Management</td>
</tr>
<tr>
<td>MDSP</td>
<td>Master Data Service Provider</td>
</tr>
<tr>
<td>PDM</td>
<td>Product Data Management</td>
</tr>
<tr>
<td>PII</td>
<td>Product Integration Information</td>
</tr>
<tr>
<td>PIM</td>
<td>Product Information Management</td>
</tr>
<tr>
<td>PLM</td>
<td>Product Lifecycle Management</td>
</tr>
<tr>
<td>PMD</td>
<td>Product Master Data</td>
</tr>
<tr>
<td>QR</td>
<td>Quick Response</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and Development</td>
</tr>
<tr>
<td>RFID</td>
<td>Radio Frequency Identification</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>ROI</td>
<td>Return On Investment</td>
</tr>
<tr>
<td>RQ</td>
<td>Research Question</td>
</tr>
<tr>
<td>SCM</td>
<td>Supply Chain Management</td>
</tr>
<tr>
<td>SDO</td>
<td>Smart Data One</td>
</tr>
<tr>
<td>SKU</td>
<td>Single Key Unit</td>
</tr>
<tr>
<td>SQ</td>
<td>Sub-Question</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Advancements in technology have led to a myriad of digitalized products and services that engage humans in all areas of life and economy (Otto and Österle, 2016). According to Amiona (2014), there are ten acknowledged life areas that digitalization has reached, influencing from the way people drive, how professionals conduct business, to how individuals search for a partner. However, the term digital in the current times has become oversimplified by only differentiating between analog and digital technologies. Thus, it is pertinent to remark how this concept revolutionizes industries: (1) It’s capability to disrupt business models, (2) Its focus on data and technology as basis of competitive advantage, (3) its emphasis on customer experience, and (4) a new and different mindset and working methods (Friedlelin, 2016).

The retail industry is an example that harnesses digitalization by offering online shops, traditional brick and mortar stores and combinations of both in order to extend their offering across platforms and provide customers with a convenient way of shopping. Furthermore, an emerging trend called “Clever Commerce” shows that customers are empowered by intelligent services that deliver robust and intuitive paths to find the right product automatically at the best price (Trendwatching, 2016). Recalling the life areas that digitalization has touched, shopping is one of the broadest with initiatives designed for searching stores, consult and compare product information, product and service buying, performing payments, using coupons, reading manuals or obtaining recommendations (Amiona, 2014).

Alongside, customers have increased their expectations in regard of products and services as technology has developed further and the amount of information available in the palm of customer’s hands grows rapidly (McGovern, 2016). Customers expect to have interactions with companies through every channel and find the same prices and promotions regardless if via online shops, emails, regular post or SMS. To accomplish that, companies must have consistent, current and complete data about the customers and the products throughout the channels being served (Schemm, 2012).
Therefore, a key resource that fuels the digital services and new business models is data. This resource can be used to engage new customers, discover pricing models and develop novel economic systems, as it accounts for a great percentage of the strategic business innovations (Otto and Österle, 2016). Taking the retail industry as example, the term Efficient Consumer Response (ECR) has been coined to refer to the collaborative strategies and operating practices between retailers and suppliers by exchanging data such as sales forecasts, brand loyalty, information on product specifications, early notifications of new models and feedback on competitors among others. Consequently, this collaboration reduces costs across the supply chain and aids at better fulfilling consumer needs (Zentes et al., 2011).

1.1 Supporting Organization and Research Cooperation

This study was funded by Mars GmbH and supervised through its complete development by the Market Supply Chain department based in Verden, Germany. Additionally, the organization GS1 Germany GmbH and its subsidiary company Smart Data One GmbH, both being based in Cologne, Germany, provided active support in the definition of the research problem as well as by facilitating the empirical study with companies active in the retail industry.

1.1.1 Supporting Organization

Mars Inc. started its operations in 1911 with a candy factory in Tacoma, Washington. From that point onwards, Mars has diversified its portfolio, see Figure 1.1, covering the segments Chocolate, with 29 brands, Petcare, with 41 brands, Wrigley, with nearly 34 brands, Food, with 12 brands, Drinks (distinguished in yellow for Figure 1.1), with 5 brands, and Symbioscience (distinguished in green for Figure 1.1), which acts as a business incubator for ideas generated across Mars’ segments (Mars, 2016b). Currently, Mars has around 80,000 associates around the globe in 78 countries. Organizationally, Mars has been kept a private and family-owned enterprise, thriving through its core five principles that influence the strategies of the company and its every day operations: Quality, responsibility, mutuality, efficiency and freedom (Mars, 2016a).
The company’s story began with the Milky Way bar, since then, Mars has grown to achieve in sales more than $35 billion USD across the globe (Mars, 2016a). Thereby, the different segments of the company have contributed with billion-dollar brands as M&M’s, Snickers, Dove / Galaxy, Mars / Milky Way and Twix and from the chocolate segment, Pedigree, IAMS (excluding Europe), Banfield, Royal Canin and Whiskas from the Petcare segment and Uncle Ben’s from the Food segment, see Figure 1.2 (Mars, 2016b).
Mars Germany has been operating since 1959 with its first sales and production facilities in Verden. Nowadays, the company has nationwide 2,600 associates from 43 different nationalities distributed across five sites: Verden, Viersen, Minden, Cologne, and Unterhaching. Thereby, Mars Germany contributes in five product segments: Petcare, Chocolate, Food, Wrigley and Drinks. During 2014, Mars GmbH reached a total turnover of 1.8 billion euros (Mars, 2016d).
1.1.2 Research Cooperation

Global Standards One (GS1) had its first standard applied in the year 1974 with the introduction of the bar code, standardizing bar codes for any item to be readable worldwide. GS1 is a global, neutral, non-profit standards organization with the goal of enhancing the supply chain through efficiency and transparency. Their standards have improved the exchange of information in some of the world’s biggest industries such as retail, healthcare, transports and logistics, providing them with traceability, resource optimization and communication capabilities (GS1, 2016a). Presently, GS1 is active in over 100 countries and it is the most important standardization body in the Fast-Moving Consumer Goods (FMCG) and retail industries (GS1, 2016f, 2016b).

GS1 Germany, with more than 171,000 clients and users, is the second largest organization within the GS1 network. Reportedly, GS1 Germany participated with different industries to develop solutions for specific requirements and processes, from supporting in the struggle against product piracy in the fashion industry, improving merchandise availability and traceability in the technical spare parts market, to optimizations in its data pool for the consumer goods and other industries in cooperation with its daughter companies 1WorldSync and Smart Data One, see Figure 1.3 (GS1 Germany, 2015; 2016).
1WorldSync was founded as a joint venture by GS1 US and GS1 Germany. This company is the leading cross-company product information network, serving more than 23,000 global brands in 60 countries. 1WorldSync provides solutions in B2B supply chain, product transparency and compliance (GS1 Germany, 2015; 2016; 1WorldSync, 2016a). 1WorldSync’s data management services foster the automatic digital communication of product master data through data pools, certified in the Global Data Synchronization Network (GDSN). The company offers access to external data pools based on GS1 standards to retailers and manufacturers (1WorldSync, 2016a).

Smart Data One (SDO) provides services in product data management, assisting retailers and manufacturers by assuring the quality of the product data. SDO is completely integrated into the systems of 1WorldSync. The service portfolio of the company range from collecting, completing and verifying data, creating product images and completely managing companies’ product data (Smart Data One, 2016). In year 2015, SDO took the data quality initiative to increase the quality of product master data to 100 percent (GS1 Germany, 2015).
1.2 Background of the Study

Supply chains are faced with new challenges as knowledge, workers, suppliers and customers are increasingly sourced from around the globe, whilst products and services are becoming more customized and turning obsolete in ever shrinking life cycles (Sengupta, 2013). Additionally, the different channels that a manufacturer can serve to reach its customers are various and their operations can change based on the commerce nature (Boykin, 2016). Further, the challenges confronted in the supply chain for Fast-Moving Consumer Goods (FMCG) can be entirely different when looked from a retailer perspective, than those faced by manufacturers (Root, 2016).

Technological progress impacts different areas of life and economy, and the management of a supply chain is not an exception. Nowadays, all enterprises have turned into digital companies, but not necessarily their supply chains, despite they might be enhanced by digital processes, since the underlying practices remain traditional. Thus, a re-invention of the traditional supply chain requires it to be: (1) Connected, traceable on demand and smooth with collaborative and operative models, (2) Intelligent, data driven, (3) Scalable, integrated, flexible and personalized, and (4) Rapid, pro-active and responsive (Hanifan et al., 2014).

Likewise, supply chains in the consumer goods industry consider all stakeholders related directly or indirectly to the process of receiving and fulfilling customer orders. Supply chains can be described as dynamic with a permanent stream of goods, finance and data across different stages (Zentes et al., 2011). The latter stream, data or information, is vital for enabling collaborative planning but also in a context of multi-channel commerce, where the information made available to the customers in online shops is key to thrive and convert the physical liabilities of brick and mortar stores into assets of a combined online-offline strategy (Downes, 2012). Similarly, the company Cloudtags describes this strategy as turning multi-channel commerce into omni-channel commerce through integration across channels with real-time information about inventories, customers, and products (Cloudtags, 2016).
Encompassing the supply chain an information-driven business concept named Product Life-cycle Management (PLM) can be found. PLM manages the information of products (product data), product innovations and product modifications through their life-cycle (Grieves, 2006; Sääksvuori and Immonen, 2008). In times of increasing product complexity, shrinking product life-cycles, and broader networked enterprises, PLM aids at shortening the time-to-market of new products and reducing costs. Grieves affirms that “Information is an intrinsic characteristic of a Product” which is proven by the progressive development of product information, initiated with the creation of standards and tools of measurement, then, blueprints and detailed product information became embedded into the product, and nowadays, digital models are created by linking a physical product with its digital equivalent, attaining a minimized waste of physical resources (Grieves, 2006).

The life-cycle of a product is divided into three major phases: Beginning of Life (BOL), Middle of Life (MOL) and End of Life (EOL). Each of these phases holds different information about the product and its environment, for example, BOL includes the information generated in the stages from the product conception to its production, while MOL includes the product delivery, usage and its maintenance, storing information about its performance in different conditions, whereas EOL covers the processes related to the retirement of the product and the respective information (Kiritsis, 2011; Terzi et al. 2010).

In order to profit from product information at a cross-company level, the discipline of Master Data Management (MDM) performs as a data hub for the exchange of product or other types of information. Gartner defines MDM as a technology based discipline that combines business practices with Information Technology (IT) to guarantee the uniformity, accuracy, governance, semantic consistency and ownership of a company’s master data (Gartner, 2016). However, if MDM incurs in poor practices, it might lead to fragmented and inconsistent product data, among other types of data that could cause SCM inefficiencies, weaker market penetration, and increasing compliance costs (Oracle, 2011).

The quality of master data has a profound impact on a company. Thus, enterprises strive to measure and improve their data quality. Data quality can be considered a multi-dimensional
concept. Therefore, there are various dimensions to measure the quality of master data, yet, the most accepted quality dimensions are: correctness, consistency, completeness, actuality and availability (Otto and Österle, 2015). Nonetheless, when master data is shared across multiple enterprises, like product master data between manufacturers and retailers for new products, each company’s metrics for data quality could differ, revealing the need for industry-wide standards. The correct interpretation of cross-company transactions can only succeed through master data synchronization, for example, by clearing which exact products will be offered or ordered, and how these products will be identified across companies. Thereby, data pools are an alternative to data exchange between companies (Schemm, 2009).

1.3 Problem Definition and Research Objectives

The following sub-sections describe the elements of the research problem and its context based on the theoretical background. Also, a solution proposed by the industry is presented and the approach of this study to analyze the feasibility of that solution. Alongside, the research objectives and aims of this study are mentioned, thus, in this section the main research question and sub-questions are introduced.

1.3.1 Problem Definition

As described in the background of the study, Data Quality Management (DQM) has a significant impact on the financial and operational performance of companies, specially, when the data is shared across multiple enterprises. Therefore, great efforts have been done to reduce the frictions in the data communication through technological platforms and standards in the FMCG and retail industries. Nonetheless, defining standards driven by the consensus of stakeholders (GS1, 2016a) can be a herculean task, considering that the stakeholders are different manufacturers with various product innovation processes, and retail companies with diverse business strategies serving multiple customer segments through numerous channels, and also, for different product categories. Hence, harmonizing the manufacturers’ capabilities to deliver Product Master Data (PMD) with the requirements from the retail companies’ demands deeper research, thus, this master’s degree thesis.
Products of the FMCG industry are sold in short times with low margins, yet, the volume of sales is high (Investopedia, 2016). Examples of products in the FMCG industry include chocolates, pet food, meat, candy, milk, beer, toilet paper or soap, among many others. Hence, due to the high volumes of merchandise, fast consumption and, based on the product category, short shelf life, SCM and Category Management (CM) partnerships based on collaboration and information sharing are key to thrive in business (Zentes et al., 2011).

In addition to market and SCM information, manufacturers exchange PMD about their new products with the retail companies. Moreover, as stated by different sources (Sara Novak, 2015; Zentes et al., 2011), PMD is requested well in advance the final product is ever manufactured. The early communication of PMD brings different benefits to the retailers, such as better CM as shelf planning and SCM cost reductions, among others. However, before the final product is manufactured, the PMD is not yet final, since various factors are still subject to change, for example, the manufacturing location in the food segment, which could impact the allergens information, and therefore the artworks (label), of the products depending on other products being manufactured in the same location.

Generally, the PMD becomes final with the beginning of production. Currently, retail companies request final data sets from 8 to 6 weeks before the new products are shipped to their facilities. However, the PMD is requested by retail companies at different points in time, ranging between 8 weeks, to up to 26 weeks prior to the first shipment (represented in Figure 1.4 through a Gordian knot), bringing the request for the delivery of PMD further away from the point in time where the data is finalized, namely, with the physical production of the product.

Consequently, the separation of PMD from the point in time when it is finalized leads to two possible approaches to be taken by the manufacturers: (1) To send “dummy” PMD when it is requested, and communicate updates once the data has been finalized, which is a poor practice of Master Data Management (MDM), since it causes low data quality and generates additional data manipulation costs. Further, (2) the manufacturer can opt for holding the PMD until it is
final, but risking to miss the opportunity of selling its new products to retailers due to the given time windows, specially by seasonal articles, which have fixed time frames for products to be made available for the end consumer.

Currently, both manufacturers and retail companies have exercised flexibility towards data delivery and its reception. However, this flexibility has diverted both parties’ resources from their main operations to a hard bargaining based on the quality of the PMD. The lack of specific and clear business terms to deal with this situation across manufacturers and retailers are partly a cause for this. An illustration of the problematic described can be seen in Figure 1.4, further explanations to this visual aid can be found in the following paragraphs.

![Figure 1.4 - Problem Definition](image)

The upper part of Figure 1.1 describes a simplified version of the product development process. Under the product development stream, the “PLM – Product Development Process” arrow refers to the accumulative creation of PMD through time, for example, the specifications of the
product, design of the label or packaging, logistical information such as units of product per package or per pallet, among others.

In the center of Figure 1.4 a time line can be found, across this axis two lines intersect the time line, a vertical dotted line with the label “26 weeks” and a vertical solid line with the label “8-6 weeks”. The time buckets written on the labels represent the time pending until the “First Shipment”, which represents the dispatch of the first possible shipment to customer once the purchase order for the new product has been processed. The product delivery times in Germany are not considered since it lasts few days while the time buckets of the figure are weeks. Additionally, the solid vertical line crosses a box with the label "Final data set", which refers to the point in time when the PMD is finalized in alignment with the production process (red dot on the top), also the data quality dimensions used in the German market can be seen, nonetheless the names of the quality dimensions vary among authors, see sub-chapter 2.3.4. The apparent distance between both vertical lines and the distance to the “First shipment” point can be ignored since it is merely symbolic and it doesn’t reflect a uniform scale of time.

At the bottom of Figure 1.4 the flow called “Listing process / Commercialization” refers to the retailer process of listing and selling a new product throughout its channels (referred as omni-channel). The process flow under the green arrow depicts the sub-processes undertaken by the retailer to complete the “Listing Process”. The “1st Listing Meeting” represents when the manufacturer contacts the retailer to present the new product (a visual representation of the product, prototype or mockup, rarely the final product is ready at this stage). However, prior to this study, the sub-processes following the “1st Listing Meeting” were unknown to the manufacturers, therefore, question marks have been placed on the following processes. Additionally, the overall “Listing Process” was not generalized across retail companies.

1.3.2 Research Objectives

The main objective of this master’s thesis is to find a compromise between the manufacturers’ PMD disposition capabilities and the data requests from the retail companies, assuring the data
quality and the enablement of the retail companies’ omni-channel operations. GS1 Germany endeavors to create a data quality standard to moderate the frictions related to inter-company PMD communication, through clear and detailed responsibilities assignment, deadlines, metrics and further business terms. Hence, Mars Germany and Smart Data One have synergized efforts by supporting this study in order to facilitate the creation of the data quality standard in collaboration with GS1 Germany.

Furthermore, this study aims at creating knowledge on the described research problem and evaluate the feasibility of a recently implemented technical functionality as an enabler to solve the research problem (Sara Novak, 2015). The technical functionality is called Preliminary Trade Item (PTI), for which there are currently no instructions that cope with business rules and end-to-end processes across companies for its implementation. The deductive approach taken to deal with this challenge can be seen in Figure 1.5, this figure includes a question mark next to PTI, and this denotes that PTI’s feasibility is still to be tested.

