

School of Industrial and Information Engineering  
Department of Management Engineering



**POLITECNICO**  
**MILANO 1863**

Data Productivity measurement in the Industry 4.0 scenario:  
Theoretical modelling and application in Supply Chain Planning

Supervisor: Prof. Giovanni Miragliotta

Co-supervisor: Prof. Massimiliano Schiraldi

Ing. Elisa Convertini

Authors:

Maria Elisa Piciacchia 877408

Cristina Pizzadili 875886

Academic year 2017/2018



# Abstract - ENG

In the new era of innovation with the hype of the Industry 4.0, businesses are facing tough challenges to succeed in a global competitive market. Indeed, organisations are digitising and integrating essential functions within their internal vertical processes, as well as with their horizontal partners along the value chain. To proactively respond to these challenges, organisations' management requires up-to-date and accurate performance information relatively to their businesses. Until now organisations following the saying "you can't manage what you don't measure" accustomed to the measurement of productivity of machinery, materials, workers and capital to control. Nowadays smart leaders across industries clearly feel the extent of the data management revolution and the actual potential of data driven decision-making. Considering this scenario and the increased relevance of data for organisations, it is clear the necessity of an appropriate and comprehensive data productivity measurement. Thus, the proposal of this Master thesis work is the definition of a metrics able to understand the current level of data exploitation in decision-making, enabling managers to improve their awareness and translating the gained knowledge into enhanced decision-making. The development of this index is done with the perspective to provide a quantitative measurement comparable between different companies and characterised by a good applicability in different sectors. The study is inspired by a previous work of Miragliotta et al. (2018) and principally relies on the parallelism with the traditional OEE (Overall Equipment Effectiveness) framework. The data productivity is broken into three main dimensions: data availability, quality and the performance of decision support system processing those data. Thanks to a thorough literature review multiple sub-dimensions have been identified together with an assessment methodology. The model is built following a maturity model approach, leading to the final computation of a numeric index. The path taken by this thesis work is towards the standardisation of the metrics proposed, with the aim to ensure the model transposition in business context, achieving a universal benchmark. The application in companies of different sizes belonging to several industries acknowledged the model to represent a good reference for organisations to assess data exploitation in decision-making and benchmark it against other business realities.

# Abstract - ITA

Nell'era dell'Industria 4.0 le realtà aziendali si trovano ad affrontare nuove sfide per affermarsi nel mercato competitivo globale. Le imprese si stanno muovendo verso una maggiore digitalizzazione, che permette da un lato l'integrazione di tutti i processi interni e dall'altro il coinvolgimento e la trasformazione dell'intera filiera produttiva. Per reagire prontamente a queste nuove sfide, le imprese devono essere in grado di monitorare in modo accurato e aggiornato le performance aziendali. Fino ad ora seguendo il detto "you can't manage what you don't measure", le aziende sono state capaci di misurare la produttività di macchinari, materiali e lavoratori. Oggigiorno i leader delle diverse industrie avvertono chiaramente la portata della rivoluzione digitale e la conseguente potenzialità rappresentata dalla possibile ottimizzazione dei processi decisionali guidati dai dati. In questo scenario, caratterizzato da una sempre maggiore rilevanza dei dati, diviene evidente la necessità di definire una metrica appropriata per misurare la produttività del patrimonio informativo aziendale. Pertanto, l'indice è stato sviluppato nell'ottica di offrire alle diverse compagnie una misura quantitativa confrontabile e facilmente applicabile nei differenti settori. La nostra ricerca prende ispirazione dal lavoro proposto da Miragliotta et al. (2018), che trova il suo punto di partenza nel parallelismo con la struttura dell'OEE (Overall Equipment Effectiveness). Precisamente l'indice proposto si scompone secondo tre principali dimensioni: la disponibilità del dato, la sua qualità e come è processato dai sistemi di supporto alle decisioni (DSS). Grazie a un'approfondita analisi della letteratura, è stato possibile identificare molteplici sottodimensioni, insieme a un'eventuale metodologia di valutazione. Il modello è stato costruito basandosi sull'approccio Maturity Model, con la finalità di ottenere una sintesi numerica. Quest'ultima è stata progettata in una prospettiva di standardizzazione, per meglio permettere la ricezione del modello nelle realtà aziendali e il successivo riconoscimento a benchmark universale. Mediante l'applicazione del modello nelle diverse industrie, è stato possibile identificare l'indice proposto come un credibile primo punto di riferimento nella valutazione dei processi decisionali basati sull'analisi dei dati.