**Figure 1.5 - Deductive Approach to the Research Problem**

Additional information needed to grasp the meaning from Figure 1.5 is the differences between product master data, product datasets and product attributes. Products in the FMCG industry can be described by approximately 500 attributes (1WorldSync, 2016b), however, not all attributes are mandatory or used by all product categories within this industry. For example,
the category for alcoholic beverages uses attributes that describe the percentage of alcohol, whereas food products have certain attributes that describe the allergens and nutrients, further examples of product attributes are the dimensions of the product, like width, height and depth, or even its color. A dataset comprises all attributes that describe a single new product, which means that if a manufacturer introduces two different new products at the same time, the product master data will contain two datasets, one for each product.

The principle behind PTI in this study, is that certain attributes of a product become final before the product is manufactured, and therefore, could be shared earlier with the retail partners, and then, once the product has been manufactured, the rest of the attributes can be sent to the retailers, completing the dataset for that given product. A cornerstone for this approach is to define, whether an incomplete dataset in an earlier point in time would enable the retail partners to start their operations whilst the rest of the attributes are finalized and shared at a later point. Further, if this approach is valid and the retail companies can initiate their operations with an incomplete dataset, then the next priority is to define, as stated in Figure 1.5, which product attributes are needed, and in which point in time during the listing process they are needed.

Based on the research objectives the main focus and research question for this master’s degree thesis is:

“**How can a manufacturer meet the required quality for its product master data on time to enable omni-channel commerce for its trading partners in the Fast-Moving Consumer Goods industry?**”

Thereby, this study considers the following Sub-Questions (SQ) to address the main Research Question (RQ):

1. **SQ-1: How can data quality be achieved for product master data?**
   This sub-question aims to define what product master data is, what data quality is, and how product master data of quality can be achieved. An objective approach to answer this sub-question is by conducting literature research. Consequently, an answer to this
question will provide a reference and ground knowledge on what the requirements on data quality from the retail companies could be.

(2) **SQ-2: What activities are performed by the retailers in order to list a new product and what is the duration of those activities?**
Once the terminology and goals of data management have been stated, this sub-question clarifies what activities determine the product master data requirements during the listing process. Additionally, the duration of the single activities enables the retroactive calculation of the point in time when the product attributes are used. A method to collect unbiased information is to gather it directly from the retail companies, obtaining first-hand information (primary data).

(3) **SQ-3: What product attributes are required for each of the activities or processes involved in listing a new product?**
This sub-question aids at aligning the data capabilities from the manufacturers with the requirements of the retailers by deepening the granularity level of the data to an attribute level. This Sub-question complements the time factor obtained in SQ-2. The information to address this sub-research question can be gathered through the same method as SQ-2.

(4) **SQ-4: What data requirements differ per channel (online & offline)?**
In order to cover an omni-channel solution, different channels should be covered and their different impact on the product master data requirements. This information should be gathered at the same time as SQ-2 and SQ-3 to obtain this complementary perspective per activity at an attribute level.

(5) **SQ-5: What is the current status of data quality for product master data sent by the manufacturers?**
This sub-question provides the picture of the current performance in data quality serving two purposes, to provide a point of comparison after quality measures are implemented and to create awareness of how critical the current situation might be. An answer to this sub-question can be sought through literature research and complemented with direct information from retail companies.
1.4 Research Scope and Delimitations

This sub-chapter frames the context of the study within its validity areas. Additionally, this section explains the elements of the research area and research problem that have been deliberately left aside. Each of this actions might lead to limited generalizability for the results, however, this study has been designed to be valid within a given framework but useful as basis for further research to expand its applicability scope.

1.4.1 Research Scope

This master's degree thesis considers the supply chain interactions between fast-moving consumer goods manufacturers and retail companies, from the conception of a new product to the delivery of the finalized product at the retailers’ designated premises. Moreover, this study contemplates an aspect of the supply chain that creates interactions and frictions between the involved stakeholders before the physical product has ever been created, the scope of this thesis considers the product information that has to be exchanged in order to:

A. Awake the interest of retail companies to sell the new product in their stores,
B. Allow optimal planning of transports, storage and shelf-layouts, among other activities,
C. Conduct profitability analyses,
D. Generate demand for the products,
E. Enable omni-channel commerce: Online commerce, brick and mortar commerce or combinations of both, and,
F. Assure legal compliance.

Product information and its exchange between manufacturers and retail companies imply numerous exceptions based on product type and exchange conditions (such as platform and other technological factors). Therefore, delimitations are needed to frame the context of this research study.
1.4.2 Research Delimitations

This study focuses in the fast-moving consumer goods sold in the German retail industry. Thus, two delimitations have been made, firstly, other types of consumer goods are not being considered. This selection of consumer goods impacts on the strategies and working methods that tailor the “fast-moving” SCM and the interactions between the involved entities in the supply chain. Secondly, the study focuses in the German retail industry. Different industry standards for data communication between commerce partners, legal obligations and transport lead times among other factors, frame the study within the German market.

Additionally, this study is conducted from the perspective of a manufacturer in the fast-moving consumer goods industry with a customer-centric approach. Therefore, as stated in sub-chapter 1.3.2, under the research objectives, understanding the data requirements from the retail companies is a cornerstone to build knowledge within the research problem area. Nonetheless, this approach limits the input regarding the product innovation processes for this thesis to only one manufacturer in the industry. Yet, this does not compromise the validity of the results since the focus is placed on the requirements from retail companies, rather than the capabilities of the manufacturers.

Also, this work is oriented to new product introductions only. Nonetheless, it is valid for seasonal, standard, promotional or bundle products, essentially, for all products that require the allocation of a new GTIN code. For information about GTIN codes, see sub-chapter 2.3.2. The reasoning behind this delimitation, is that the product master data of recurrent products has been previously shared from manufacturers to retail companies, therefore, skipping the listing process, which is a cornerstone of this study.

All participating retail companies in this study synchronize, partially or entirely, product master data through a GS1 certified Global Data Synchronization Network (GDSN). However, not all manufacturers synchronize their data with the retail companies through a GDSN platform, since
the size of their portfolio or the number of customers that they serve does not justify the investment, hence, these manufacturers use different alternatives. Nonetheless, the solution to be tested as part of this study, the Preliminary Trade Item (PTI), is a technical functionality enabled in a GDSN platform, therefore, the participating companies should be active in this platform. Nevertheless, PTI could be enabled for various manufactures across different platforms in the future. For more information regarding data pools and GDSN, see sub-chapter 2.3.2.

The thesis is public and therefore can be read without limitations on the audience, however, the names of the companies that took part in the study cases will remain anonymous and the results of the interviews were generalized and anonymized. This is part of a non-disclosure agreement defined between the participating, sponsoring and collaborating organizations.

Lastly, this study analyzes the ability of a specific technical feature named Preliminary Trade Item to solve the research problem. This tactic delimits the proposed results to one solution, however, this technical feature has been prioritized by experts in the industry as an enabler for data quality in the context of new product introductions and their data communication across companies. Preliminary Trade Item was designed to alleviate the product data communication between enterprises, which is the research problem of this study as described in sub-chapter 1.3.1, thus, the focus on this solution for feasibility.

1.5 Structure of the Study

This sub-chapter provides guidance about the structure of the thesis to enable the reader to find specific content and to smooth the analysis of the chapters. This study has been divided into six chapters with a logical flow that aggregate knowledge in order to fulfill the research objectives. Thereby, the design of this thesis has been developed considering the best approach to answer the research questions. A visualization of the thesis structure can be seen in Figure 1.6.
The first chapter introduces the study providing rich information about the research problem, objectives and the context in which the study was conducted. The second chapter analyses and presents a compilation of current knowledge in the research area, presenting and linking the theoretical disciplines. The third chapter describes the methodology selected to approach the research problem combining different sources of knowledge. The fourth chapter presents the primary data obtained from the empirical study conducted as part of the research strategy. The fifth chapter analyses the results obtained in the previous chapter and frame them within the research problem. Lastly, the sixth chapter utilizes the aggregated knowledge from the previous chapters to answer the research questions and present the managerial implications, recommendations, limitations and further research areas.
2. THEORETICAL BACKGROUND

The purpose of this chapter is to show the state of knowledge in the different disciplines that encompass this study and build the intellectual basis on which the empirical study is settled. The major disciplines that frame the master’s degree thesis are Supply Chain Management (SCM), Master Data Management (MDM) and Product Life-cycle Management (PLM). Thereof, deeper specialized domains of research are analyzed to assess diverse perspectives of experts in the various fields of the industry, for further detail please refer to Figure 2.1.

Figure 2.1 - Focus Area of Research

While reviewing the focus area of research in Figure 2.1 it should be noticed that there are different perspectives among authors regarding the fitness of some specialized domains within their framing major disciplines in MDM and SCM. Namely, MDM and Data Quality Management (DQM) can be considered independent or combined spheres of research (Power, 2016; Mosley
et al., 2010). Also, other authors refer to DQM as “Corporate Data Quality” when data quality is focused entirely on master data (Otto and Österle, 2015). Nonetheless, these differences will be discussed deeper in sub-chapter 2.3.3.

Similarly, SCM can be considered a uniform discipline with applications across industries (Farfan, 2016), however, some authors find sufficient elements to consider the Retail SCM a deeper specialized domain of SCM (HCL, 2016; Infor, 2009; Zentes et al., 2011; Günther and Meyr, 2009; Scott et al., 2011). Additionally, others authors state that traditional SCM should be revolutionized to serve omni-channel retail (Chaturvedi et al., 2016; Hatamura and Roussos, 2006), see more information about SCM in sub-chapter 2.1.1. Moreover, the remaining sphere of research “PLM” showed no discrepancies in the theoretical boundaries used for this study, hence no remarks are presented.

The theoretical background was consolidated based on meaningful and current literature out of the different disciplines and domains of research for this study. The various sources utilized to gather knowledge included peer-reviewed literature databases such as Scopus and ScienceDirect from Elsevier, and Web of Science from Thomson Reuters. Further, the Finnish research platforms Finna and Nelli were used, which together offer material from museums, libraries, archives, databases, journals, books and dictionaries (Finna, 2016; Nelli, 2013). Additionally, content from university courses, conferences, white papers, magazines, specialized books and both, official and corporate websites were considered.

The research strategy was based in keywords resulting from deepening iterative literature review in the context of the research problem. The main keywords and strings used for research were “Master Data”, “Data Quality”, “Supply Chain Management”, “Omni-channel Commerce”, “Retail Supply Chain”, “Fast-Moving Consumer Goods Industry”, “Efficient Consumer Response”, “Supply Chain Management AND Product Life-cycle Management” and “Product Life-cycle Management AND Master Data Management”. The criteria considered for the selection of the content was based on the applicability of the knowledge to the research problem, relevancy and validity of the information, diversity in authors perspectives and projection of trends in the framing disciplines and industry.
2.1 Supply Chain Management

Since the era of the industrial revolution firms have struggled to source and supply their materials and products. Initially, obtaining the necessary resources and distributing the final goods was an effortless task, since suppliers and customers were in a close proximity to the company. However, the ability of completing these activities diminishes as the locations of the suppliers, manufacturers and customers scatter around the globe (Ross, 2015). Nowadays, in very seldom circumstances, the sourcing, production and consumption of any product occurs in the same location (Caplice, 2016a).

There are many different definitions of SCM and logistics, generally referring to both as being conceptually equal, yet, being logistics an older term (Marien, Edward J., 2003; Henkoff, 1994). However, there are different streams of literature that define SCM as the collective activities that increase customer value through collaborative partnerships, based in improving the competitive advantage of the holistic network. In addition, logistics management is defined as the compound of tasks that steward inventories, warehousing, and transportation resources in an efficient and cost-effective manner. Logistics allow firms to meet the daily product and service demands from their supply chains (Ross, 2015; Council of Supply Chain Management Professionals, 2016).

Looking ahead, several factors or megatrends will dictate the performance of SCM such as customer centricity and the servitization of supply chains (and products), were the pre- and post-sales customer service becomes a competitive advantage for companies (Sengupta, 2013; Ross, 2015). Likewise, new ways of working as “buy from anywhere, ship from anywhere” by leveraging multi-channel and omni-channel fulfillment (Friedlein, 2016; Ross, 2015). Additionally, different factors such as climate change and geopolitical instability increase the risk of supply chain disruptions. Nonetheless, technological developments like digitalization, drones, big data, autonomous driving, virtual reality and artificial intelligence will redefine the way SCM operates (The British Standards Institution, 2016; Felgendreher, 2016).
2.1.1 Supply Chain in the Retail Industry

In this study, the FMCG industry and the retail industry have been referred as separated industries, yet, they are intrinsically correlated (Thain, 2014; Zentes et al., 2011). The FMCG industry is comprised of manufacturers producing Consumer Packaged Goods (CPG), and as described in sub-chapter 1.3.1, FMCG products are sold quickly in high volumes. The FMCG companies are at the forefront of new retail developments, emerging markets, online commerce and online engagement, also, these firms drive the world’s advertising industry (Thain, 2014). Alongside, retail is one of the biggest industries in the world, being closer to the end-consumers than companies in the FMCG industry. Retail companies are no longer exclusively intermediaries for merchandise between manufacturers and customers, they have taken an active and key role in marketing, running distribution channels and SCM (Zentes et al., 2011).

In the retail industry logistics and SCM are divided by the scope of their tasks. Logistics comprehends the management of storage facilities, inventory, transportation and recycling (Fernie and Sparks, 2009). In turn, retail SCM involves several parties as CPG manufacturers, suppliers, wholesalers, retailers, third-party service providers and customers (Chopra and Meindl, 2010). The Retail SCM operates through a continuous flow of products, information and finance (Zentes et al., 2011). In the world of SCM, the key elements that differentiates retail SCM are:

- Volume of product movement,
- Fast moving nature of the products in the retail industry (HCL, 2016).
- Additionally, the particular challenges that the retail SCM facea include the ever growing number of suppliers, logistics providers, channels and products, increasing fuel costs and tight labor markets, as well as changing consumer preferences (Infor, 2009).

2.1.2 Enhancement of Collaboration

In order to avoid inefficiencies in the supply chain as the bullwhip effect among other causes for inefficiencies, the concept of Efficient Consumer Response (ECR) has been developed, it
prevents planning isolation in the supply chain through cooperation and collaboration across the parties involved. ECR aids manufacturers and retailers to fulfil consumer needs through various strategies and practices. Conceptually, ECR shifts push-oriented supply chain planning towards pull-oriented management, actioned by orders and just-in-time tactics, hence, this is a demand driving approach (Kracklauer et al., 2004).

ECR exploits three mains areas of collaboration among parties With the aim of minimizing isolation in the supply chain (Global Scorecard, 2011):

- **Responsive supply**: Aligns the distribution of products in the retailer premises with the production of the goods at the manufacturers’ site. The replenishment is triggered by real consumption aiming to a high service level with the minimum inventory costs.
- **Integrated demand-driven supply**: This area focuses one step further in the upward supply chain by synchronizing the procurement of materials with the requirements to meet the demand of production, which in turn aligns with the real time sales from the retail companies.
- **Operational excellence**: Excellence is achieved by establishing industry relevant standards and methods to enable same-language communication between enterprises and thus increase efficiency and the reliability of operations, a cornerstone for this goal is the Electronic Data Interchange (EDI) and other technologies.

Examples of concepts developed and applied in ECR are Quick Response (QR), which is a type of continuous replenishment model applied for merchandise. QR aims at reducing time in product flows in the supply chain, especially for products with high variability in demand and short cycles (Coyle et al., 2003).

Functionally, ECR depends on technical capabilities to perform EDI in real-time. A major part of these capabilities are achieved through standardization and automation. Specific aspects that have to be considered for ECR are (Hertel et al., 2011):

- **Automated identification systems**, bar codes or Radio Frequency Identification (RFID). The identification results from coding systems as the Global Trade Item Numbers
(GTIN), which is mapped to an item or Single Key Unit (SKU). Additionally, the different parties in the supply chain can be identified via a Global Location Number (GLN).

- Communication standards, including the standards defined by the organization GS1. This organization standardizes formats for data to assure inter-company communication and the correct interpretation by the different IT systems, for example the EANCOM or the GS1-XML, which are flexible message format.
- Master data: This is the main data used in information processing. Customer and supplier master data tend to be independent within a firm, however, product master data is exchanged between companies, and therefore, standardization is a prime requirement to support operational processes.

### 2.1.3 Integrated Channel Experience

The past two decades have seen significant changes in the retail arena, where different markets have harnessed the capabilities of digitalization through the online channel, yet, in some markets, the online channel has outweighed the traditional channels, like in the travel industry, where customers now skip intermediaries. However, there are other markets like the FMCG industry where the change in commerce has not been as disruptive as in the travel industry, nonetheless, FMCG retail companies have reshaped their retail marketing mix strategy, including channel selection and channel management to adapt their business models (Verhoef et al., 2015).

The variety of channels affect both types of companies, purely online retailers and traditional brick and mortar retailers. Both should decide if they incur in multi-channel practices, considering the challenges of managing customers online and offline as well as integrating their business through all the channels (Neslin et al., 2006). Multi-channel retailers offer products in local stores as well as in online sites. Serving customers in different channels implies deciding on separate or joint strategies, such as pricing. Nonetheless, depending on the channel, the cost structures and processes could vary significantly (Trenz, 2015).
Nowadays, the channels continue to evolve with the emergence of mobile devices, tablets, and the social media, which in turn, develop interconnectivity with electronic and traditional retailing. Remarkable research on the transition from multi-channel and omni-channel commerce has been conducted by Verhoef, Kannan, and Inman (2015) as they describe different streams of literature from the year 2006 to the year 2014. They define omni-channel management as an inclusive and holistic integration of the various channels and customer touchpoints. Thereby, creating a seamless and optimized customer experience. Omni-channel includes different advantages over traditional multi-channel that can be seen in Table 2.1.

### Table 2.1 - Multi-channel versus Omni-channel Management

<table>
<thead>
<tr>
<th></th>
<th>Multi-channel management</th>
<th>Omni-channel management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel focus</td>
<td>Interactive channels only</td>
<td>Interactive and mass-communication channels</td>
</tr>
<tr>
<td>Channel scope</td>
<td>Retail channels: store, online website, and direct marketing (catalog)</td>
<td>Retail channels: store, online website, and direct marketing, mobile channels (i.e., smartphones, tablets, apps), social media Customer Touchpoints (incl. mass communication channels: TV, Radio, etc.)</td>
</tr>
<tr>
<td>Separation of channels</td>
<td>Separate channels with no overlap</td>
<td>Integrated channels providing seamless retail experiences</td>
</tr>
<tr>
<td>Brand versus channel customer relationship focus</td>
<td>Customer – Retail channel focus</td>
<td>Customer – Retail channel – Brand focus</td>
</tr>
<tr>
<td>Channel management</td>
<td>Per channel</td>
<td>Cross-channel objectives (i.e., overall retail customer experience, total sales over channels)</td>
</tr>
<tr>
<td>Objectives</td>
<td>Channel objectives (i.e., sales)</td>
<td></td>
</tr>
</tbody>
</table>

**Source:** Adapted from Verhoef et al., 2015.

Furthermore, other authors emphasize the relevance of the customer experience in omni-channel. For example, a mobile app should correspond the responsive design of the company’s website, which in turn should reflect the aesthetics and atmosphere of the physical store (Cloudtags, 2016). Additionally, the different platforms per channel should be integrated and
enable the customer service agents and sales representatives to know about previous interactions from customers with the company, also, to enable the customer to find real-time inventories in stores to pick the products physically later or have them delivered to their preferred addresses (Rouse, 2014).

2.1.4 Added Value in Enterprise Management

Value adding activities drive entrepreneurship and are tightly connected to a product or service. In Figure 2.2 four different schemes for value creation in a company can be seen: (1) Product Life-cycle Management (PLM), (2) Finance management, (3) Production and logistics, and (4) SCM. Each of these schemes has a percentage number next to it, representing the estimated value added percentage of these activities to enterprise management as an average across industries (Ivanov and Sokolov, 2010).

Figure 2.2 - Main Elements of Enterprise Management

Source: Ivanov and Sokolov, 2010.
As it can be seen, different functional departments as marketing, procurement, sales or Research and Development (R&D) are not listed, however, they have a direct impact on these schemes. For example, marketing and sales impact SCM through the accuracy of a sales forecast (Rudolph et al., 2007). As depicted in Figure 2.2, (2) “Finance management” and (4) “Supply chain management” represent the most value-adding activities in enterprise management, balancing supply with demand and controlling direct and indirect financial flows respectively (Ivanov and Sokolov, 2010).

Lastly, (1) “Product life cycle management” and (3) “Production and logistics management and engineering” account for 25% and 15% of added value respectively. Regarding PLM, its contribution to the company growth varies significantly depending on the industries and branches. However, the interactions that PLM and SCM have with suppliers and customers through material, informational and financial flows, increase the effectiveness and efficiency of all schemes (Ivanov and Sokolov, 2010).

### 2.2 Management of Product Cycles

Product Life-cycle Management (PLM) has different definitions, focusing on different benefits that PLM offers to enterprises, however, there are two major streams that conceptualize PLM as a business strategy or a system. The first stream defines PLM as a business strategy that enables the management of a product in the most efficient way. This happens through the product’s complete lifecycle which encompasses people, processes and technology. Thereby, PLM shortens the time-to-market, increase product revenues, reduce product-related costs, maximize the value of the product portfolio and the current and future value of the product for customers and shareholders (Vadoudi et al., 2014; Stark, 2015).

The second stream of literature focuses on the core activities of PLM, the creation, preservation and storage of product’s and firm’s data. Additionally, this stream of PLM ensures the communication of information to the extended company in an efficient way during the complete life-cycle, and also facilitates the development of new products. Simply put, the intellectual work
from previous products should remain available and exploitable converting the data into company assets (Kiritsis, 2011; Sääksvuori and Immonen, 2008).

While referring to life-cycle planning, all stages of a product should be considered and documented within its expected life, from its early design to the disposal of the product, hence, Kiritsis proposes three major phases that categorize the product information creation according to its maturity and usage, see Table 2.2 (Kiritsis, 2011; Terzi et al., 2010). Moreover, a product’s life-cycle can also be analyzed by its market penetration as described by the model for diffusion of innovations (Rogers, 2003), however, this method focuses on the adoption of a product rather than the information about a product and its environment, therefore, this model is not considered.

Table 2.2 - Product Lifecycle Phases

<table>
<thead>
<tr>
<th>Phase</th>
<th>Functions performed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning of life (BOL)</td>
<td>Conceptualization, definition and realization</td>
</tr>
<tr>
<td>Middle of life (MOL)</td>
<td>Delivery, usage, service and maintenance</td>
</tr>
<tr>
<td>End of life (EOL)</td>
<td>Reuse of the product, reuse of components, or disposal.</td>
</tr>
</tbody>
</table>

Source: Adapted from Kiritsis, 2011.

PLM was originally used in the 1980s as Product Data Management (PDM), before it advanced to a lifecycle approach by facilitating engineers to store their design files or Computer Aided Designs (CAD), and also to identify relationships between components and assemblies. However, goods’ manufacturers tried to find new monetary income opportunities, especially in the after-market services. The idea behind this alternate business approach was to offer value adding services throughout the lifecycle of the products, since depending on the industry products could remain active up to 30 years, and hence, produce steady incomes from maintenance contracts. Conversely, in different industries like the FMCG, lifecycles are shortening. Moreover, PLM increases product development’s efficiency, and thereby it reduces the overall time-to-market via a seamless innovation process (Sääksvuori and Immonen, 2008).
An extension of PLM on product information includes closed-loop information streams across the lifecycle and different parties (Kiritsis, 2011). Further research disciplines as SCM and MDM benefit from the best practices in PLM and share key concepts that fuel each discipline’s targets. For example, optimal SCM practices involve information sharing between customers and suppliers about product delivery, which is an activity considered as well within the product’s lifecycle. Likewise, in MDM, product information is a business object, which in both disciplines SCM and PLM is shared with the customers (Otto et al., 2012; Zentes et al., 2011).

Lastly, the future of PLM comprehends innovative trends towards sustainability and further closed loops lifecycle management (Vadoudi et al., 2014). These trends include the integration and documentation of mechanical and electronic components, with different versions of software and operating systems for products (Shilovitsky, 2016). Additionally, the re-invention of the front-end or user interfaces of PLM systems act as innovation platforms for companies to enable a systematic approach to capture, select and invest in the right projects (Vagdati, 2015).

### 2.2.1 Beginning of Life

This stage manages the activities related to product design and manufacturing. Product design activities comprehend product’s design, process’ design and plant’s design, whereas manufacturing include the processes of production and internal logistics (Terzi et al., 2007; Kiritsis et al., 2008). Design and manufacturing are fundamentally different, design involves a repetitive and looped intellectual activity measured by efficacy, where designers and engineers take part, whilst manufacturing is a recursive transactional-based action measured by efficiency with no further intellectual inputs (Terzi et al., 2010).

During this stage product design data is generated and shared across the involved parties, namely, further designers and engineers to assure efficient manufacturing (Terzi et al., 2010). However, the full potential of PLM is reached when other departments involved in the new product development as marketing and production collaborate with R&D, thereby shortening
the time-to-market (Oh et al., 2015). Different types of collaboration started to be exploited in PLM involving customers in the new product design with the introduction of Collaborative Product Definition (CPD), enabling concurrent engineering and increasing efficiency (Terzi et al., 2010). In the FMCG and retail industries, product category specialist from retail companies gather with multidisciplinary teams from manufacturing companies to discuss about new trends in the market, customers’ preferences, materials and packaging designs to smooth the logistics between both companies and improve the product adoption.

The development of a product is an ever evolving process driven by different trends in technology, markets and society. Technological leaps can be seen through innovative products, markets’ impact with changing product requirements, and lastly, societal influence based on rising concerns for environmental footprint, sustainability, health and safety (Persson, 2016). However, through direct collaboration with customers the product uncertainty can be reduced and risks on all three drivers can be mitigated.

### 2.2.2 Middle of Life

The intermediate phase of PLM considers several stages of the product, from the point where it is sent to a customer until it offers no more benefits to the customer. The transition between BOL and MOL occurs once the product has been finalized, internally stored and made available for shipment, which might include transportation suppliers and after-sales personnel suppliers (Ciceri et al., 2009). Therefore, the activities that are comprehended by MOL are distribution (external logistics), product’s use and support or maintenance. The related product information that can be gathered during this phase includes distribution routes, usage environment, and depending on the industry, malfunctions and maintenance (preventive and corrective) to keep an accurate report about the performance and conditions of the product (Terzi et al., 2010).

The stakeholders in this phase include the customer directly and additional parties, which might involve product users, end-consumers, service providers, after-sales assistants, maintenance specialist and logistics providers (Terzi et al., 2010; Ciceri et al., 2009). The entities involved in
the MOL phase vary significantly among industries, for example, in the retail industry the delivery might flow through several instances, including externally or internally managed distribution centers, then to be further transported to a retail store, however, until this point, the product has not yet reached the usage stage by an end-customer, see Figure 2.3.

![Figure 2.3 - Delivery of Goods in the Retail Industry](image)

**Source:** Zentes et al., 2011.

The information gathered on product performance can be used to improve the product design, features or materials in the BOL phase, yet, for further versions of the product (Ciceri *et al.*, 2009). However, most of the information from the usage stage is not captured or transmitted back to the manufacturing company, interrupting the closed-loop product information flow, due to the lack of PLM systems utilization in this phase or the lack of affordable and appropriate technologies that facilitate the information flows (Kiritsis *et al.*, 2008; Vadoudi *et al.*, 2014; Verdugo Cedeño, 2016). Furthermore, companies out of the FMCG industry, namely with long-lasting products, pursue business opportunities throughout the complete lifecycle of the products via after-sales services, hence, collecting the information of the performance of the products during the delivery, use and maintenance stages has gained significant relevance (Sääksvuori and Immonen, 2008).
2.2.3 End of Life

This stage involves the activities of retiring a product which is no longer useful for the customer, yet, depending on the industry, the components or the complete product might still have value for the manufacturing company. Additionally, the product might contain elements that require special disposal processes as with batteries due to their chemical compounds. This phase starts when the product no longer fulfills its purpose, hence it is retired or recollected from the customer (reverse logistic) and the different steps that can occur for the reprocessing of this item, see Table 2.2 (Terzi et al., 2010). The EOL stage, has gained importance as society becomes aware of environmental impact, since even products that offer no value to the manufacturers through their components or re-use possibilities, could have a negative ecological impact if not disposed properly (Niemann et al., 2009). Likewise, fast moving consumer goods producers are influenced to develop environmental friendly packaging for reducing the environmental impact after the consumption of the products.

Product information such as components, materials, manufacturers, time of usage, as well as further information useful for its re-use are the most important in this phase (Terzi et al., 2010). Also, increasing technologies in sensors make able to track the wear of single components and provide performance information to manufacturers (Persson, 2016). This information can aid companies at predicting product’s expected life and increase the overall Return of Investment (ROI) for customers by utilizing this information to improve the selection of materials and to enhance the design, thus, extending the product life.

2.2.4 Integration Architecture of PLM and SCM

In the previous sub-chapters, the conceptual alignment of SCM and PLM have been described through the key benefits that collaboration and information sharing offer for manufacturers and customers. Yet, both of these disciplines are supported by people, resources, processes and information (Terzi et al., 2010), thereof, systems and platforms enable and empower people to conduct their operations (Sääksvuori and Immonen, 2008; Zentes et al., 2011; Hefu et al., 2016).
A recent system's model for new product developments proposes a robust architecture between SCM and PLM that allow enterprises to shorten the time-to-market despite product uncertainty. According to this model, firstly, a new product project is created in the PLM system, which controls the schedule of the development. Then, through project management, project schedule monitoring and product development controlling, PLM activities are integrated through MDM with SCM, see Figure 2.4 (Oh et al., 2015).

**Figure 2.4 - PLM, MDM and SCM Architecture Integration**

**Source:** Oh et al., 2015.
The intermediate step between PLM and SCM is conducted within the discipline of Master Data Management (MDM), which acting as an internal and external data hub, consolidates the product information or Product Master Data (PMD) following industry standards that assure the correct interpretation of the data across companies. The integration of the PMD takes place in a module of the PLM system called Product Integration Information (PII), or more widely known in the industry as a Product Information Management (PIM) system. Data created from both, the SCM system and PLM system, is consolidated through the PIM system and then can be used internally or shared externally (Oh et al., 2015).

Although the majority of the elements from Figure 2.4 are well depicted, this model focus on the upward stakeholders in the supply chain, which does not cover the interfaces for PMD communication between manufacturer and customer through PII or PIM systems, as it occurs in the FMCG and retail industry. Additionally, a reciprocal stream of information has not been considered between manufacturers and the respective suppliers and partners that after receiving the materials supply plan, would confirm to the manufacturer their ability to fulfil the orders placed.

### 2.3 Management of Master Data

Master Data Management is the discipline that combines Information Technology (IT) and business practices to cover activities such as creating, changing and deleting a master data class, attribute or object (Smith and McKeen, 2008). The goal of MDM is to achieve data quality for the master data, and therefore enable all business processes (Loshin, 2009). Examples of MDM activities include modeling, maintaining and provisioning the master data (Otto et al., 2012). Additionally, as seen in Figure 2.4, MDM performs as a data hub where the master data is integrated from different systems as PLM and SCM, in order to be re-distributed within the firm or externally to suppliers or customers (Oh et al., 2015).
Master data defines all relevant corporate entities on which a firm relies to conduct business. Examples of these entities are customers, products or even employees (Smith and McKeen, 2008). According to Bosch, entities or classes of master data include customer hierarchies, materials, suppliers, charts of accounts and organizational units (Hatz, 2008). Essentially, master data consists of three hierarchical elements: Master data class, master data attributes and master data objects (Loshin, 2009), see for further information Table 2.3.

Table 2.3 - Master Data Elements

<table>
<thead>
<tr>
<th>Master Data Elements</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master Data Class</td>
<td>Business objects or entities. For example, customers, products or suppliers.</td>
</tr>
<tr>
<td>Master Data Attributes</td>
<td>Characteristics that describe instances of a master data class. For example, color, features, price or dimensions.</td>
</tr>
<tr>
<td>Master Data Object</td>
<td>An instance of a master data class. Represents a concise unique product, elaborated at a certain plant at a certain point in time.</td>
</tr>
</tbody>
</table>

Source: Adapted from (Otto et al., 2012)

In companies transactions are referenced to master data. However, the definitions of the master data entities should be consistent across the different functional departments, otherwise the master data is valid only locally, for example, to one business division. Additionally, master data is referred as “static data” because it is not subject to modifications frequently and increases its volume at a slower pace compared to transaction or inventory data (Otto and Österle, 2015). For example, a specific product could be referenced in hundreds of thousands of purchase orders or invoices. The amount of transactions increased with the same number of orders or invoices, whereas the product master data did not increase, because it was the same product in all transactions (GS1, 2016g).
2.3.1 Product Information

As described in the previous sub-chapter, Product Master Data (PDM) is a master data class. Yet, this class of master data is highly complex since it can use approximately 500 master data attributes depending on the product category to describe a single product (1WorldSync, 2016b) and the nature of this master data class is to be shared internally or externally. Additionally, as covered in sub-chapter 2.2, PLM gathers and manages product information, however not all product information becomes product master data. In Table 2.4 a sample of the different types of product information that PLM stores and manages across the lifecycle is shown.

Table 2.4 - Examples of PLM Product Information

<table>
<thead>
<tr>
<th>Examples of PLM's Product Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer’s feedback</td>
</tr>
<tr>
<td>Maintenance</td>
</tr>
<tr>
<td>Procurement</td>
</tr>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>Procedures</td>
</tr>
<tr>
<td>Drawings</td>
</tr>
<tr>
<td>Production</td>
</tr>
<tr>
<td>Prototypes</td>
</tr>
<tr>
<td>Processes</td>
</tr>
<tr>
<td>Concept designers</td>
</tr>
<tr>
<td>Marketing specifications</td>
</tr>
<tr>
<td>After sales service</td>
</tr>
</tbody>
</table>

Source: Adapted from Terzi *et al.*, 2010.

Furthermore, Figure 2.5 shows different examples of PMD. As it can be observed, some information or product attributes seem similar with Table 2.4. Nonetheless, both approaches and categorizations are valid since each firm tailors their MDM and PLM strategies independently, as it best suits the needs of the company (Otto and Österle, 2015).
However, each industry and product category have different obligations to share product information in order to trade their goods. Each country or politico-economic union of countries specify their own regulations on information to be shared to customers and/or end-consumers. An example of this is the Food Information Regulation (FIR): (EU) No 1169/2011, for the European Union. This regulation obligates all food product manufacturers to deliver specific information about their products, with the aim of offering to end-consumers the same information in online commerce as they would find it in physical stores (Van Sambeck, 2016). Besides country legal regulations, international organizations and their country branches as GS1, aid at defining standards to assure the correct interpretation of the data across enterprises (GS1, 2011).

The product category dictates for standardized data pools which master data attributes must be exchanged in order to do commerce. Therefore, the Global Product Classification (GPC)
code defines through a code, the taxonomy of all products (also known as trade items). GPC classifies products in groups depending on their essential characteristics (GS1, 2016c). For example, products in the GPC “Alcoholic Beverages” would use different attributes to be described as products in the categories “Clothing Accessories” or “Cheese/Cheese Substitutes”. GPC hierarchical structure can be seen in Figure 2.6, after the last level “Brick”, a unique set of attributes is specified. For further visualization aid towards GPCs and its interlinking with a product and attributes, see Appendix - 1: GTIN - GPC - Attributes - Values.