# Table of content

- 1 Executive Summary ..... 12
- 2 Introduction..... 28
- 3 Literature Review ..... 29
  - 3.1 Introduction ..... 29
  - 3.2 Literature Methodology..... 31
  - 3.3 Literature Analysis..... 33
  - 3.4 Information and Data productivity - Current study ..... 35
    - 3.4.1 Information productivity – Quantitative financial approach..... 35
    - 3.4.2 Data productivity - Qualitative approach ..... 35
    - 3.4.3 Data productivity - Quantitative approach ..... 36
  - 3.5 Background: Industry 4.0..... 39
    - 3.5.1 Big Data ..... 44
    - 3.5.2 Impact on the organisations ..... 45
  - 3.6 Data management and Decision support system ..... 46
    - 3.6.1 Hierarchical view of Data Information Knowledge & Wisdom..... 46
    - 3.6.2 Knowledge management vs Business Intelligence ..... 49
  - 3.7 Data Quality Assessment..... 57
    - 3.7.1 Data quality assessment methodologies ..... 58
    - 3.7.2 Application context and User’s perspective ..... 63
    - 3.7.3 Strategy to enhance data quality ..... 67
  - 3.8 Maturity Model..... 69
    - 3.8.1 Origins and overview of maturity models ..... 70
    - 3.8.2 Maturity model development framework..... 73
  - 3.9 Performance measurement in manufacturing..... 78

3.9.1	Manufacturing Productivity .....	80
3.10	Decision-making process .....	87
3.10.1	Decision-making process within supply chain .....	92
4	Objectives .....	95
4.1	Literature gaps .....	95
4.2	Research Questions.....	98
4.2.1	Academic Objectives .....	98
4.2.2	Practical Objectives .....	99
5	Research Methodology .....	100
6	Model building .....	105
6.1	Model Structure.....	105
6.1.1	User perspective and task dependence .....	108
6.1.2	Data Availability .....	109
6.1.3	Data Quality .....	113
6.1.4	DSS Performance.....	121
6.2	Data Productivity Maturity Model - DPMM.....	127
6.2.1	Research contributions on the topic .....	127
6.2.2	Data Productivity Maturity levels.....	132
6.3	Data Productivity Index 2.0 – DPI 2.0 .....	138
6.3.1	From maturity to index.....	138
6.3.2	Data Relevance.....	140
6.3.3	Data Availability .....	141
6.3.4	Data quality .....	144
6.3.5	DSS Performance.....	148
7	Model Testing.....	152
7.1	Context application .....	153

7.2	Designing Case study .....	155
7.2.1	The Interview.....	156
7.2.2	Training and Preparation .....	161
7.2.3	Case Study Protocol .....	161
7.2.4	Screenings case study nominations.....	163
7.2.5	Pilot Case studies.....	164
7.3	Case study.....	165
7.3.1	CASE A – 1 <sup>st</sup> Pilot Case study .....	165
7.3.2	CASE B – 2nd Pilot Case study .....	166
7.3.3	CASE C .....	174
7.3.4	CASE D.....	177
7.3.5	CASE E.....	179
7.3.6	CASE F.....	181
7.3.7	CASE G.....	183
7.3.8	CASE H.....	186
7.3.9	Overall results .....	188
7.3.10	Judge the quality of the study.....	192
8	Conclusions .....	194
8.1	Findings.....	194
8.2	Improvement actions.....	197
8.2.1	The Data Productivity Improvement matrix.....	199
8.3	Model Limitations & Future Research .....	209
9	Bibliography.....	214
Annex	.....	223

# List of Figure

<i>Figure 1 Thinking behind our research objectives.....</i>	<i>13</i>
<i>Figure 2 Data Productivity factors.....</i>	<i>17</i>
<i>Figure 3 Data Productivity factors and sub-factors.....</i>	<i>18</i>
<i>Figure 4 Data Productivity Maturity Model.....</i>	<i>20</i>
<i>Figure 5 Model Application.....</i>	<i>22</i>
<i>Figure 6 Data Productivity Improvement matrix.....</i>	<i>26</i>
<i>Figure 7 Example of literature selecting and identifying procedures.....</i>	<i>31</i>
<i>Figure 8 Year of Publication distribution.....</i>	<i>33</i>
<i>Figure 9 Main topics of the Literature review.....</i>	<i>34</i>
<i>Figure 10 Model Structure (Cianca 2018).....</i>	<i>36</i>
<i>Figure 11 Data Availability factor (Miragliotta et al 2018).....</i>	<i>37</i>
<i>Figure 12 Data Quality factor (Miragliotta et al. 2018).....</i>	<i>37</i>
<i>Figure 13 DSS Performance (Miragliotta et al. 2018).....</i>	<i>38</i>
<i>Figure 14 The fourth industrial revolution.....</i>	<i>39</i>
<i>Figure 15 Replacement of equipment of the Industry 4.0.....</i>	<i>40</i>
<i>Figure 16 The three main pillars of Industry 4.0.....</i>	<i>40</i>
<i>Figure 17 Smart manufacturing technologies (Politecnico di Milano).....</i>	<i>42</i>
<i>Figure 18. DIKW Hierarchy (Rowley, J., 2007).....</i>	<i>47</i>
<i>Figure 19 From Tacit to Explicit Knowledge (Nemati et al., 2002).....</i>	<i>50</i>
<i>Figure 20 Components of a Business Intelligence systems (C.Vercellis, 2009).....</i>	<i>51</i>
<i>Figure 21 Data taxonomy (Cleven &amp; Wortmann, 2010).....</i>	<i>53</i>
<i>Figure 22 KDD process (Fayyad, U., Piatetsky-Shapiro, G., &amp; Smyth, P., 1996).....</i>	<i>56</i>
<i>Figure 23 Data Quality Hierarchy (Wang et al. 1995).....</i>	<i>58</i>
<i>Figure 24 Fishbone of the Evolutional data Quality (Liu &amp; Chi 2002).....</i>	<i>59</i>
<i>Figure 25 The information quality model.....</i>	<i>60</i>
<i>Figure 26 Data Quality Hierarchy (Bovee et al. 2003).....</i>	<i>61</i>