![Figure 2.6 - GPC Hierarchical Structure](image)

**Source:** GS1, 2016e.

Lastly, an important attribute in the inter-company PMD exchange which is not related to the product, but rather to the identification of the issuing organization is the Global Location Number (GLN). The GLN can also specify a location, for example, a warehouse or berth in a port, a firm’s legal and functional entities performing business transaction as a buyer, seller or carrier. Also this attribute can be encoded into an Electronic Product Code (EPC) or RFID (GS1, 2015a).
2.3.2 Inter-company Data Synchronization

Data synchronization enables cross-company exchange of data. Technically, there are different types of structures for master data synchronization between manufacturers and retail companies, where the structures are differentiated by the storage of the data, centralized or decentralized. Nonetheless, these differences impact the reach of the synchronization (Van Sambeck, 2016). Decentralized data keeping works with independent data silos, where each company keeps its own data and communicates it upon agreement, however, in a one-to-one approach. Conversely, centralized synchronization uses a central data pool as data hub, where all parties share their data and can retrieve it, in a one-to-many approach (Bussler and Sprague, 2002). Additionally, there is a hybrid synchronization form, which combines both types of data keeping. Hybrid synchronization manages centrally a registry with reference data (also known as metadata), that provide indications on how to identify the master data within each company’s data silo (Loshin, 2009).

In the practice, the solutions are divided by their one-to-one or one-to-many approach, which are supported by bilateral synchronization tools like individual web-portals (decentralized synchronization) or data pools (centralized and / or hybrid-centralized synchronization) respectively. A data pool utilizes the data hub principle to hold the data and distribute it to multiple parties, and additionally, a solution called Global Data Synchronization Network (GDSN), follows the same principle, yet, using multiple data pools as data hubs and synchronizing them with one another (Schemm, 2009). Figure 2.7 depicts the bilateral synchronization and the role of GDSN and data pools in the multilateral master data synchronization.
Master data pools are solutions offered by third party organizations which facilitate data communication through electronic standardized services between manufacturers (referred as suppliers as in Figure 2.7) and retailers (Schemm, 2009). 1WorldSync offers the largest product master data solution via a Global Data Synchronization Network (GDSN). This platform enables trading partners worldwide to share their product master data in a standardized manner. GSDN is a standard from GS1 and is implemented by 1WorldSync (1WorldSync, 2016a).

Figure 2.8 describes the process of data synchronization. The first step is performed by the seller, which in the context of this study is the role of the manufacturer, who registers the data of a product to the GS1 Global Registry, obtaining a Global Trade Item Number (GTIN) for that product. Then, the manufacturer uploads the product master data to a data pool using the GTIN as reference key. Thereafter, the retailer subscribes to that newly created GTIN. Once the subscription to a target market or GLN has taken place, the retail company will be able to receive all the product data (and its updates) published by the manufacturer to his target market.
retailers for defined product categories. Lastly, the retail company issues a confirmation of reception, closing the process flow.

![GDSN Data Communication Process](image)

**Figure 2.8 - GDSN Data Communication Process**

**Source:** GS1, 2016d.

GTINs aid to uniquely identify all products in across companies, also known as trade items in the retail industry. GS1 defines GTINs as “products or services that are priced, ordered or invoiced at any point in the supply chain”, they can be used to identify products regardless the packaging level or hierarchy, from a pallet to the smallest consumer unit. Information such as the expiry date and similar information can be identified via the GTIN code. This code is worldwide recognizable and can be encoded as a barcode or an EPC / RFID tag (GS1, 2015b).

### 2.3.3 Management of Data Quality

A compilation of research on Data Quality Management (DQM) has found significant variations in the discipline, from defining data quality dimensions in order to measure the quality of the
data, to practical industry applications or academic theoretical backgrounds. However, the DQM discipline has been consistent in defining what data quality is: Data’s fitness for use. This is a subjective definition that varies depending on the purpose or context of the operations (Otto et al., 2012). Essentially, DQM aims at continuously enhancing the quality of the data (Batini and Scannapieca, 2006). Hence, defining, measuring, analyzing and improving data quality (English, 1999; Eppler and Helfert, 2004).

As stated by the definition of Master Data Management, data quality is one of its goals. Therefore, DQM has been considered by some authors as one function of MDM (Dreibelbis, 2008). However, this doesn’t cover in full the activities that DQM entail, since this approach is restricted to the reactive functions of DQM. Besides, there are further practices of DQM that influence the way MDM is structured.

Figure 2.9 - MDM and DQM

Source: Adapted from Otto et al., 2012; Heinrich et al., 2009.

As it can be seen in Figure 2.9, DQM and MDM interact with each other, where preventive data quality measures as data governance, business metadata management and the development of data quality metrics impact the way master data structure is defined, creating therefore master data of quality. Moreover, there are further aspects beyond MDM, as reactive data
quality measures such as the identification and correction of data defects, to improve the current quality of the data (Otto et al., 2012; M. Hüner et al., 2011). This framework has been built specifically for master data, there are however other types of data that are impacted by the practices of DQM. For further visualization of the connections between data, types of data, data quality and DQM, see Appendix - 2: Data and Data Quality Conceptual Model.

2.3.4 Data's Fitness for Use

As stated earlier, data of quality is data that meets a company’s operational requirements and business needs. However, to determine its fitness, there are different approaches utilized by researchers: Data quality metrics or dimensions through measuring points, measuring techniques and measuring scales (Batini and Scannapieca, 2006), other authors define data quality via procedures and practices resulting from interviews and surveys (Lee et al., 2002; Nicolaou and McKnight, 2006), additionally, other authors use validation rules to define data’s fitness (Fan et al., 2008).

Defining quality dimensions through measuring points, techniques and scales could be a challenging task, however, this method offers considerably more objective and valid results than those based on employees’ perceptions or evolving business needs through static validation rules. Nonetheless, defining which dimensions to measure, taking into consideration their business impact requires intensive planning.

Furthermore, research conducted by Richard Wang and Diane Strong (1996) has found by surveying data-working professionals and M.B.A students, that more than 170 different quality dimensions could be measured, thereafter, a second study considering 1,500 data-using participants. In the second study all dimensions were weighted to keep only the quality dimensions regarded as the most popular in the practice. In Table 2.5 the top 15 rated data quality dimensions are shown by category of quality dimensions (being 1 the perceived most important and 15 the least important quality dimension).
Table 2.5 - Top 20 Quality Dimensions

<table>
<thead>
<tr>
<th>Category</th>
<th>Dimensions (Ranking)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic Data Quality</td>
<td>Believability (1), Accuracy (4), Objectivity (8)) and Reputation (11).</td>
</tr>
<tr>
<td>Contextual Data Quality</td>
<td>Value-added (2), Relevancy (3), Timeliness – age of data (9), Completeness (10), Appropriate amount of data (15)</td>
</tr>
<tr>
<td>Representational Data Quality</td>
<td>Interpretability (5), Ease of understanding (6), Representational consistency - format (12) and Concise representation (13).</td>
</tr>
<tr>
<td>Accessibility Data Quality</td>
<td>Accessibility – availability of data (7) and Access security (14).</td>
</tr>
</tbody>
</table>

Source: Adapted from Wang and Strong, 1996.

Nonetheless, several authors (Otto et al., 2012; English, 1999; Mosley et al., 2010; Wang et al., 1995; 1996; Ballou and Pazer, 1985; Delone and Mclean, 1992; Jarke and Vassiliou, 1997) have consistently used the following five dimensions, making them, if not the most important quality dimensions, the most popular among authors:

- **Correctness**: Also known as accuracy, data factually represents the characteristics of the real world object.
- **Consistency**: Uniform, self-consistent representation of data across different systems.
- **Completeness**: The breadth, depth, and scope of the real world object is contained in all of the values or attributes.
- **Actuality**: Also known as timeliness, the data corresponds to the current status of the real object. Regularly, adaptability measures are required when state changes occur.
- **Availability**: Also known as accessibility or punctuality, access to the data is provided at the desired point in time by data users (Otto et al., 2012).

However, in order to assure data quality, preventive and reactive practices of DQM should be followed. Boris Otto and Hubert Österle propose a framework for corporate Data Quality Management based on Business Engineering, a method to streamline firms in the digital age (Österle and Winter, 2003). This framework is developed in three levels: strategic, organizational and information systems. The first strategic level requires defining goals for data
quality and guidelines, assuring that organizationally, each department is aware of their role in data quality. Secondly, the organizational level, requires metrics to be in place since “You cannot control what you cannot measure” as DeMarco stated in 1982. Also, to include those metrics in the annual goals, assign clear ownership per department, and also, associate in data quality tasks. Additionally, to plan for managing the master data across its lifecycle. Lastly, at an information systems level, clear business objects should be defined, further, data models should be stated based on the business objects together with an architecture for data retention and distribution, see Figure 2.4.

2.3.5 Data’s Time Factor

Based on the different data quality dimensions covered in sub-chapter 2.3.4, there are two that are influenced by the temporal factor: (1) Timeliness, the age of the data and its validity at a given point in time, and (2) Availability, data is accessible, retrievable, at an appropriate speed and at the point in time when it is requested (Otto et al., 2012; Wang et al., 1996). The latter quality dimension is the core of this study. Although different authors have used the concept Timeliness for referring to both definitions, (1) and (2). Moreover, other terms as punctuality or accessibility have been regarded as the same data quality dimension, however, the latest streams of research have been consistent with the definitions stated at the beginning of the paragraph and will be used in this sense further in the research (Otto and Österle, 2015, 2016).

Similar research to this thesis has been conducted by Hüner, Schierning, Otto and Österle (2011) in the company Beiersdorf focused on identifying critical product data defects for business, and proposing metrics that allow the company to measure the data quality, although the authors considered the temporal dimension in their study, as well as the exchange of product data between manufacturers and retail companies. The previous approach taken was not based on measuring how to deliver product data on time, but rather on assuring that precise data attributes have no delays due to potential internal inefficiencies. The scope of this master’s degree thesis takes a customer-centric approach by clarifying the master data expectations of the retail companies in regard of the “timely” disposition of data, namely, defining when exactly
“timely” takes place and which product data is needed at that point in time to enable retail companies to perform all processes related to the commercialization of a new product.

The costs of poor data quality in cross-company data communication have been covered widely regarding transactional data such as sales forecasts. However, cross-company data communication of product data has been somewhat left aside, despite the criticality of exchanging data across supply chains, for example, to assure the correct interpretation of data (M. Hüner et al., 2011). This study aligns the stages of creation of the data, or Beginning of Life (BOL) from PLM, with the delivery of the product and its data, or the Middle of Life (MOL) stage, within the framework of collaboration from the Retail SCM and assuring data quality.
3. DATA COLLECTION AND METHODOLOGY

Research design resembles the work of an architect where different variables play a role in defining how the final form of a project will be shaped, for example, the preferences and philosophy of the researcher, as well as the predilections and resources provided by the funding organization. Ultimately, the design should result in the optimal methodology to address the research questions, or in the case of the architect, to create a stable, functional and appealing edifice (Hakim, 2000).

The strategy in this master’s degree thesis pursuing to answer the main research question “How can a manufacturer meet the required quality for its product master data on time to enable omni-channel commerce for its trading partners in the Fast-Moving Consumer Goods industry?” combines exploratory and descriptive methodologies. As mentioned in the sub-chapter 1.3.1, significant variations among retailers in regard to the point in time when they request Product Master Data, as well as differences in the channels being served impact the attributes of the requested PMD. Therefore, this context suggests that research is needed to:

1. Create initial knowledge about the nature of the listing process, since this specific process between manufacturers and retailers has not been documented before.
2. Facilitate a retail industry-wide standard, specifying the point in time for the disposition of PMD, and which PDM should be communicated at this point in time.

Hence, goal number (1) corresponds to the explorative framework definition by (Robson, 2002) “what is happening; to seek new insights; to ask questions and to assess phenomena in a new light”. By defining what the retailers’ listing process is and what activities does it cover, although these activities were unknown and different among retailers, the questions of “what is happening” is being answered while the study benefits from the flexibility and robustness that the exploratory research design provides towards uncertain results (Adams and Schvaneveldt, 1991).
Additionally, this study through goal number (2) aims at finding a precise point in time in which the product data should be made available for the retailers, as well as, defining which product data is needed at this given point in time. However, the explorative methodology is limited in portraying an accurate profile of the exact data needed per activity of the listing process against a timeline. Therefore, as recommended by (Saunders et al., 2009), descriptive research is used to have a clear picture of the phenomena. Thus, descriptive and exploratory methodologies have been combined in order to define what the listing process is, what activities take place during the listing process, what product data is needed for the listing process, when during the listing process is the product data needed, and lastly, what the current status of data quality is, answering thus the sub-questions #2 to #5, see Table 3.1 for further detail.

This study is a precursor for further research, however, due to the early stage of research in the corporate data (master data) exchange in the retail industry, the aim of this study is to show the existence and relevance of the time factor in the data communication process rather than directly measuring its influence on data quality with quantitative research.

As mentioned earlier in this chapter, there is no historical data recorded that shows the point in time when the product data has been made available in relation to the first shipment of a new product and the resulting quality of the product data. Nonetheless, this study contributes to the development of metrics or Key Performance Indicators (KPIs) to measure the temporal quality dimension of the data by providing a measuring point (Batini and Scannapieca, 2006), and hence, enable further explorative research.

The contents of this section include the detailed description of the research design and the research strategies employed to gather information and the justification of the strategies for answering the research questions. Also, the structure, development and techniques for data collection are mentioned. Lastly, the measures taken to assure the reliability of the results are covered.
3.1 Research Design

According to (Yin, 2003), different research strategies, or combinations of them, can be employed to conduct exploratory, descriptive and explanatory research. Nonetheless, the selection should consider that each strategy aims at answering different questions. For example, who, what, how, where, why, how much and how many. Therefore, the strategy should be carefully tailored to cover the research objectives, and therefore, the research questions.

A summarized visualization of the research objectives and research sub-questions with their strategies can be found in Table 3.1:

Table 3.1 - Research Design

<table>
<thead>
<tr>
<th>Research Sub-Question</th>
<th>Research Objective</th>
<th>Research Method</th>
<th>Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ-1: How can data quality be achieved for product master data?</td>
<td>Define product master data, data quality and how product master data of quality can be achieved</td>
<td>• Literature research  • Explorative research</td>
<td>• Scientific papers in journals  • Conferences  • Specialized  • Books</td>
</tr>
<tr>
<td>SQ-2: What activities are performed by the retailers in order to list a new product and what is the duration of those activities?</td>
<td>Determine which activities entail the listing process and when they take place, using the first shipment as reference</td>
<td>• Qualitative multiple case study  • Descriptive research</td>
<td>• Semi-structured interviews  • Process Mapping (Workflow diagram)</td>
</tr>
<tr>
<td>SQ-3: What product attributes are required for each of the activities or processes involved in listing a new product?</td>
<td>Determine which PMD attributes are needed in each activity of the listing process</td>
<td>• Qualitative multiple case study  • Descriptive research</td>
<td>• Semi-structured interviews  • Attribute clusters mapping</td>
</tr>
<tr>
<td>SQ-4: What data requirements differ per channel PMD</td>
<td>Determine the omni-channel PMD</td>
<td>• Qualitative multiple case study  • Descriptive research</td>
<td>• Semi-structured interviews</td>
</tr>
<tr>
<td>channel (online &amp; offline)?</td>
<td>requirements and their timings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ-5: What is the current status of data quality for product master data sent by the manufacturers?</td>
<td>Determine current performance of product master data quality</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| • Multi-method approach. |
| • Qualitative multiple case study |
| • Explorative research |
| • Semi-structured interviews |
| • Scientific papers in journals |
| • White papers |

Scientific research practices support the use of case study while investigating organizational issues related to information systems, in this case the inter-company exchange of product master data between manufacturers and retail companies, to handle phenomena within its real life context, as the data exchange between data pools regulated by international standard organizations. Additionally, this thesis conducts a multiple case study approach to improve the generalizability of the results (Yin, 2003). This type of research strategy is suitable if previous research on the subject is limited, or in case that a phenomenon has not been understood theoretically, as it is the scenario with this area of research on the impact of the temporal factor on data quality in the data exchange performed by retail companies (van der Blonk 2003).

### 3.2 Description of Study Cases

Initially, six leading companies in the retail industry in Germany were invited to participate in the study, however, due to different limitations to allocate resources on the topic during the time window for the interviews, the number of participant companies was reduced to four. The participating companies are active internationally in the retail industry, yet, their answers and processes correspond solely to the German market. The names of the companies remain anonymous and the results generalized in accordance with a non-disclosure agreement, as mentioned in the delimitations of the study in sub-chapter 1.4.

Altogether, the four participating companies are categorized within the following areas: (1) Pure online retailer, (2) online and offline retailer and (3) master data service providers. Due to the international nature of the participating companies, it is possible that in different countries their commerce presence and strategy differ. The first category, pure online retailer, represents retail
companies that conduct solely online commerce, whereas the second category, represents retailers active in brick and mortar commerce as well as online commerce, or combinations of both.

Additionally, Master Data Service Providers (MDSP) take the role of an intermediary between manufacturers and retail companies, regardless of their market presence strategy, online or offline. These providers offer master data management services that are wide ranging, an example of the services includes accessing data pools to retrieve the product data of the desired items to then communicate them through a different system platform to a retail company, additionally, the MDSP can control the data for quality. Simply put, the MDSPs could take responsibility for their customers in activities related to Master Data Management (MDM), to see more information about MDM, refer to sub-chapter 2.3. Although the MDSPs do not sell products to end-consumers as the retail companies do, they are key stakeholders in the listing process due to their interactions with the product master data and their retail partners.

Lastly, each of the participating companies have different market strategies and could offer different product categories to customers. Moreover, reflecting their strategies by utilizing different approaches to serve end-customers, and therefore, having different business requirements that might be extended to the products and data delivered by manufacturers.

3.3 Data Collection

Due to the multi-research method used in this study, both explorative and descriptive strategies have led to different types of sub-questions being proposed to answer the main research question. As described in Table 3.1, some sub-questions are answered by referencing to literature whereas other sub-questions are answered by a qualitative study, or a combination of both. Regarding the reference to literature, chapter 2 describes in detail the resources and strategy used to gather knowledge on various topics. Figure 3.1 depicts different sources of knowledge used for building the study framework that aggregate to 118 sources. The major source being online documents, which were fundamental to cite standard concepts from
different organizations, articles from digital magazines, consulting firms’ publications, finding major trends in each of the research disciplines, approaches from leading companies to services and products within the research disciplines, as well as for obtaining information and content from the sponsoring and collaborating organizations. Secondly, specialized books in both, printed and digital editions were used to robust theoretical concepts and frameworks. Thirdly, journal articles were consulted to gather the voices of different experts in the different fields covered. Lastly, a combination of conference procedures, theses, university lectures and white papers were used. For a visual representation of the sources see Figure 3.1.

Figure 3.1 - Knowledge Sources

In order to best collect primary data from the companies considering different operations from the participating companies, not all elements of a fixed questionnaire would apply for all, therefore, a semi-structured interview design was used, since this method offers flexibility to the researcher and to the interviewee to utilize the natural flow of the interview to select the sequence of questions to be discussed (Saunders et al., 2009). Moreover, this design allowed a seamless flow in the interview, considering that all questions were shared as optional, to allow the interviewees to focus in the areas of the listing process that they considered more
pertinent. The elements of the interviews, their sequence and detailed description can be found in the following sub-chapters.

### 3.3.1 Invitation and Questionnaire

Based on suggestions from experts in the sponsoring and collaborating organizations, as well as following recommended practices to conduct interviews and process mappings, a questionnaire with 34 questions was designed. The complete set of questions can be found in Appendix - 3: Questionnaire for Interviews. A summary with the simplified content of the questions can be found in Table 3.2.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 10</td>
<td>Listing process insights – Overview for the creation of the research context</td>
</tr>
<tr>
<td>11 to 18</td>
<td>Mapping of the listing process – Activities, attributes used and durations, among other details</td>
</tr>
<tr>
<td>19 to 32</td>
<td>Status of data quality</td>
</tr>
<tr>
<td>33 to 34</td>
<td>Improvement of the questionnaire and the attribute clusters</td>
</tr>
</tbody>
</table>

Following recommended practices for interviews (Saunders et al., 2009), the first set of questions were general and introductory to the topic in order to develop a pleasant atmosphere. Thereafter, the core of the questionnaire based in the listing process and data quality. Lastly, two questions were addressed to improve the questionnaire, since further explanatory research could follow, and improvements to the questionnaire are critical to automatize the method through electronic survey tools and broaden the audience for further generalization purposes.

Furthermore, a cover letter for the questionnaire was redacted with the aim of introducing the research, the scope and the overall guidelines of the questionnaire and interviews. The summarized points of the guidelines for the questionnaire are the following:
• Non-disclosure-Agreement for sensitive information as names of companies
• Note for possible recording of the interview
• Reference to the source of attribute names for the process mapping
• Expected duration of the interview, overall process of the interview and invitation to answer only the desired questions
• Presentation of sponsoring and supporting organization, as well as university guiding the master’s thesis
• Definitions of concepts used in the questionnaire

Lastly, after the design of the questionnaire was completed, an invitation to the six companies for the interviews was sent via the supporting organizations GS1 Germany and Smart Data One. Next to the invitation and the questionnaire, a presentation introducing the project, its goals and the researcher was attached. The invitations were sent approximately three weeks prior to the time window to conduct the interviews. Once the companies replied with a point of contact, the researcher contacted the person by phone and explained again the goals of the project as well as the approach, then an interview appointment was scheduled.

3.3.2 Interviews

The interviews were conducted in a Face-to-Face basis as suggested by research practices when complex information such as processes have to be mapped (Saunders et al., 2009). All interviews took place in the premises of the companies being interviewed and had an average duration of four hours including short breaks. In some companies the interviews were conducted in a one-to-one basis whereas with other companies on a one-to-many basis. All companies allowed the interviews to be recorded.

At the beginning of the interviews the researcher introduced himself and presented selected slides of the project to give continuity and answer questions about the project, the interview and managing of the results. While conducting the interview, all written answers were projected or shown in real time to the interviewee to foster transparency and avoid the wrong wording for
activities or processes. The process mapping was captured in a digital spreadsheet directly in the interview. Finally, after concluding the interview a follow-up call was scheduled to discuss any additions or changes to the results of the interview.

Additionally, during the process mapping attribute clusters or attribute groups were used to facilitate the process mapping by allocating one or more group of attributes to each of the activities in the process mapping, hence, identifying which attributes are needed per activity or process. The usage of attribute clusters was inspired by the approach taken from M. Hüner, K., Schierning, A., Otto, B. and Österle H. (2011), in a study conducted to improve master data quality. The usage of attribute clusters is essential to facilitate the process mapping since as mentioned in sub-chapter 1.3.2, there are approximately 500 attributes that describe products, which could lead to an herculean task to perform if the process mapping was conducted in that level of detail. Therefore, with assistance from the sponsoring and collaborating organizations, 14 attributes clusters containing approximately 115 attributes were created, see Table 3.3 for further detail on the clusters.

<table>
<thead>
<tr>
<th>Attribute Cluster</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Party / product owner</td>
<td>Informations to the brand owner, producer and information provider</td>
<td>Manufacturer GLN</td>
</tr>
<tr>
<td>Identification</td>
<td>Attributes that differentiate a product from others</td>
<td>Trade item identification: GTIN</td>
</tr>
<tr>
<td>Classification</td>
<td>Global Product Classification (GPC) information among others</td>
<td>GPC Brick code</td>
</tr>
<tr>
<td>Description</td>
<td>Free form descriptions of product and line of trade</td>
<td>Trade item description / Language</td>
</tr>
<tr>
<td>Hierarchy / indicators</td>
<td>Quantities &amp; GTIN of items in the next lower packaging level, Indicators (Yes/No) for base unit, shopping units among others</td>
<td>Quantity of children</td>
</tr>
<tr>
<td>Logistics data: dimensions &amp; temperature</td>
<td>Logistical measurements, units and temperatures for handling the product.</td>
<td>Gross weight / UOM</td>
</tr>
<tr>
<td>Category</td>
<td>Description</td>
<td>Additional Information</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Additional logistics data: transport units</strong></td>
<td>Pallet information (rented, exchangeable), number of layers of base trade item, if pallet has no GTIN, quantities of items.</td>
<td>Case Level Non GTIN: Quantity of trade items per pallet</td>
</tr>
<tr>
<td><strong>Dangerous goods</strong></td>
<td>Information about dangerous goods code and products within each item</td>
<td>Dangerous goods</td>
</tr>
<tr>
<td><strong>EU regulation on product label additions</strong></td>
<td>Information on the packaging and needed for the implementation of EU-Regulations (Ingredients, Allergens…etc.)</td>
<td>Allergen type code</td>
</tr>
<tr>
<td><strong>Packaging &amp; returnable assets</strong></td>
<td>Item is bar-coded, expiration date on packaging, returnable packaging indicator…etc.</td>
<td>Data carrier type code</td>
</tr>
<tr>
<td><strong>Taxes</strong></td>
<td>Information about VAT, and use taxation rules</td>
<td>VAT</td>
</tr>
<tr>
<td><strong>Customs</strong></td>
<td>Specific country of origin</td>
<td>Import classification: Type</td>
</tr>
<tr>
<td><strong>Non id-attribute information</strong></td>
<td>Pictures and Artworks</td>
<td>Product images split by shot types: ID, out of pack, marketing left, marketing right, etc.</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>Date times, for validity of data, availability for supplier, availability for synchronization</td>
<td>Effective date time</td>
</tr>
</tbody>
</table>

The attribute clusters were designed based on a logical grouping, for example, all attributes that describe the morphological characteristics of a product were grouped in the “Logistics data” cluster. The basis for this clusters was provided from the collaborating organization taking a list of attributes that were defined as the essentials to conduct the listing process. Due to changes in GDSN attribute structures that took place in May 2016 this documented was partially outdated, therefore adoptions were done and additional attributes and groupings were added based on literature (M. Hüner et al., 2011) and advice from experts in the field. However, an ideal cluster of attributes would be built based on business processes with precise activities, yet, this information was not available prior to this study.
3.3.3 Processing the Results

After concluding the interviews, within the next two to five days the researcher processed the results of the interviews, enriching the answers to the questionnaire based on written notes and recorded material from the interviews. Additionally, the results of the interviews were anonymized and generalized with the scope of keeping the anonymity agreed in the non-disclosure agreement, also, specific names for departments or processes were changed to neutral names in order to avoid the identification of the companies by this information.

Furthermore, additional visualization aids that can be seen in chapter 4 were developed to aid the correct interpretation from the process mappings in regards of durations, sequence of processes, parallel running activities and corresponding product data attributes for each activity. The results were sent to the interviewees and reviewed internally, to then be discussed, modified or complemented in the follow-up calls.

3.3.4 Follow-up calls

As agreed during the interviews, follow-up calls took place with the interviewees. The content and detail of the answers, together with the process mapping was discussed. Regularly, minor to moderate adjustments took place in this phase. The aim behind this activity was to obtain results that correspond to the activities and processes of the retail companies, especially since detailed information as the activities, durations and attributes were captured in the interview.

Lastly, the modifications requested by the interviewees were applied and sent again to the interviewees. Based on these results, the interviewees provided their written sign-offs through emails to use this information as input for (1) an analysis, (2) use the results in this thesis and (3) present a comprehensive overview of the topic in the GS1 GDSN Workgroup moderated by GS1 Germany representatives, were different companies from the FMCG and Retail industries were present. The latter to become the basis for further process standardization.
3.4 Data Reliability

The major possible threat to reliability posed by the research methodology could be caused by language barriers. The questionnaire, presentations of the topic, language used during the interviews, follow-up calls and presentations of results were conducted in German language, which is the second foreign language from the researcher. Although the researcher is fluent in the German language, additional measures were taken to mitigate the risk of misinterpretations. Mainly, guidance and support provided by native speakers that are active in the topic and therefore know the correct terminologies. This aid was taken for all steps that involved written documentation. Further, in order to assure reliability of the results from the interviews, the interviews were recorded, the results were processed and sent to the interviewees, afterwards, the interviewees confirmed the results as described through the previous sub-chapter 3.3.4 “Follow-up calls”.

59
4. CASE STUDIES

The case studies found in this section were funded and supported by Mars GmbH (Germany) and developed in collaboration with GS1 Germany and its subsidiaries Smart Data One and 1WorldSync. The case studies involved the participation of personnel from the companies with knowledge of the listing process and its cross-departmental interactions within the company.

The entirety of the information gathered from the interviews regarding processes, strategies, requirements and used channels was shared to enable the realization of this work. Hence, assuring a clear picture of the aimed business process through primary data. Prior to the collection of the information, a description of the listing or commercialization process was mentioned to relate the interviewees with the same set of activities, since different companies might refer to the same process under different names. Particularly, since this process was not documented across-companies before. The definitions used to guide the study in the intended direction can be found in Table 4.1.

Table 4.1 - Core Case Study Definitions

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listing process</td>
<td>Refers in this study to all sub-processes that start with the 1st listing meeting (Presentation of a new product) between procurement and sales and finish with the product activation in the retailer system / start of ordering process.</td>
</tr>
<tr>
<td>New product</td>
<td>Refers in this study to every product that obtains a new GTIN assigned (This includes products with a predecessor that based on a re-launch and GTIN allocation rules have been assigned a new GTIN).</td>
</tr>
</tbody>
</table>

This chapter is divided in four sub-chapters, each sub-chapter is dedicated to one case study. Within each sub-chapter the most important elements of the listing process for each company are mentioned, also the process diagram from the listing process for that company including the attributes needed is displayed, and lastly, the current status of data quality in the company.
is shared. Not all questions were answered by all participating companies, therefore, the results might vary in content from case study to case study.

4.1 Company A

Company “A” is active in brick and mortar stores and in online commerce in Germany. This company performs their sourcing activities centrally organized for Germany. However, each subsidiary has the freedom to select which products from the company’s catalog prefers to offer, important factors that subsidiaries consider is the size of the store for choosing the assortment. Product data from Germany is not shared with stores from other countries. Company “A” is an internationally active company, however its German segment is completely independent from other functional areas or stores in different countries.

4.1.1 The Listing Process

The listing process at Company “A” starts with the presentation of a new product with the procurement team in the called “1st Listing Meeting”. In Company “A” the procurement team has autonomy on the decision to include a product in the firm’s portfolio, which reduces the time needed to close initial negotiations with the manufacturers, since 70 percent to 80 percent of the times the decision to list a product takes place during the 1st listing meeting. Traditionally, all products that are chosen to be listed for physical stores are also listed in the online shop, however, from eight percent to ten percent of the times it is decided that the products will be sold solely in the online shop.

Company “A” utilizes the services of a master data service provider which performs as a data intermediary between the manufacturers and Company “A”, however, direct contact between the retailer and the manufacturer takes place to accelerate processes, nonetheless, the data follows the standard path through the master data service provider. Company “A” participates yearly in strategic meetings with manufacturers to discuss about innovations for products and improve their market adoption. Regularly, the 1st listing meeting takes place six months before
the first shipment of the merchandise, although the ideal time for the presentation of the product for company “A” would be one year before the first sale.

4.1.2  Process Diagram and Attributes

The process flow depicting the listing process of Company “A” can be seen in Figure 4.1, whereas the description, duration and attributes needed by the sub-processes are explained in Table 4.2.

![Figure 4.1 - Listing Process in Company "A"

For the correct interpretation of Figure 4.1 note that the arrows indicate that the previous step runs further in parallel with the sub-processes below or above the arrow. Also, for the correct interpretation of Table 4.2, in the column “Duration” when no content is shown, it represents a milestone that does not involve processing time. In the column “Attribute clusters” the attributes will be listed when they are used, which is different as when the data is requested (currently). In addition, when the content of this column is written in “italics” it represents data or information not considered within the product attribute clusters explained in Table 3.3, alternatively if no content is found, it represents that no attributes are used in this sub-process. Also, from this point onwards the term “product” and the term “item”, as in Global Trade Item Number (GTIN), will be used interchangeably since the term item relates to the virtual product stored in a system.
### Table 4.2 - Listing process of Company "A"

<table>
<thead>
<tr>
<th>#</th>
<th>Sub-process</th>
<th>Description</th>
<th>Duration (wks.)</th>
<th>Attribute clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; Listing Meeting</td>
<td>New product is presented to the procurement team from the retail company. Negotiations about different terms and conditions take place. It will be discussed about packaging shape and materials.</td>
<td>1 - 4</td>
<td>Prototype or Mock-up from the product.</td>
</tr>
<tr>
<td>2</td>
<td>Written confirmation of agreements is sent</td>
<td>Agreed terms and conditions of commerce are written and confirmed by both parties.</td>
<td>1*</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Filling information in the listing data base</td>
<td>Most relevant informations about the new product listing are stored in retail internal databases.</td>
<td>1*</td>
<td>Retail internal data</td>
</tr>
<tr>
<td>4</td>
<td>Capturing data for marketing planning</td>
<td>Creation of the product retail specific identification number. Data from internal databases is transferred to the company’s ERP system to plan marketing strategy.</td>
<td>1*</td>
<td>Retail internal data</td>
</tr>
<tr>
<td>5</td>
<td>Marking processes</td>
<td>It is decided if the product will be listed online and offline or entirely online. If the product will be sold in both, printed media is considered. Printed media requires 4 months to be completed due to coordination with agencies and regular lead times.</td>
<td>12</td>
<td>Classification, Logistics data, Party / Product Owner, Packaging &amp; returnable assets, Date, Identification, Description, EU regulation product label additions and Non id-attribute information</td>
</tr>
<tr>
<td>6</td>
<td>Data provision for the listing</td>
<td>Remaining datasets are filled in the internal database for the listing. Information includes calculations, taxation rules and channel strategies (online/offline).</td>
<td>1*</td>
<td>-</td>
</tr>
<tr>
<td>#</td>
<td>Activity Description</td>
<td>Details</td>
<td>Duration</td>
<td>Notes</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>7</td>
<td>Data storage processes</td>
<td>The item is completely stored in the system through retrieving data from the master data service provider. Data integration from PIM-ERP-Internal data bases. Data quality validations take place.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Logistics planning</td>
<td>Warehouse is selected, quantities are planned, first order is planned and repackaging is considered.</td>
<td>2-4</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Shelf Planning</td>
<td>Layout of the shelf is planned for the new product.</td>
<td>1*</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Activation of the Item ID – Introduction in online and offline channels</td>
<td>Validations regarding visualization of the product in online commerce. Item is activated for selling after product data is correct and complete</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

(*) Used to denote those activities that last less than a week.

As it can be seen from the column “Duration”, the longest timing is impacted by printed media, a strategy primarily relevant for brick and mortar commerce. Also, different sub-process as #2, #3, #4, #6 and #9 last less than a full week. Taking this into consideration in addition to total average duration for the listing process suggested by the interviewees, the resulting duration of the listing process has been considered to last between 15 and 17 weeks to guarantee robustness against most variations in the entire process. Whilst the duration of the listing process for only online commerce related activities (without printed media) lasts between 6 and 8 weeks.