<i>Figure 27 Data Quality Hierarchy (Cai &amp; Zhu 2015)</i> .....	62
<i>Figure 28 Measuring the perceived value of information (Ahituv, 1989)</i> .....	65
<i>Figure 29 Measuring the realistic value of information (Ahituv, 1989)</i> .....	65
<i>Figure 30 Positioning of maturity models' in-between models and methods (Mettler, 2011)</i> .....	70
<i>Figure 31 Model Development Phases (De Bruin et al. 2005)</i> .....	74
<i>Figure 32 Procedure model for developing maturity models (Becker et al. 2009)</i> .....	77
<i>Figure 33 Performances tree (Sianesi 2014)</i> .....	79
<i>Figure 34 OEE (Nalajima 1988)</i> .....	82
<i>Figure 35 OEE - Availability (OEE Foundation website)</i> .....	84
<i>Figure 36 OEE - Performance rate (OEE Foundation website)</i> .....	85
<i>Figure 37 OEE - Quality (OEE Foundation website)</i> .....	86
<i>Figure 38 Decisions taxonomy</i> .....	87
<i>Figure 39 Organizational roles and decisions (Spina 2012)</i> .....	89
<i>Figure 40 Phases of the decision-making process (Vercellis, 2009)</i> .....	90
<i>Figure 41 SCOR version 4.0 (Supply-Chain Council 2000)</i> .....	92
<i>Figure 42 Supply chain decision categories mapped to the SCOR model ()</i> .....	94
<i>Figure 43 Papers' chart</i> .....	95
<i>Figure 44 Thinking behind our research objectives</i> .....	96
<i>Figure 45 Data Productivity Factors</i> .....	106
<i>Figure 46 Elaboration of model dimensions</i> .....	107
<i>Figure 47 Data productivity factors and sub-factors</i> .....	107
<i>Figure 48 Data Availability</i> .....	109
<i>Figure 49 Model Structure (Cianca 2018)</i> .....	132
<i>Figure 50 The five level of DPMM</i> .....	133
<i>Figure 51 Data Productivity Model</i> .....	134
<i>Figure 52 Model application</i> .....	139
<i>Figure 53 Excel Dashboard Computation</i> .....	160

*Figure 54 CASE A Results - Model Version 2.0..... 171*

*Figure 55 CASE B Results - Model Version 2.0.....173*

*Figure 56 CASE C Results.....175*

*Figure 57 CASE D Results.....178*

*Figure 58 CASE E Results ..... 180*

*Figure 59 CASE F Results..... 183*

*Figure 60 CASE G Results..... 184*

*Figure 61 CASE H Results ..... 186*

*Figure 62 Data Productivity Index 2.0 - Dimensions..... 194*

*Figure 63 Phases for Data Productivity Assessment and Improvement .....197*

*Figure 64 Data Productivity Improvement Matrix.....205*

*Figure 65 Data Productivity Improvement Matrix Application.....208*

# List of Table

<i>Table 1 Sub-factors definitions</i> .....	19
<i>Table 2 Simplified Version of the Model input</i> .....	23
<i>Table 3 Case studies results</i> .....	24
<i>Table 4 Papers selected for each macro arguments</i> .....	32
<i>Table 5 5V's Big data</i> .....	44
<i>Table 6 Data Quality Methodologies</i> .....	63
<i>Table 7 Part of the Quality Management Maturity Grid (Fraser, Moultrie &amp; Gregory, 2002)</i> .....	71
<i>Table 8 The Quality Management Process Maturity Grid (Fraser, Moultrie &amp; Gregory, 2002)</i> .....	71
<i>Table 9 Maturity levels of the software CMM (Fraser, Moultrie &amp; Gregory, 2002)</i> .....	72
<i>Table 10 Methodologies followed</i> .....	100
<i>Table 11 Data Authorization definition</i> .....	110
<i>Table 12 Data Accessibility definitions</i> .....	111
<i>Table 13 Data Timeliness definitions</i> .....	112
<i>Table 14 Data Accuracy definitions</i> .....	114
<i>Table 15 Data Credibility definitions</i> .....	115
<i>Table 16 Data Completeness definitions</i> .....	117
<i>Table 17 Data Consistency definitions</i> .....	118
<i>Table 18 Data Presentation Quality definitions</i> .....	119
<i>Table 19 Data Relevance definitions</i> .....	120
<i>Table 20 Model dimensions definitions</i> .....	126
<i>Table 21 Research Contribution Maturity models - Business Intelligence</i> .....	128
<i>Table 22 Description of the derived MD3M Maturity levels (Spruit et al. 2014)</i> .....	129
<i>Table 23 Key topics and focus area in Master Data Management (Spruit et al. 2014)</i> ...	129
<i>Table 24 Research contribution Maturity model - Big Data</i> .....	130