4.1.3 Data Quality and Quality Measures

According to the results from the interview, Company “A” controls the quality of the product data in collaboration with the master data service provider. Manual and automatic data quality validations take place, there are tools that give a hint to the retailer on which products to check.
manually for wrong data. There is no team assigned solely to the quality validation of the data, however, while the company associates conduct their data related tasks they also validate the content of the data. All GTINs (referring to all products) are verified for the quality dimensions completeness, consistency (format, units used...Etc.) and correctness. Almost all product attributes are validated, the attributes are validated when they are used.

In average, the data quality error rate found by Company “A” in the items is between 60 percent and 70 percent. In average, the product master data has to be retrieved from the GDSN data pool between two to three times until the dataset is correct and/or complete. When corrections have to be applied to the data, the waiting time until the updates have been sent can take up to one day.

4.2 Company B

Company “B” is a pure online commerce retailer in Germany. This company executes its sourcing activities, including the management of Product Master Data (PMD), separately per product category within Germany. Nonetheless, Company “B” collaborates actively with the subsidiaries from other countries, where PMD can be captured from products originated in foreign countries but can be sold through the German subsidiary, however, PMD originated in foreign countries must fulfil the data requirements from the German market, such as the language among others. International operations for Company “B” are not limited to commercial activities, they also include administrative tasks relevant to the listing process that are spread across the globe.

4.2.1 The Listing Process

The listing process for Company “B” starts with the 1st listing meeting for new products were prices and different quantities are discussed. Initially, a major requirement for manufacturers to list a product with Company “B” is to have a contract. The arrangement of the contract details can take place during the 1st listing meeting. Afterwards, the manufacturers can independently
list all their products through a supplier online platform. There are two different paths in order to upload the PMD into the supplier online platform, Company “B” differentiates between food product categories and others. Due to the Food Information Regulation (FIR), food products and their data are subject to higher scrutiny that compel manufacturers to provide specific mandatory attributes, for further information see sub-chapter 2.3.1. Thus, manufacturers of food products fill product data templates with their PMD, and only the mandatory FIR attributes are filled through the GDSN data pool, whereas non-food manufacturers will only fill the product data templates with their PMD. The templates utilized to upload PMD to the supplier online platform are product category specific.

Thereafter, Company “B” analyzes the different products for profitability and assesses the next steps accordingly. Further major requirements for the listing process are pictures of the product, to be made available for end-consumers, without pictures the product will not be activated in the system. Due to the nature of pure online commerce, there are no physical shelf planning activities, nonetheless, the early communication of PMD allows Company “B” to manage their product forecasts and warehouses more efficiently.

4.2.2 Process Diagram and Attributes

The process flow depicting the listing process of Company “B” can be seen in Figure 4.2, whereas the description, duration and attributes needed by the sub-processes are explained in Table 4.3.
For the correct interpretation of Figure 4.2 note that the arrows indicate that the previous step runs further in parallel with the sub-processes below or above the arrow. Also, for the correct interpretation of Table 4.3, in the column “Duration” when no content is found, it represents a milestone that does not involve processing time. In the column “Attribute clusters” the attributes will be listed when they are used, which is differently as when the data is requested. Also when the content of this column is written in “italics” it represents data or information not considered within the product attribute clusters explained in Table 3.3, alternatively if no content is found, it represents that no attributes are used in this sub-process.

**Table 4.3 - Listing process of Company "B"**

<table>
<thead>
<tr>
<th>#</th>
<th>Sub-process</th>
<th>Description</th>
<th>Duration (wks.)</th>
<th>Attribute clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1st Listing Meeting</td>
<td>Contractual terms arranged – Only required when no contract is available.</td>
<td>1 - 8</td>
<td>Product catalog</td>
</tr>
<tr>
<td>2</td>
<td>New product introduction desired</td>
<td>Manufacturer decides to list a new product or Category Management asks manufacturer to list a missing product from the manufacturer's catalog.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Filling the product data template</td>
<td>The supplier online template is filled by the manufacturer according to the product category, if the product is food, FIR attributes are filled in via GDSN. Automatic validation rules take place.</td>
<td>1</td>
<td>Classification. Retail specific attributes.</td>
</tr>
<tr>
<td>4</td>
<td>Provision of product data template</td>
<td>The manufacturer uploads the template in the supplier online platform. Automatic validation rules take place.</td>
<td>-</td>
<td>Identification</td>
</tr>
<tr>
<td>5</td>
<td>Processing product data</td>
<td>Creation of a calculation spreadsheet. Cross-departmental validation of the calculation spreadsheets.</td>
<td>1 - 2</td>
<td>Logistics data, Tax &amp; Customs, Hierarchy / Indicators, Party / Product Owner, Date, Description and EU regulation</td>
</tr>
<tr>
<td></td>
<td>Profitability validations and product bundling</td>
<td>Profitability test are conducted per item. If the desired profitability is not reached negotiations with manufacturers or product bundles are done. Manual validations for data quality performed.</td>
<td>1* -</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Creation of the retail specific item ID</td>
<td>The product data are stored and included in the retail product catalog. A retail specific ID is created.</td>
<td>1* -</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Product pictures are uploaded</td>
<td>Manufacturer receives a confirmation of the retailer ID creation. Manufacturer uploads the product pictures through the supplier online platform.</td>
<td>1* Non id-attribute information</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Validation of product pictures</td>
<td>Product pictures are tested for compliance.</td>
<td>1* -</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Validation for dangerous goods</td>
<td>Validations for dangerous goods are performed, if the product is a dangerous good, transportations and warehousing will be planned accordingly.</td>
<td>1* Dangerous goods</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Activation of the Item ID</td>
<td>The product is activated and can be ordered and purchased.</td>
<td>- Additional logistics data and Packaging &amp; returnable assets</td>
<td></td>
</tr>
</tbody>
</table>

(*) *Used to denote those activities that last less than a week.*

As it can be seen from the column “Duration” there are major variations in sub-process #1 that account for more than the duration of the remaining listing sub-processes. Therefore, the average time of 1 to 2 weeks will be considered for sub-process #1. Also, different sub-process as #6, #7, #8, #9 and #10 last less than a full week. Taking this into consideration in addition to total average duration for the listing process suggested by the interviewees, the complete
duration of the listing process has been considered to last between 4 and 6 weeks to guarantee robustness against most variations in the entire process.

4.2.3 Data Quality and Quality Measures

According to the results from the interview, Company “B” controls the quality of the product data through manual and automatic validations. All GTINs are verified for quality dimensions completeness, correctness, formatting and plausibility. The validation of all items contained in a product data template, depending on the amount of items, could last up to six weeks, while it takes in average 2 minutes to validate each item for quality.

In general, the perceived error rate found in the items is high, although currently there are no historical records available. The error rate varies among manufacturers and product categories, however, the error rate of manufacturers that upload their data through the GDSN data pool is considerably lower, since they do not have to manually capture approximately 65 FIR-Attributes for food products.

4.3 Company C

Company “C” is a master data service provider that provides MDM services to different customers. All customers from the company are active in online and offline commerce, thus with multi-channel strategies. However, not all customers from Company “C” have the same data requirements. Some customers from Company “C” require solely the attributes provided by the manufacturers through the GDSN data pool whereas other customers require additional retail specific attributes.

The master data service provider adapts its operations and processes to the data needs from the customers, therefore, differences in the timings for data provision as well as variations in the types of data quality controls can occur. For example, a customer might request that the
product master data is sent directly after it has been received, regardless if errors or bad quality data was found. On the contrary, other customers might request that Company “C” holds the data until the manufacturer has corrected it, leaving the duties to assure that the data has been corrected to the master data service provider.

Company “C” is based in Germany and works centrally and decentralized. Centrally since it provides data from German manufacturers to customers active in the German market, and decentralized because Company “C” serves customers from other countries providing them with product master data from German manufacturers. Yet, the data has to be appropriate by having the right flag “Target Market” to fulfill the requirements of different countries.

4.3.1 The Listing Process

Company “C” does not conduct any listing process, although it is active in the retail industry by providing services to retail companies it does not commercialize products. The company rather assists retailers to conduct their listing processes. Thereby, taking ownership of the data management steps to then share the processed and controlled data to its clients.

Normally, the retailers request for product master data six weeks before the introducing the new products in the market. In the case that six weeks are not enough, the retailers contact Company “C” and ask for the missing items. However, for seasonal products, there are retail companies that request the product master data up to nine months prior to the introduction of the product in the market. The master data service provider does not differentiate between channels while requesting data and evaluating data quality, however, the FIR-Attributes are particularly relevant for online-commerce.

Major requirements for the provision of data are the product categories, defined by the GPC code. Since depending on the product category different product attributes will be required. For retail companies are both, standard (GDSN) and retail specific attributes relevant, however, not
all manufacturers are able to deliver all retail specific attributes. The criteria for product quality has to be fulfilled considering timeliness (actuality), availability (timely delivery), completeness, correctness (or accuracy) and correspond to the validation rules determined by Company “C”.

4.3.2 Process Diagram and Attributes

The process flow depicting the data services of Company “C” can be seen in Figure 4.3, whereas the description, duration and attributes needed by the sub-processes are explained in Table 4.4.

![Process Diagram](image)

**Figure 4.3 - Listing Process in Company "C"

For the correct interpretation of Table 4.4, in the column “Duration” when no content is found, it represents a milestone that does not involve processing time. In the column “Attribute clusters” the attributes will be listed when they are requested by the retail companies to the master data service provider. The master data service provider does not use the data, hence the column “Attribute clusters” shows how the disposition of data currently works for the master data service provider, namely, all attributes received follow that process.

**Table 4.4 - Listing process of Company "C"

<table>
<thead>
<tr>
<th>#</th>
<th>Sub-process</th>
<th>Description</th>
<th>Duration (wks.)</th>
<th>Attribute clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial data provision</td>
<td>Manufacturer provides the PMD via GDSN to the master data service provider. Retail specific attributes are communicated</td>
<td>1*</td>
<td>Classification, Logistics data, Party / Product Owner, Packaging &amp;</td>
</tr>
</tbody>
</table>
through a supplier online platform. Data is requested between 6 and 12 weeks prior to the offering of the product.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Validation of the product data for quality</td>
<td>PMD is stored in a separate database and analyzed with own systems for data quality. Technical quality aspects are tested extensively, content is tested manually, depending on the arrangements with the retail company. Data can be directly sent to the retailers without corrections.</td>
</tr>
<tr>
<td>3</td>
<td>Product data is sent to the retailer</td>
<td>Once the data quality controls have been successfully completed (Service contracts with retail companies do not specify content or standard procedures for data) the data is prepared and exported to the system of the retailer (Through GDSN standard interfaces).</td>
</tr>
<tr>
<td>4</td>
<td>Feedback from data receiver</td>
<td>This step takes place solely if the retail company finds errors in the data. The retailer communicates all data errors to the service provider and then the service provider solves the issue with the manufacturer, repeating the process flow from step #1 again once the manufacturer delivers the corrected data.</td>
</tr>
</tbody>
</table>

returnable assets, Date, Identification, Description, EU regulation product label additions, Non id-attribute information, Additional logistics data, Hierarchy / Indicators, Dangerous goods, Tax & Customs
(*) Used to denote those activities that last less than a week.

As it can be seen from the column “Duration” all sub-processes last less than a full week. Taking this into consideration in addition to total average duration for the listing process suggested by the interviewee. The complete duration of the listing process has been consider to last between 1 and 2 weeks to guarantee robustness against most variations in the entire process.

4.3.3 Data Quality and Quality Measures

According to the results from the interview, Company “C” controls the quality of the product data through manual and automatic validations. However, in the future all validations will be performed automatically. Automatic controls test technical aspects as data structure, semantic and field specifications, then the content is controlled manually for attributes such as dimensions, weight, volume and dependencies.

Approximately 20 associates are responsible for technically testing the data quality for the FMCG category. All GTINs are tested and the logistic attributes, basic attributes, taxation attributes, dangerous goods attributes are carefully controlled, and additionally B2C attributes are tested to a lesser degree. The error rate from the data is between 30 percent and 50 percent. When errors are found in the data, the product data has to be retrieved from the GDSN data pool two times in average.

4.4 Company D

Company “D” is active in brick and mortar stores and in online commerce in Germany. This company performs their sourcing activities centrally organized for Germany. However, for imported merchandise they have delivery and trading offices in different countries. For Company “D”, “listing” is a sub-process in the commercialization activities of a new product, which is independent from purchasing and / or category management. The company is currently in a transition phase to redesign its listing sub-processes, currently, the data
requirements are different between online commerce and offline commerce. Namely, for online commerce the product pictures are requested and depending on the product category different attributes are needed. However, with the new design of the listing sub-processes there will not be differences per channel regarding data requirements. Nonetheless, data requirements are determined based on legal and business reasons.

4.4.1 The Listing Process

The listing sub-process is conducted at company “D” by a data management team. This sub-process contemplates product data retrieval from GDSN data pools and data gathering from product data templates (in case that the manufacturer does not utilize GDSN data pools). Once the product data has been stored internally, the PMD is validated for data quality, and if the product data fulfills all requirements, then the product is activated in the company’s catalog to be purchased and sold.

The data management team triggers its activities with the request from the procurement team to process new given items. The listing process for new products currently depends on the product category, depending on the category there are specific time windows during the year where the new products can be listed. However, through the new design of processes this working method will be modified.

Lastly, the duration of the various activities from the listing process at Company “D” is not impacted by the product category or selection of the online or offline channel. Nonetheless, depending on the nature of the product, like promotional products, seasonal products, re-launch products, bundle products or standard products the data requirements and / or planning activities differ. For promotional products no shelf layout or planograms are developed, whilst for re-launch products the information on the predecessor-successor product is very important, likewise for sales activation programs or bundle products, where the reference to the standard product is of prime relevance.
4.4.2 Process Diagram and Attributes

The process flow depicting the data services of Company “D” can be seen in Figure 4.4, whereas the description, duration and attributes needed by the sub-processes are explained in Table 4.5.

![Process Diagram](image.png)

Figure 4.4 - Listing Process in Company "D"

For the correct interpretation of Table 4.5, in the column “Duration” when no content is found, it represents a milestone that does not involve processing time, unless it is indicated differently, for example “Information not available”. In the column “Attribute clusters” the attributes will be listed when they are used, if no content is found in the column “Attribute clusters”, it represents that no attributes are used in this sub-process. Additionally, in Figure 4.4 the elements depicted with a dotted border line represent the activities that are out of the area of the responsibilities and expertise from the interviewee.

<table>
<thead>
<tr>
<th>#</th>
<th>Sub-process</th>
<th>Description</th>
<th>Duration (wks.)</th>
<th>Attribute clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; Listing Meeting and further procurement activities.</td>
<td>Activities conducted by the Category Management and / or procurement team are not specified, however in this step is decided if the product will be listed.</td>
<td>- Information not available</td>
<td>- Information not available</td>
</tr>
<tr>
<td>2</td>
<td>Request for product</td>
<td>Major requirement for participating in the listing process is a contract with the</td>
<td>1&lt;sup&gt;*&lt;/sup&gt;</td>
<td>Identification, Classification, Hierarchy /</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>activation in</strong></td>
<td><strong>retail company or to be active in the retailer’s system. Request is raised for the data management team to activate a product, request includes information on product, product category and the target subsidiaries for the activation.</strong></td>
<td><strong>Indicators, Logistics data. Also internal information</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3 Product data retrieval</strong></td>
<td><strong>Product master data is retrieved from the GDSN data pool or product data templates. Data quality validations take place in this step, if errors are found in the data, communication for corrections with the manufacturers is started.</strong></td>
<td><strong>1 - 2 Party / Product Owner, Packaging &amp; returnable assets, Date, Description, EU regulation product label additions, Non id-attribute information, Additional logistics data, Dangerous goods, Tax &amp; Customs. Also internal information.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4 Further product data driven activities</strong></td>
<td><strong>This sub-process considers all activities performed by additional departments besides procurement and the data management team. For example, marketing or logistics. Thereby considering activities such as having printed media or planograms, however, some of these activities occur after the item activation, yet, before ordering the first shipment of product.</strong></td>
<td><strong>Information not available</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5 Activation of product in retailer’s systems.</strong></td>
<td><strong>If data quality is ok, the item will be activated to be purchased and sold. As part of the activation process, the product is allocated within an internal product category and to a specific set of subsidiaries</strong></td>
<td><em><em>1</em> Information not available</em>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(*) Used to denote those activities that last less than a week.
As it can be seen on Figure 4.4 and Table 4.5, sub-processes #1 and #4 could not be covered during the interview. Also, different sub-process as #2 and #5 last less than a full week. Taking this into consideration in addition to the total average duration for the listing process suggested by the interviewees, the partial duration, due to the missing information from sub-process #1 and sub-process #4, of the listing process has been considered to last between 2 and 4 weeks to guarantee robustness against most variations in the partial processes.

4.4.3 Data Quality and Quality Measures

In Company “D” the data quality is controlled through manual and automatic validations. Currently the company has a data quality monitoring tool which will be developed further. Company “D” tests all GTINs and all of their attributes. The major quality dimensions that the company controls are completeness, correctness (accuracy), consistency (formatting, units of measure…Etc.) and plausibility of the values through validation rules.
5. ANALYSIS OF RESULTS

The information obtained by the complete data collection strategy provided valuable insights on how the retail companies bring new products to end-consumers. This empirical study also showed how retail companies have embedded data validations and data corrections within their regular operations, even outsourcing these tasks to master data management service providers. Literature and research support the idea that data acts as fuel for businesses in the advanced digitalization era. Therefore, data quality management practices should be implemented within companies as well as across companies. However, all stakeholders should be aware of the requirements and capabilities of both ends in order to define feasible business practices.