<i>Table 25 Research contributors for DPMM.....</i>	<i>132</i>
<i>Table 26 Maturity Levels .....</i>	<i>137</i>
<i>Table 27 Relevance Assessment .....</i>	<i>141</i>
<i>Table 28 Existence Assessment.....</i>	<i>142</i>
<i>Table 29 Accessibility &amp; Authorization assessment.....</i>	<i>143</i>
<i>Table 30 Timeliness assessment .....</i>	<i>144</i>
<i>Table 31 Accuracy &amp; Credibility assessment .....</i>	<i>145</i>
<i>Table 32 Completeness assessment.....</i>	<i>146</i>
<i>Table 33 Presentation Quality &amp; Consistency dimensions .....</i>	<i>147</i>
<i>Table 34 Data Usage assessment.....</i>	<i>148</i>
<i>Table 35 Data Integration assessment.....</i>	<i>149</i>
<i>Table 36 Data Processing assessment .....</i>	<i>149</i>
<i>Table 37 Supporting Infrastructure assessment.....</i>	<i>150</i>
<i>Table 38 Model input - Interesting Data for the Distribution Planning process.....</i>	<i>158</i>
<i>Table 39 Company sample table.....</i>	<i>163</i>
<i>Table 40 CASE A Results - Model Version 1.0.....</i>	<i>166</i>
<i>Table 41 CASE B Results - Model Version 1.0.....</i>	<i>167</i>
<i>Table 42 Model version 1.0 – Maturity level scores.....</i>	<i>169</i>
<i>Table 43 Case studies results.....</i>	<i>188</i>
<i>Table 44 Company size classification.....</i>	<i>192</i>

# 1 Executive Summary

This thesis has the objective to define a metrics to quantitatively measure the organisations' data productivity in decision-making processes within the Industry 4.0 paradigm. The research study has been performed following four main phases: the identification of the theoretical background thanks to the (1) Literature review, the model development in the (2) Model building part and the subsequent (3) Model testing. Finally, the model limitations and future research are presented in the (4) Conclusions part.

The **Literature Review** has been conducted initially investigating the current knowledge level of the specific issue through an open research phase to obtain a first overview. Only few preliminary steps have been undertaken addressing closely the theme treated in this dissertation. Strassman (2004) approached the data productivity issue under a financial perspective, providing a quantitative measure to assess the information investment. Two early academic studies have been considered relevant, laying the foundations for the development of a quantitative data productivity metrics. Cianca (2018) proposed a qualitative approach, based on maturity models, to globally evaluate company' datasets in Industry 4.0 context. Miragliotta et al. (2018) introduced a metrics named Data Productivity Index, capable to measure how data support a specific decision-making process. This model was developed starting from the OEE index, breaking down the data assessment into three components: data availability, data quality and decision support system (DSS) performance. Therefore, considering the relevance of the topic both for academia as well as for the industry, it was decided to develop a model taking inspiration from these two studies. Considering their potential, the quantitative and qualitative approaches were merged together in order to obtain the *Data Productivity Index 2.0*.

Particularly, Miragliotta et al. (2018) was the real starting point of our work, taking into consideration the potential limitations identifiable. The previous work is largely marked by great originality at the expense of the business applicability. The application of this

measurement tool in the organisations is characterized by high complexity, requiring significant investments in terms of time and resources.

The proposal of this thesis work is to overcome this issue, trying to figure out a new compromise, based on enhanced applicability without sacrificing the previous work originality. The effort should be made especially on the standardization, with the purpose to make this index comparable between different companies working in different sectors and realities.

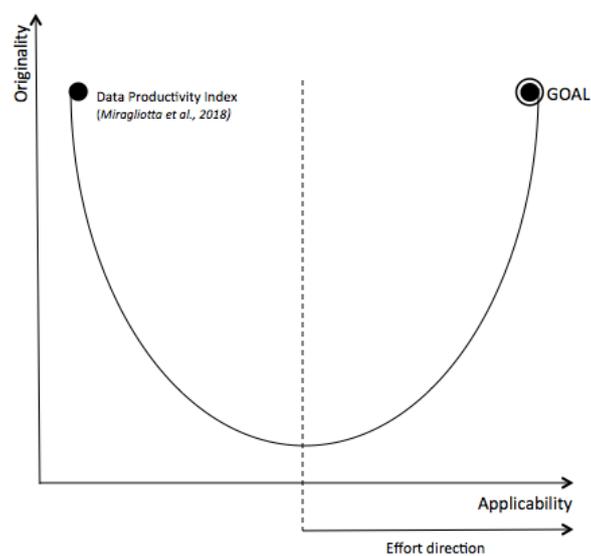


Figure 1 Thinking behind our research objectives

Therefore, the research objectives of this thesis work are of both academic and practical nature. The academic objectives concern, on one side, the identification of the most appropriate features required to assess data productivity and, on the other side, the most suitable approach to evaluate in a structured way the data dimensions identified. Concerning the practical objectives, the main question refers to the possible way to enhance the model applicability in business context.

Whereupon, to better understand the background environment, the current Industry 4.0 scenario has been analysed. It is possible to define this concept as *“a vision of the future of Industry and Manufacturing in which Information Technologies are going to boost competitiveness and efficiency by interconnecting every resource (data, people and*

*machinery) in the Value Chain”* (Politecnico di Milano, 2017). Thus, the main aim of Industry 4.0 is to merge two different worlds, creating a synergy between operational and information technologies, unlocking new value. After having investigated the context of application, it has been considered necessary to focus on the advent of Big Data and on the strategic significance of data exploitation for companies. In order to optimize performances, industrial businesses should adopt data-driven strategies and decisions based on real-time and historical process information, leveraging Big Data as the heart of efficiency and competitiveness. The final scope is to improve the cooperation and interconnection along the value chain, focusing on the end-to-end digitalisation.

Data, as the main subject of analysis, have been investigated thoroughly considering their management and role in decision-making. Firstly, the hierarchical view of data, information, knowledge and wisdom have been discussed under the existent different perspectives. These concepts are generally viewed sequentially, describing data as the raw material of information and information as the raw material of knowledge. Wisdom could be broadly defined as the kind of knowledge that has been processed basing on personal experience. Then the concept of Knowledge Management and Business Intelligence have been discussed in order to understand how data are processed and converted into information and knowledge. From the literature, a slight difference between these two disciplines could be found, noticing that Knowledge Management focuses more on the processing of unstructured and implicit information, while Business Intelligence deals mostly with structured and organized databases. The business intelligence components have been then detailed in order to describe how data are collected, stored, processed and treated in order to become supportive in decision-making processes. Through these studies, the concept of Decision Support Systems (DSS) is emphasised, underlining the importance for companies of having auxiliary tools in supporting decision makers.

After data management overview, a focus on the data assessment topic was required. In literature data assessment studies are centred on the data quality theme, playing a crucial role in business context. Multiple assessment methodologies have been proposed, defining data quality as a multi-dimensional and hierarchical concept, where data quality dimensions

are related to particular data aspects. The different approaches have been analysed in order to detect which are the most relevant dimensions that must be considered in data assessment. Moreover, for our study, it has been crucial to perform a contextualization of data quality in Big Data, detecting which are the new challenges triggered by their advent. The new issue is to extract and collect high quality data from the multiple sources available in this context. Analysing data assessment literature, data quality is not generally considered an objective evaluation just related to its features, but it is strictly connected to two main questions: the application context and the data user. On one side, multiple user could have different quality requirements and expectations, on the other side different circumstances of the data usage could affect data quality. This literature review was required in order to develop a comprehensive theoretical structure of the model, involving a wide range of quality dimensions.

In order to obtain a metrics capable to return to organisations a quantitative measurement of data productivity, it has been necessary to identify an assessment instrument. As a first step, it was taken the decision to resort to a qualitative approach, to be compliant with the subjective nature of the majority of data assessment features. The assessment method selected as reference was the Maturity Model framework, since its growing importance and utilisation, its wide range of application and its generally recognized validity. Maturity model consists of a sequence of maturity levels defined as archetypical states of maturity, characterized by a descriptor and the relative detailed description, with the aim to capture the as-is situation. After a first investigation of maturity models' origins and related development frameworks, the research contributions on close topics were deepen analysed. At first, business intelligence and data warehousing maturity models were examined, considering their significance in supporting the organisational decision-making processes. Then the master data management maturity model was reviewed, considering its precise aim to assess the maturity of organisation's current state of master data management. Overcoming the concept treated by previous maturity models, the Big Data maturity model was studied, given its purpose to evaluate the organisations ability to leverage Big Data and the related value created. In addition, it was deemed necessary to extend our research to a maturity model focusing on the assessment of the state-of-development of manufacturing

company in Industry 4.0, considered the background. Finally, a further research contribution related to process-oriented, knowledge-based decision-making maturity model was considered.

Since introducing a new metrics in the company performance measurements system is the main objective of this work, a first study on the performance measurement was done. A proper set of indicators is required to support managers' activities, guiding them in future decisions, learning from company history performance measurements. Indicators are capable to trigger the motivation of company's resources, being the foundations of improvement process. Companies are getting used to measure productivity of materials, machines, worker and capital to control and improve their own businesses. Therefore, a broad classification of the existing metrics in manufacturing context is performed, focusing particularly on the productivity ones. Moreover, an overview of the OEE metrics is given since, inspired by the work of Miragliotta et al (2018), the starting point of the model is to develop the data productivity index through a parallelism with the OEE.

Data productivity must be assessed in relation with a certain decision-making process, detecting if data are actually supportive. Therefore, it was decided to make a study on the business processes to have an overview on the type of decisions that a company could face. In this part the important concept of limited rationality is introduced, stressing that decisions are not usually undertaken under optimal circumstances. Indeed, decision makers do not always have a clear understanding of the objectives to be attained, information available or potentially required. Moreover, the companies decision-making processes were investigated in order to detect the processes for which data, playing a crucial role, could be fully exploited.