The results of this study show how the participating retail companies pursue diverse strategies through various channels and utilizing different systems to reach end-consumers. This variety impacts the retailers’ activities for commercializing products, and thus their data requirements. Furthermore, information provided by a retail company revealed that more than the half of the product master data shared from manufacturers to retailers is B2C oriented, whereas all the data is transmitted through a B2B flow. This fact gives an initial hint on the possible causes why inter-company product data communication is a challenging process to streamline. For a visualization aid to this fact, see Figure 5.1.

Figure 5.1 - Cross-company Product Data
This chapter is structured in two sub-chapters, firstly describing the analyses performed to the listing process documented from the four participating companies. In this section a proposal for a generalized listing process can be found, this information will be used as basis to analyze the suitability of the proposed solution by industry experts. Secondly, the feasibility of Preliminary Trade Item as an approach for data quality is analyzed and timing specifications for its implementation are suggested. In both sub-chapters the corresponding methods used for data analysis are provided in order to give transparency to the proposals.

5.1 The Listing Process

The listing process as explained in chapter 4 comprehends all activities driven by product data from the first product introduction in the 1st listing meeting until the product (or item) has been activated in the retail internal system to be ordered and sold. As it can be seen in Figure 5.2, different companies conduct different processes.

![Figure 5.2 - Listing Process Compilation](image-url)
Through mapping the listing process, research sub-questions #2, #3 and #4 can be answered. This approach is customer-centric since firstly defines what the customer data requirements are based on, as well as, the activities and attributes used during the listing process. Secondly, it measures the time required to conduct the business processes to calculate the complete duration. Thereafter, idle times for particular data attributes can be identified and alternatives for improving the quality of the data can be sought based on a similar delivery of data as in “Just in Time”, delivering the data at the point in time when it is needed.

Nonetheless, a solution for the research problem, described in sub-chapter 1.3.1, can only be feasible if it is implementable across companies through standard business practices. Hence, the next challenge for the research was to find a generalized representation of the listing process based on the information gathered from the participating companies that is robust enough to be valid across channels and companies, by considering activities, attributes and durations.

5.1.1 Data Analysis

The listing process has been analyzed per channel in order to make the activities comparable. The first companies to be reviewed serve offline and online commerce, therefore, analyzing the operations from Company “A” and Company “D”. However, the information acquired from Company “D” does not reflect the entirety of the company’s listing process, merely the activities that manage the product data in the systems, without inter-departmental interactions, and thus excluding procurement or category management, marketing and logistics, and possibly other areas.

Company “A” and Company “D” serve customers in online and offline channels and both offer products from the FMCG industry, with same product categories although their overall product portfolios differ. Also, Company “D” is re-designing its data management processes to include a master data service provider, as Company “A” does. Therefore, similarity of both listing
processes and requirements can be expected. Since the information of the inter-departmental interactions of Company “D” are missing, an extrapolation of the available information from Company “A” will be conducted, yet, the available information from data management of Company “D” will be kept. Please see Figure 5.3 to obtain an overview from the extrapolation, comparisons can be made with Figure 5.2.

**Figure 5.3 - Extrapolation from Company "A" to "D"**

In order to generalize the duration of the processes starting from the 1st Listing Meeting till the product is ready to be sold, the maximum and minimum duration times per channel and companies were analyzed and then a unified duration per channel was proposed. See Table 5.1 for a comparison between timings across companies and about the timing generalization. Nonetheless, due to the nature of the processes from Company “C” as a master data service provider its timings have not been included in this comparison. However, Company’s “C” timings have been considered as part of the timings from Company “A” and Company “D”. Further considerations in the generalization of the timings for the listing process can be found under the table.
Table 5.1 - Proposed General Durations for Listing Process

<table>
<thead>
<tr>
<th>Channel</th>
<th>Company</th>
<th>Min. Duration (wks.)</th>
<th>Max. Duration (wks.)</th>
<th>Proposed General Duration (wks.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offline</td>
<td>Company “A”</td>
<td>15</td>
<td>17</td>
<td>15 to 19</td>
</tr>
<tr>
<td></td>
<td>*Company “D”</td>
<td>17</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>Company “B”</td>
<td>4</td>
<td>6</td>
<td>7 to 8</td>
</tr>
<tr>
<td></td>
<td>**Company “A”</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

(*) Timings from Company “D” were modified according with the extrapolation of activities.

(**) There is sufficient information from Company “A” to assess its operations from a pure online perspective.

As it can be seen in the offline channel section from Table 5.1, the timings of Company “D” last longer than the timings of Company “A”, although most of the process duration of Company “D” were extrapolated from Company “A”. This is justified due to the uncertainty caused by the extrapolation of sub-processes and the re-design stage of data processes in Company “D”. Therefore, in order to offer a robust time range two additional weeks were considered.

The proposed general durations for both offline and online listing processes were planned to include the longest duration of all companies in each channel, and thus offer inclusive planning times. For Company “B” in online commerce there were significant variations for the duration of the 1st listing meeting, therefore the proposed general duration for online commerce is above the minimum and maximum timings of Company “B”.

In order to generalize the sub-processes performed by the different retail companies, the most common activities where categorized within the following actions:

- Product presentation
- Data provision and retrieval (Internal data and product data)
- Product data validation and correction
- Item creation and activation
- Planning of activities and development of analyses and calculations
Lastly, the attribute clusters were assigned to the generalized sub-processes. A complete data cluster was allocated to the first sub-process that used any of the attributes comprehended within that cluster. For example, if different attributes from the data cluster “EU regulation on product label additions” were used in sub-processes #3 and #5, the complete data cluster would have been allocated to sub-process #3, despite the number of attributes of that data cluster used in sub-process #5. Thereby, fully enabling the operations of the retail company.
5.1.2 Standardized Listing Process

The “One size fits all” generalized listing process is valid for offline and online commerce in regard of the activities being performed and the attributes needed. A graphic representation of this process can be seen in Figure 5.4:

![Generalized Listing Process Diagram](image)

**Figure 5.4 - Generalized Listing Process**

The activities covered within each of the sub-processes are described below, each of the number of the sub-processes matches the numbering of the descriptions. From process number (4) onwards continuous data validations take place once the data is used within its context.
(1) **1st Listing Meeting**: Presentation of the new product. Negotiations about the terms and conditions (for example, packaging form (Shelf-Ready), units per package, Prices...Etc.). Acceptance of the new product for the listing. Prerequisite: Established cooperation between retailer and manufacturer (for example a contract).

(2) **Provision of product data**: The manufacturer transmits the product data via Templates, Provider-Online-Platforms or GDSN Data Pools. In case of questions or technical problems the manufacturer contacts the retailer or the Master Data Service Provider and vice versa. Generally, the first automatic data quality validation takes place in this step.

(3) **Validation of product data**: Manual / automatic data quality validations are performed. In case of discrepancies, the manufacturer is contacted. If corrections / additional data is needed the process repeats from sub-process #2 after the disposition of the new data.

(4) **Processing of the data**: Creation of the retailer specific Item-ID for the new product (Item is not yet activated). Retailer internal calculations are conducted (for example, profitability, bundling options...Etc.). Allocation per sales channel (Pure online, Brick & Mortar or both). Selection of marketing channels (for example, printed media / online media). In case of discrepancies, the manufacturer is contacted. If corrections / additional data is needed the sub-processes #2 and #3 repeat after the disposition of the new data.

(5) **Access and validation of additional data**: Additional data (exempli gratia, retailer specific data, and retailer internal data or additional GDSN-Data...Etc.) is retrieved via internal / external Data pools and automatically / manually validated for data quality (Completeness, Accuracy and Consistency). In case of discrepancies, the manufacturer is contacted. If corrections / additional data is needed the sub-processes #2 and #3 repeat after the disposition of the new data.

(6) **Planning of different activities**: Validation of the additional data to perform different activities (exempli gratia, handling of dangerous goods, compliance of data contents with guidelines, allocation of the product per category and subsidiary, Demand Planning, Shelf planning...etc.). In case of discrepancies, the manufacturer is contacted. If corrections / additional data is needed the steps #2 and #3 repeat after the disposition of the new data.
(7) **Milestone:** Activation of the Item-ID in the retailer’s system (New product is orderable / sellable).

(8) **Supply Chain Management:** Orders are placed. Warehouse planning (Centrally, regionally or per subsidiary).

Additionally, the durations of the listing process have been standardized. However, due to the significant differences in the complete duration of the listing process per channel. Online and offline channels have been mapped separately. For detail information about the time factor in the offline channel see Figure 5.5 and for the online channel see Figure 5.6.

**Figure 5.5 - Time Factor Offline Commerce**

In offline or “brick and mortar” commerce the complete duration ranges from 15 to 19 weeks as justified in sub-chapter 5.1.1. However, in Figure 5.5 only the minimum duration is depicted. The main step that extends the duration of the listing process is sub-process #4 - Processing of the data, due to printed media since its lead time lasts approximately 12 weeks. Printed media is considered part of the advertising activities, which are normal practice for the introduction of new products.

**Figure 5.6 - Time Factor Online Commerce**
In online commerce the complete duration ranges from 7 to 8 weeks as justified in sub-chapter 5.1.1. However, in Figure 5.6 only the minimum duration is depicted. The main reason for online commerce to last about the half of the time compared to offline commerce is that activities as printed media are not needed since posting advertisement of a product in a digital platform has no processing or waiting times.

Lastly, the sub-processes that use attributes can be recognized in Figure 5.7 by the red glow around the number of the sub-processes.

**Figure 5.7 - Attribute Usage in the Listing Process**

The attribute clusters that are used with its corresponding sub-processes can be found in the list below. For more information regarding the description of each attribute cluster, see Table 3.3.

- **(1) 1st Listing meeting**: Dummy / Mock-up (for example, prototype of the new product). Units per packaging and similar informations regarding functional description of the new product and Prices (all attributes in this sub-process are non GDSN attributes).
- **(4) Processing of the data**: Classification, Identification, Logistics data: Dimensions & Temperature, Prices (suggested Retail / price, catalogue Price / basis Per Unit and Net-price - these attributes are not considered in the current attribute clusters), Taxes, Party / Product owner, Description, Date, Hierarchy / Indicators, EU Regulations / Product label and Packaging and returnable assets.
- **(6) Planning of different activities**: Additional logistics data - Transport Units, Dangerous goods and Customs.

Nonetheless, different challenges were encountered while working with the attribute clusters, for example, the lack of product category specification within the clusters delimits its usage,
since not all attributes are the same within product categories. Additionally, the logical grouping of attributes was not the best solution to identify which attributes belong to which sub-process since most of the times, several attributes within the same cluster were scattered in various sub-processes. Therefore, business process oriented groupings should be developed to replace the logical grouping approach.

5.2 Preliminary Trade Item: How soon is soon enough?

As described in sub-chapter 1.3.2, the challenge in data communication between manufacturers and retailers involves that the dataset is requested to be complete, final and sent at an early stage of the listing process. At this time manufacturers do not have all required data in a structured format and final approved content available.

PTI is a technical functionality enabled in the GDSN data pool from 1WorldSync that allows manufacturers to flag GTINs as preliminary. However, PTI has not been properly defined as a concept supported by business rules. Therefore, PTI could be considered in two different ways regarding the exchange of product data. Firstly, as a functionality to send an incomplete dataset marked as “preliminary” with only the product attributes that are final at an early point in time, and then flagging the product dataset as “final” when the missing attributes have become final and have been shared.

Secondly, PTI can also be defined as a functionality to mark a complete dataset as “Preliminary”, sending all available product attributes at an early point in time. However, all attributes would be subject to change until the manufacturer changes the flag to “final” at a later point in time. With this approach more information would be sent at an early point in time compared to the first approach. Nonetheless, it would have to be define if “Draft” data is useful to the retail companies, and which data might be useful as draft.
The first approach of PTI considers sending solely final data, and therefore correct data. Hence, this study has considered the first approach as the solution to be tested. Nonetheless, the second approach of PTI could offer benefits to the retail companies, therefore, might be valuable to consider it for further research. Based on the generalized listing process and having defined PTI as a two-phased delivery of final product data. Now it is possible to clear the question regarding the point in time when data is timely delivered, based on the point in time when the data is needed and the data attributes that are needed.

In Figure 5.8 a visualization of PTI’s framework can be seen. The approach for PTI involves requirements that in turn originate challenges to attain benefits. The core of PTI is assuring data quality through the provision of product data in the correct point in time. A detailed information chart to this framework stating the perspectives of the involved parties along the data communication chain, manufacturer, a potential data quality instance, data pool and retailer can be found in Appendix - 4: PTI: Challenges - Requirements - Benefits.

Figure 5.8 - PTI: Requirements - Benefits - Challenges
5.2.1 Data Analysis

Based on the answers obtained through the interview, the participating companies responded that they retrieved the product data in more than one occasion from the GDSN data pool to obtain a final correct dataset. The number of times can be seen in Table 5.2.

Table 5.2 - Number of Times Data Is Retrieved

<table>
<thead>
<tr>
<th>Company</th>
<th>No. of times data is retrieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company “A”</td>
<td>From 2 to 3 times</td>
</tr>
<tr>
<td>Company “C”</td>
<td>In average 2 times</td>
</tr>
</tbody>
</table>

Although not all participant companies responded the question, Company “C” as a master data service provider, answers in average for all of the retail companies that Company “C” manages as clients in the FMCG industry. This fact shows clearly that there is a mismatch between the point in time when the data is provisioned (and requested) and the point in time when the data is correct (final).

Additionally, as shown in sub-chapter 5.1.2, not all product attributes are needed from the beginning of the listing process. Hence, the option of a two-phased delivery of data from PTI is feasible. Thus, there are three sub-processes in which the two data deliveries can be distributed, see options below:

- **Sub-process #1: 1st Listing meeting** – 19 to 15 wks. prior to the first shipment.
- **Sub-process #4: Processing of the data** – 12 wks. prior to the first shipment.
- **Sub-process #6: Planning of different activities** – 10 wks. prior to the first shipment.

The number of weeks prior to the first shipment can be seen in Figure 5.5. Only the offline commerce “weeks to first shipment” have been considered, since they represent the longest duration times. Therefore, if the times for offline commerce can be fulfilled, the time windows for data delivery in online commerce (no longer than 8 weeks) will be met.
5.2.2 Approach to Preliminary Trade Item

As stated in the research problem in sub-chapter 1.3.1, retail companies request for the complete product data between 6 and 26 weeks prior to the first shipment. Based on the information gathered from four participating companies active in the retail industry, there is no evidence that suggests that the product data is needed with more than 19 weeks in advance nor that the complete data set is needed at this point in time.

Nonetheless, the information collected through the empirical study also showed that the current error rate is high. The product master data communicated to the retail companies has a perceived error rate between 30 percent and 70 percent of the GTINs when all product attributes are considered across the FMCG industry. However, this is a perceived value from the interviewees since there is no historical data available, additionally, the data error rate varies among manufacturers and product categories.

Additionally, literature shows in further detail the level of quality for product data, categorizing it by consumer units and traded units. A trade unit comprehends regularly more than one consumer unit and are the units used for commercializing products in a B2B basis, whereas consumer unit is used for a B2C basis. The findings in the literature describe that in consumer units 51 percent of the products contained inaccurate data, whilst the trade units presented 34 percent of product data inaccuracy (Capgemini, 2004).

Based on the previous facts about the overstated anticipation for the data delivery and the current error rate for product data, the given information suggests that retail companies plan for delays in their listing processes, as well as they have data validations embedded in their operations. Hence, the retail companies include time buffers while requesting for the data. Moreover, this approach of including time buffers separates the product data even more from the point in time when the content is proven to be finalized on the manufacturer side, which is
when the physical production of the consumer goods start. Consequently, causing the data to be more prone to errors, as in the vicious cycle depicted in Figure 5.9.

![Figure 5.9 - Vicious Circle of Defective Data](image)

The proposal for PTI begins with separating offline commerce from online commerce for the data delivery, since the duration of both listing processes is significantly different. Nevertheless, manufacturers do not differentiate per channels for the data disposition, they share all data with the retailers to enable their operations regardless of the channel. Nonetheless, this study covers both channels separately to assure that both requirements are covered when a PTI proposal is made.

**Offline commerce:**

- For offline commerce a first delivery of data is proposed to take place 19 weeks prior to the first shipment, which would allow the retail company to start all of its operations considering the maximum duration of the listing process, and it is in alignment with the 1st Listing meeting.

- Subsequently, a second and final delivery of data should take place 12 weeks before the first shipment. This timing for the second delivery of data has been considered to be in alignment with sub-process #4, where activities such as organizing printed media occur, and thus providing enough time to the retailers to have complete and final data to be included in their printed media.
• Also, this approach gives the manufacturer seven weeks of time to finalize its data from the first delivery.

**Online commerce:**

• For online commerce, the longest processing time lasts 8 weeks. However, based on the approach for offline commerce, a complete final data set will be available up to 12 weeks prior to the first shipment. Therefore, the data will be available earlier than 8 weeks prior to the first shipment, and thus the same 12 weeks have been proposed for online commerce.

• Likewise, there is no preliminary delivery of data since the duration of the entire process is shorter and delivering product data earlier than 12 weeks in advance to the first shipment is not needed.

Referring to the same visual aid used to explain the research problem, the proposal for PTI has been placed in Figure 5.10.

![Figure 5.10 - PTI Proposal](image-url)
In order to align the data deliveries with the business processes from the retailers, the disposition of the final dataset has been proposed to be sent 12 weeks prior to the first shipment. As it can be compared with Figure 1.4, the current point in time to send the final data to the retailers takes place between 8 to 6 weeks prior to the first shipment. The proposal for sending the final data 12 weeks prior to the first shipment, pushes the finalization of data four weeks earlier than how it currently is, fostering the manufacturers to streamline their innovation processes and align with the retail companies in a customer-centric approach. However, there is a compromise in both ends, since retail companies would have to process data retrievals two times.