The **Model Building** and design was done in accordance with guidelines and frameworks from academia (Hevner et al. 2004; Becker et al. 2009; Garcia et al. 2012) in order to maintain the work consistent with the academic requirements. Thus, an iterative procedure and multi-methodological approach were followed, maintaining rigor in the research. After having carry out a comprehensive literature study, the following step was the definition of

the model architecture and relative parameters. Taking as reference the OEE factors, three main dimensions of assessment were identified:

*Data Availability* aims to detect if data are accessible and available to the user when required, supporting the decision maker.

*Data Quality* aims to evaluate if data available to the user are of proper quality to support the specific decision.

*DSS Performance* aims to evaluate the decision support system, assessing the information system capability to integrate and process data, in order to properly support the user within a decision-making process.



Figure 2 Data Productivity factors

Each dimension is described by a set of subfactors, resulting from the previous literature studies, with the aim to develop the theoretical model. The final model structure is showed in Figure 3.

The subfactors selection for the Data Availability & Data Quality components was mainly driven by the data quality assessment methodologies proposed in literature (Wang et al. 1995; Wand et al. 1996; Lee et al. 2002; Liu et al. 2002; Pipino et al. 2002; Bovee et al. 2003; Cai et al. 2015). The model structure concerning these first two dimensions is mainly inspired by the one introduced by Cai et al. (2015), since it discusses the data quality assessment under the Big Data advent, taking into consideration the new challenges triggered.

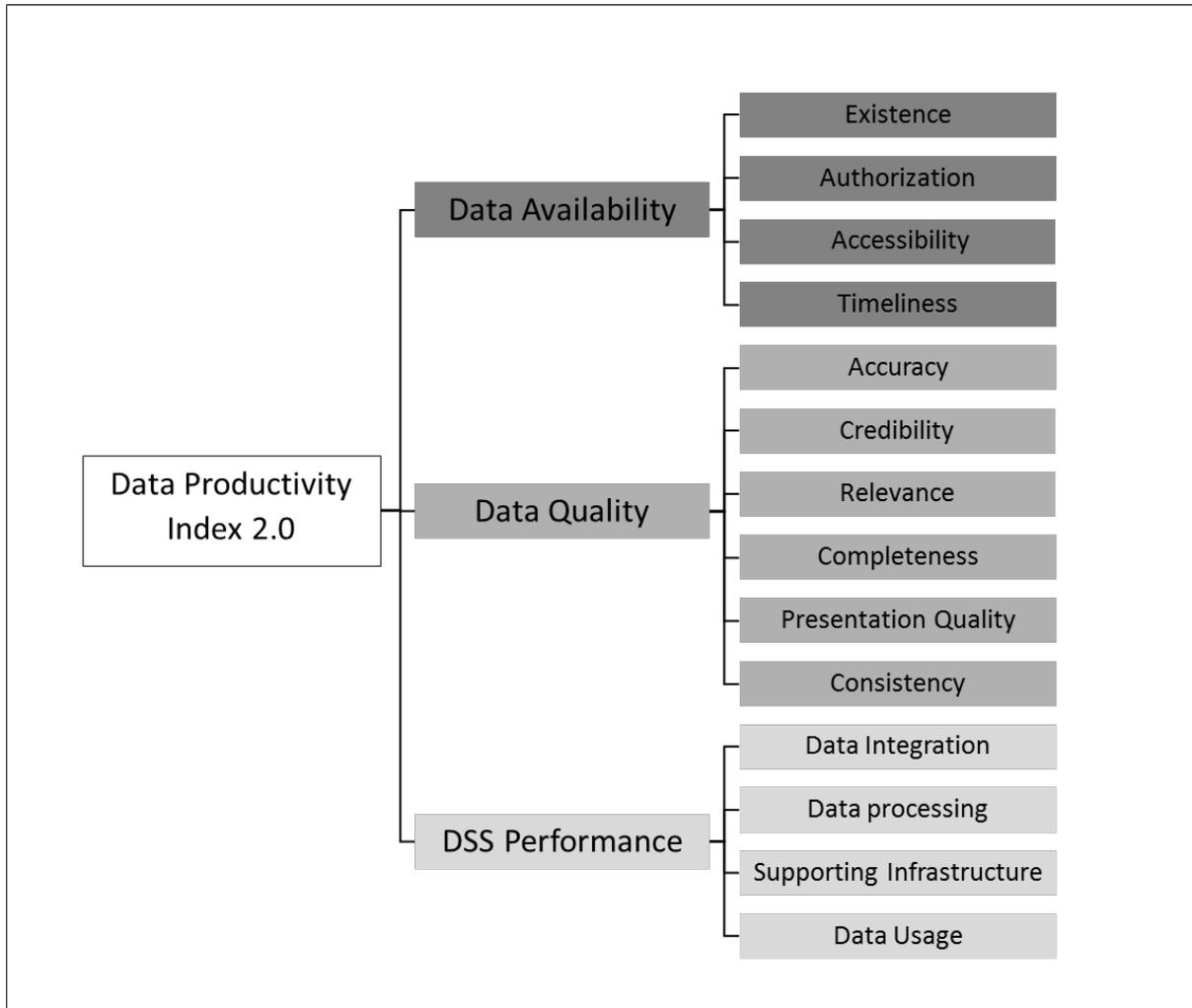


Figure 3 Data Productivity factors and sub-factors

Concerning the last component, DSS performance, the literature does not provide a consolidated subset of factors to assess the Decision Support System of an enterprise. Therefore, stemming from the literature on the data management, knowledge management, business intelligence and DSS, four dimensions characterising the DSS Performance were defined: the data integration across the organisation, the capability of data processing and analysis, the support given by the underlying infrastructure and the data effectively used in decision-making. *Table 1* summarizes the definition given to each model feature.

Dimension	Definition
Data Existence	Existence refers to the extent to which data are effectively generated and collected in a company.
Accessibility	Accessibility refers to the extent to which is difficult for users to obtain data: if they are easily and quickly retrievable and if a data access interface is provided.
Authorization	Authorization refers to the extent to which an individual or organization has the right to use the data.
Timeliness	Timeliness refers to the extent to which data are regularly updated for the task at hand and whether the data arrive on time.
Accuracy	Accuracy refers to the extent to which data are correct and free-of-error.
Credibility	Credibility refers to the extent to which data sources are checked, certified and trustworthy. But also, to the extent to which data is unbiased, unprejudiced and impartial.
Completeness	Completeness refers to the extent to which data is not missing and of sufficient breadth, depth, scope for the task at hand.
Presentation Quality	Presentation quality refers to the extent to which data are clear and understandable (understandability) and compactly presented (concise representation).
Consistency	Consistency refers to the extent to which data are: (1) presented in the same format, being the same or closely similar across time or space, (2) after processing their concepts, value domains and formats still match as before.
Relevance	Relevance refers to the extent to which data match the theme and are helpful for the task at hand.
Data Integration	Data Integration refers to (1) the interoperability, defined as the capability to cooperate and exchange information, and to (2) the level of integration among data coming from multiple sources (e.g. different databases, sensors).
Data Processing	Data Processing refers to (1) the data treatment by state-of-the-art data processes and algorithms with the aim to exploit their potential, and to (2) the ability to analyse data on various dimensions, aggregating, disaggregating and presenting data in form of indicators to better support the decision makers.
Supporting Infrastructure	Supporting Infrastructure refers to (1) the level of adequacy of the underlying hardware and software infrastructure to run data analysis techniques and to (2) the capability to process data on time to properly support the decision maker.
Data Usage	Data Usage refers to extent to which data are actually used in a specific decision-making process by the decision maker.

*Table 1 Sub-factors definitions*

After the sub-dimensions' definition, the next phase consists in the maturity assessment; for each sub-dimension, five maturity levels were identified. The maturity levels are non-existent, initial, defined, managed & measured, and strategically optimized. Considered the

novelty of this topic for organisations, it was considered suitable to start the maturity assessment from level 0 (non-existent=maturity level 0).

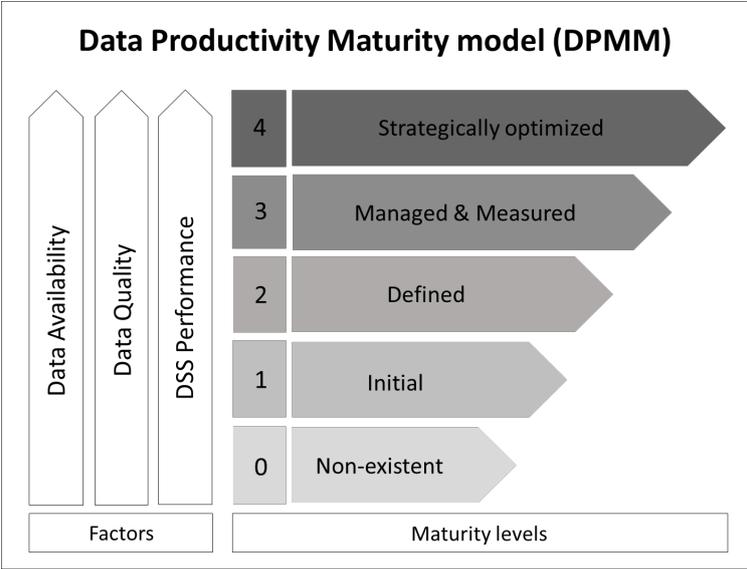


Figure 4 Data Productivity Maturity Model

**Non-Existent** - Level 0 is the lowest one, characterized by ill-defined individualistic decision-making and no real sense of awareness on the topic. Companies have not yet realized the great potentiality represented by data in supporting decision-making. Indeed, decisions are based upon judgement, following the individual knowledge resulted from unrecorded and tacit personal experience

**Initial** – A first general awareness of the topic is present in organisations. Decision-making processes are based on file drawer-type system of personal protocols. A primal management of the process is present, and the data management is in an irregular and undefined form.