Regarding the provision of attribute clusters per sub-process, the attribute clusters should be grouped in preliminary and final according to the data clusters needed at that point in time. The approach taken by using logically grouped attributes proved to be inflexible to the real business oriented usage of attributes. Previously, documentation about the listing process and its steps was not available, and therefore the development of attribute clusters based on business processes was not possible. However, after this study, the attribute clusters can be re-designed and tailored to be product category responsive in order to define which attributes should correspond to the preliminary dataset and to the final dataset.

Furthermore, in order to successfully implement PTI a technical solution as well as a conceptual solution through different factors should be considered. These factors include appropriate business rules and defined responsibilities. This is collaborative task as well, since clear business rules are the pre-requisite to successful trade operations.
6. CONCLUSIONS

Increasing technology and the digitalization of several areas of life and economy have brought advancements to society and industry. However, with new disciplines or the virtualization of existing disciplines new challenges appear, such as the digitalization of products and the omni-presence of retail companies. Additionally, customers’ preferences change rapidly and due to the myriad of information available in any smart device, customer expectations have grown significantly. Hence, companies in the FMCG and retail industry have to increase their responsiveness to evolving market requirements, which can only be attained through close collaboration.

The aim of this study is to determine how manufacturers in the FMCG industry can meet the required data quality for its product master data on time to enable omni-channel commerce for its trading partners. As a result from extensive academic and empirical research the aim of this study can be addressed. In order to cover the main research question in a structured manner, the sub questions will be answered:

(1) SQ-1: How can data quality be achieved for product master data?

As it has been defined, data of quality is data that fits for the use that it is intended for. Therefore, data owners should be aware of the applications and interfaces that their data will be used for, to understand the data requirements and adapt the data to this purpose. Once the fitness specifications for the data have been defined, the data has to be kept in accordance to these specifications. For this purpose there are several practices of Data Quality Management that can be implemented, mainly categorized in preventive quality measures and reactive quality measures.

Reactive quality measures imply taking corrective actions to adjust the level of quality of the data to meet the fitness specifications, a common practice to perform corrective activities is to prioritize the data object to be corrected by business impact criticality. Preventive quality measures include data governance that adapts to the company business
strategies and data metrics. The main data quality metrics or dimensions are correctness, availability, completeness, timeliness and consistency. Within the framework of preventive quality measures, Data Quality Management intersects with practices from Master Data Management. Which applies data quality principles to master data objects and classes.

Therefore, to achieve data quality for the product master data class, the requirements for data fitness have to be cleared in answering what is correct product data, when should the data be made available, what is a complete product dataset, what is current or timely valid product data, and lastly, in which format the product data should be shared. Based on the context of this study, the only open quality dimension with unspecified requirements is the temporal quality dimension of availability.

(2) **SQ-2:** What activities are performed by the retailers in order to list a new product and what is the duration of those activities?

Empirical research was conducted with four leading companies active in the retail industry in Germany, thereby, the listing process and its activities with their durations have been defined and generalized across companies. However, the duration of the complete listing process is rather a time range than a fixed value, and therefore flexibility of interpretation for the times should be given as long as they fit within the described range. The calculated duration range for offline commerce is between 15 and 19 weeks, whereas for online commerce is between 7 and 8 weeks. The sub-processes or activities that entail the Listing process are:

1. **1st Listing Meeting:** Approx. duration 1 - 3 wks.
2. **Provision of product data:** Approx. duration 1 - 2 wks.
3. **Validation of product data:** Approx. duration 1 to 2 wks.
4. **Processing of the data:** Approx. duration 12 wks.
5. **Access and validation of additional data:** Approx. duration 1 wk. (parallel task)
6. **Planning of different activities:** Approx. duration 2 wks. (parallel task)
(3) **SQ-3: What product attributes are required for each of the activities or processes involved in listing a new product?**

Product data clusters have been developed to answer this question. The data clusters fulfilled their purpose to identify the important logical groups of attributes per activity in the listing process. However, this format of data clusters has not been useful for the proposed solution of Preliminary Trade Item. Thereby, further refinements are needed. Using the numbering for the six steps of the listing process the attributes clusters per activity are:

1. **1st Listing meeting:** Dummy / Mock-up (for example, prototype of the new product). Units per packaging and similar information regarding functional description of the new product and Prices (all attributes in this sub-process are non GDSN attributes).

4. **Processing of the data:** Classification, Identification, Logistics data: Dimensions & Temperature, Prices (suggested Retail / price, catalogue Price / basis Per Unit and Net-price - these attributes are not considered in the current attribute clusters), Taxes, Party / Product owner, Description, Date, Hierarchy / Indicators, EU Regulations / Product label and Packaging and returnable assets.

6. **Planning of different activities:** Additional logistics data - Transport Units, Dangerous goods and Customs.

(4) **SQ-4: What data requirements differ per channel (online & offline)?**

The data requirements have proved to be the same across channels. The main difference is the time needed to complete the listing process and the importance of specific attributes. For example, for online commerce and food products, the Food Information Regulation Attributes are of utmost importance, whereas regardless of the product category, the images are relevant for online commerce. Additionally, depending on the nature of the new product, seasonal, promotional, bundle product, sales activation, predecessor-successor or standard product, the reference to the original or standard product is critical.
(5) **SQ-5: What is the current status of data quality for product master data sent by the manufacturers?**

The current status of product master data varies depending on the manufacturer and the product category. Nonetheless, the perceived product data error rate is high when all attributes and quality metrics are considered per GTIN. Since there are no historical records the interviewees shared their perceived value of the error rate which ranges from 30 percent to 70 percent. Additionally, based on the multi-method approach designed for this sub-question, reviewed literature confirmed a range of inaccuracy for data between 34 percent and 51 percent. Thereof, the inaccuracy range varied depending on the level or hierarchy of the products packaging (consumer unit or trade unit). Nonetheless, the current amount of product attributes has increased significantly based on evolving regulations and market requirements from the point in time when the referenced study was conducted, which augments the complexity of data communication and thus the possibilities of data inaccuracies. Additionally, as shared by some interviewees the error rate on attributes gathered from templates is higher than the error rate of data collected from a GDSN data pool.

### 6.1 Theoretical Contribution

The theoretical contribution of this study begins with its focus on cross-company data exchange. Whilst this research area has been widely covered for transactional data. For example, retail companies sharing their sales figures with manufacturers to improve their demand planning, there are just a handful of scientific studies in this field considering the exchange of master data needed to allow operational processes to be conducted.

Secondly, this study explores the temporal data quality dimensions under a new perspective. Namely, temporal quality dimensions have been divided in two main concepts: Availability and timeliness. Furthermore, this study does not consider solely if data was made available on time, but also elaborates further on defining the impact of data disposition on other quality dimensions like correctness depending on the point in time that the product master data was
made available, which has not been considered by the literature on data quality prior to this study.

6.2 Managerial Implications and Recommendations

The business impact of information exchange between companies has been focused mainly on the impact of data communication to reduce costs in the supply chain, reduce uncertainty for new product introductions or reducing the time-to-market for new products. However, the exchange of master data for firms and practitioners has a significant business impact as well, for example: If legally required information about dangerous goods or allergens for products is wrong or absent, compliance fees might be charged to the manufacturer or the reputation of the company might be in risk. Similarly, if product morphological data like dimensions is wrong or was not shared on time with the retail companies, a business opportunity might be lost due to delays, fees could be applied from the retail companies due to poor data quality, and generally, frictions between both companies might be caused.

Poor master data quality causes companies to waste time and resources in data validations and correcting data. The aim of this study is to increase the data quality by harmonizing the manufacturers’ capabilities to deliver data with the retailers’ requirements on data. The results of this thesis propose implementing the PTI concept as a two-phased delivery of final data attributes, which in principle would increase the tasks performed by the retail companies, since they would have to process data twice. However, due to poor data quality, currently retail companies already have to retrieve the data to include corrections between two and three times. Therefore, this solution alleviates the work of performing validations and corrections since the data sent would be final, and consequently with the reduced time invested in validating data and correcting it, the listing process duration will be reduced.

The implementation of PTI involves a compromise from both parties, manufacturers and retail companies. Since the manufacturers would have to assure the quality of the data and streamline their internal processes to meet the demanded time windows to deliver the
preliminary and final datasets (GTINs). Likewise, retail companies would have to adjust their internal processes to operate with a two-phased data delivery.

The metrics for data quality should be able to include Preliminary Trade Item seamlessly, since depending on the attributes sent as part of the preliminary dataset, the appropriate validation rules for those attributes should be conducted, and for the final dataset all the validation rules can be applied. Additionally, an availability or timeliness metric can be incorporated based on the delivery times arranged as part of PTI, to evaluate if the data was delivered on time. A possible reference that can be used to measure the punctuality of the data delivery considers the availability date of the product to be ordered, the date in which the dataset (or GTIN) was made available in the data pool, and in case the dataset was modified, the date when the modifications took place.

Lastly, the use of retailer specific product attributes (attributes not considered in the GDSN pool) will increase as forecasted by one company during the interviews. Therefore, fostering the mapping of GDSN attributes with retail specific attributes should be a priority for the industry in order to avoid the usage of third party platforms or filling templates manually. Thereby, reducing data quality inaccuracies due to the data transmission method.

6.3 Limitations

Due to the selected target market (Germany) for the research study the results cannot be generalized to other countries. Also, this study has considered only four retail companies, hence, the sample of retail companies is not representative for all the retail companies active in Germany. However, the four companies fulfilled the explorative and descriptive research goals of the thesis. Additionally, the information regarding the product innovation process only considered the sponsoring organization. Nonetheless, the study focused on gathering the data requirements from the retail companies, therefore, the input from various manufacturers in the FMCG industry were not needed to achieve the goals of this research.
Lastly, the scope used for collecting the needed attributes per activity in the listing process did not consider product categories. Also, the logic grouping of the clusters could not offer accurate information on which attributes should be used for a preliminary dataset and a final dataset.

6.4 Further Research

Further research should evaluate if draft attributes (subject to change) are useful for retail companies. Also, more manufacturing companies and retailers should be involved to analyze the suitability of the proposed Listing process and timings for a preliminary and final dataset provision.

Also, the disposition of product images and their metadata should be object of further analyses, since they have not yet been considered for GDSN data communication in Germany but are a major part for online as well as offline commerce. The further definition of product images guidelines could enable online retailers to start their operations earlier with a preliminary or prototype picture of the product.

Additionally, data quality metrics should be defined for considering the temporal factor since other quality dimensions cannot be evaluated unless they are based on final data. Consequently, quantitative research should be performed on the impact of the time factor exercised on the quality dimensions in the FMCG business.
REFERENCES


Caplice, C. (2016a), CTL.SC1x Supply Chain Fundamentals: Introduction to Logistics and Supply Chain Management.


Verdugo Cedeño, J.M. (2016), Developing Smart Services by Internet of Things in Manufacturing Business, Lappeenranta University of Technology.


APPENDICES

This section contains different appendices with the purpose of illustrating concepts in detail or to show elements used as part of the empirical study.

Appendix - 1: GTIN - GPC - Attributes - Values

Source: GS1, 2016e
Appendix - 2: Data and Data Quality Conceptual Model

Source: Otto and Österle, 2015
Appendix - 3: Questionnaire for Interviews

1) Which sales channels do you serve (i.e. Online-trading, stationary trade…etc.)?

2) How is the Procurement / Procurement process structured (i.e. centralized, decentralized, regional…etc.)?

3) Describe briefly the listing process of a new product in your company.

4) What are the main pre-requisites for the start of the listing process?

5) From an end-consumer and retailer perspective, what are outcomes of the listing process for each sales channel (i.e. Online catalog of a new product, efficient planning of the shelf…etc.)?

6) What resources (i.e. Persons, Systems…etc.) are needed / involved in order to perform the listing process?

7) What / Who (External and internal influences / limitations) control or limit the process of listing a new product?

8) How often each year does the process for the listing of a product takes place for all sales channels (i.e. Online-trading, stationary trade…etc.)?

9) When does the process for the listing of a product takes place for all sales channels (i.e. Online-trading, stationary trade…etc.)?

10) Do you use a system-supported workflow manager for the listing process (i.e. Workflow tools)?

11) Which business processes / sub-processes are required to carry out the listing of a new product (i.e. Data creation, Category Management – Shelf layout…etc.)?

12) Which sequence do these business processes / sub-processes follow (i.e. Sequential or parallel)?

13) Which departments are involved in the sub-processes of the listing process for new products (i.e. Process responsible / Process participant)?

14) Where do the involved departments get their data to perform the business process / sub-process from?

15) Which product attributes (data clusters) are needed to complete each sub-process?

16) What is the duration range (Min and Max: i.e. 1 week to 3 weeks), and average (i.e. 1 week) for each business process / sub-process from the product listing of a new product?

17) To which extend do sub-processes for the listing of a product differ for each sales channel (i.e. Online-trading, stationary trade…etc.)? (If applicable)

18) Which variations of the sub-processes can occur based on the product type / product use? See table below. The goal is to identify possible requirements / differences in the lead times.

<table>
<thead>
<tr>
<th>Product type / product use</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>Long term item</td>
</tr>
<tr>
<td></td>
<td>Re-launch with predecessor GTIN</td>
</tr>
<tr>
<td>Retailer - Event driven product</td>
<td>Promotional item</td>
</tr>
<tr>
<td></td>
<td>Seasonal item</td>
</tr>
</tbody>
</table>
19) How many new FMCG-GTINs are monthly received (Time definitions from Questions #8 and #9)?

20) Do you perform product data quality controls? (If yes; Questions #21 to #27, if no, continue with question #29)

21) Are the data quality controls manually or automatically performed?

22) How many employees control the quality?

23) How do you determine which GTINs should be controlled?

24) Which attributes / attribute clusters are controlled?

25) Which quality dimensions are controlled (i.e. Completeness, accuracy…etc.)?

26) How much time is invested for the quality control of each GTIN?

27) What is the average defect rate found through the quality controls (if applicable, explain how the rate is measured)?

28) How many resources (Time / Activities) are needed to correct the data with the manufacturer (Information process, correction process, new control of data quality…etc.)?

29) In average, how often does Data from a new GTIN has to be retrieved from the 1WorldSync-Pool, in order to have a complete and correct dataset for the first time in the retailer system?

30) How would you quantify the advantages of a controlled data communication process (Two communications versus “X-Communications”)?

31) Which benefits would controlled product data quality bring you (Name one or more benefits based on your priorities)?

32) How would you quantify the benefits of product data quality?

33) Which questions would you add to complement this questionnaire from a retailer view?

34) How would you complement / re-design the product attribute clusters?
### Appendix - 4: PTI: Challenges - Requirements - Benefits

<table>
<thead>
<tr>
<th><strong>CHALLENGES</strong></th>
<th><strong>REQUIREMENTS</strong></th>
<th><strong>BENEFITS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>First time right approach.</td>
<td>Identify cost’s root cause &amp; actors.</td>
<td>Supply Chain &amp; Dataset communication cost reduction.</td>
</tr>
<tr>
<td>Harmonize global and regional data policies.</td>
<td>Data quality assurance &amp; data protection.</td>
<td>Increased customer (B2B) satisfaction by serving Omni-channel requirements.</td>
</tr>
<tr>
<td>Automate data maintenance.</td>
<td>Continuous measurement of data status via standard metrics (KPIs).</td>
<td>Streamlining of internal processes.</td>
</tr>
<tr>
<td>Include different business model operations.</td>
<td>Operations road map for retail &amp; manufacturer.</td>
<td>Tailor made services for retail &amp; manufacturer.</td>
</tr>
<tr>
<td>Meet all stakeholders requirements &amp; capabilities.</td>
<td>Smooth implementation steps for &quot;DQ Gate&quot;.</td>
<td>Increased DQ Gate adoption.</td>
</tr>
<tr>
<td>Address Preliminary Trade Item’s complexity via stakeholders synergy.</td>
<td>Include time-phased data availability in DQ scope.</td>
<td>PTI enabling technology to broaden view on DQ.</td>
</tr>
<tr>
<td>Approach cross-industry product complexity &amp; multiple PIM systems.</td>
<td>Robust, flexible &amp; scalable platform</td>
<td>Robust DQ Gate.</td>
</tr>
<tr>
<td>Find root causes of data corrections &amp; develop standard preventive plans.</td>
<td>Ad hoc data standards for special products.</td>
<td>Smooth data flow through the network.</td>
</tr>
<tr>
<td>Harmonize business functions with two-stage data deliveries.</td>
<td>Identify product data requirements / Attribute sets per business process and dependencies.</td>
<td>Less corrections performed on the network.</td>
</tr>
<tr>
<td>Mapping time-phased data requirements from different supply chain strategies (Omni-channel).</td>
<td>Enable business processes to start with an incomplete data set.</td>
<td>Broader acceptance / trust of GDSN by data users.</td>
</tr>
<tr>
<td>Coordinate all stakeholders requirements &amp; capabilities for SLAs.</td>
<td>Define common workable SLAs with manufacturer, DQ instances &amp; GDSN.</td>
<td>Data delivery tailored to retail business processes.</td>
</tr>
<tr>
<td>Stop internal DQ validations.</td>
<td>Transfer DQ assurance responsibility to manufacturer or external DQ Instances.</td>
<td>DQ: Accuracy, Completeness, Consistency &amp; Timeliness.</td>
</tr>
</tbody>
</table>

**PRELIMINARY TRADE ITEM**
- Manufacturer
- Smart DQ instance: GDSN
- Preliminary Trade Item
- Retail