**Defined** – Developed awareness of the importance of structured data management in the decision-making process, thanks to the implementation of primary structures in the organisation, not yet institutionalised. The process could be described as formal and disciplined and data are managed through basic and standardized procedures.

**Managed & Measured** – Best practices are in places for managing data in decision-making, using developed, standardized and well-defined procedures, measured and controlled across the organisation. Data management becomes a company core activity.

*Strategically Optimized* – Optimized handling and exploitation of data in supporting decision-making processes. Data-driven decision-making process is well structured, controlled and collaborative. The process is perfectly aligned with the company strategic business goals and the system is able to improve the learning culture.

**From maturity to index:** after having generally defined the maturity levels, the transition from a qualitative approach - Maturity Models - to a quantitative one is made, to obtain a comprehensive numeric index assessing the data productivity. The proposed index is named *Data Productivity Index 2.0 (DPI 2.0)*. For every model's feature, the transition between the two approaches is made by assigning a numeric value ranging from 0 to 1 to each maturity level. In this way the interviewee - the data user – proceeds in the selection of the sentence better describing the current state of data according to its perception; and beyond, each sentence is represented by a numeric value. These values led firstly to the achievement of each single feature evaluation, then to the aggregation into the three macro dimensions – Data Availability, Data Quality, DSS Performance – and eventually to the computation of the comprehensive index, DPI 2.0.

More in detail, the starting point of the model application is represented by a list of data deemed interesting for the specific decision-making process. This is provided as input with the aim to submit a comprehensive and supportive list of data for the decision maker. Firstly, the interviewee must evaluate the relevance of each data of the list, indicating how much data are significant and helpful, matching the task at hand. The relevance dimension has a dominant role in the DPI 2.0 computation, being deployed as weight for each data of the list provided as input. This evaluation is given regardless of what are the data collected by the company. The relevance weight is applied to assess the dimensions related to the Data Availability and Data Quality.

Following the macro dimensions are evaluated. The assessment of each feature identified relative to the Data Availability dimension is done taking as object of analysis each single dimension specified in the list of *Interesting Data*. Data Existence aims to identify which is the subset of Interesting data actually collected and generated by the company. After that, all the other dimensions are evaluated starting from the existent data. Particularly, the

Authorization and Accessibility features were merged together for the maturity assessment. This has been possible after a first application of the model in the business environment resorting to case studies.

For what concerns the Data Quality dimension, the assessment of each feature is done taking as object of analysis each single item specified in the list of *Interesting Data*. Even in this case, the features Accuracy & Credibility and Presentation Quality and Consistency were merged together. The Data Quality features are evaluated only for the Existent Data. As regards the DSS Performance, the features assessment is done taking as object of analysis the whole dataset available to the decision-maker to take the decision. Except for the feature Data Usage, for which the evaluation is done taking as reference each single data existent.

0	<b>Non-existent</b>	No real sense of awareness on the topic	Relevance	Data Availability			Data Quality			DSS Performance			
1	<b>Initial</b>	A first awareness for issues regarding the use of data in the decision making process is achieved		Existence	Authorization & Accessibility	Timeliness	Accuracy & Credibility	Completeness	Presentation Quality & Consistency	Data Architecture & Integration	Data Processing	Supporting Infrastructure	Data Usage
2	<b>Defined</b>	Developed awareness of the necessary importance of a structured data management in the decision making process											
3	<b>Managed and Measured</b>	Best practices are in place for managing data in the decision making process											
4	<b>Strategically Optimized</b>	Optimized handling of data in taking decisions. The organization's efficiency in using data in decision-making process has been improved											

Figure 5 Model Application

Properly aggregating the sub-dimensions, a numeric value for each macro factor – Data Availability, Data Quality and DSS performance - is obtained. At the end, taking inspiration by the OEE metrics, the DPI 2.0 is computed multiplying these three values.

$$Data\ Productivity\ Index\ 2.0 = Data\ Availability(\%) * Data\ Quality(\%) * DSS\ Performance(\%)$$

The **Model Testing**, and the related evaluation and consolidation, was done using an observational methodology, resorting to the case study approach (Yin, 2009). The field of

application considered is represented by the manufacturing and distribution companies. With the purpose to standardize and enhance the model applicability, it was necessary to restrict the field of application to one specific decision-making process. Particularly the model is applied in the context of Distribution planning process, since its highly data-intensive, deterministic and repetitive nature. Furthermore, the respondents were selected carefully, addressing managers familiar with both the decision process under analysis and the relative required data.

In the case study design phase, after the definition and development of the theoretical model to validate, the detection of the suitable companies according to precise criteria was performed. The resulting sample of this selection process encompasses 8 companies. Two pilot cases were conducted before proceeding further with other case studies, in order to revise and retest the initial theoretical assumptions according to the possible discoveries during data collection. This loop procedure and the redesign were considered an essential step for the development of the effective practical model. To apply the model and compute DPI 2.0, a spreadsheet was developed.

Descriptive data	Product descriptive characteristics and Product physical characteristics
	Product characteristics affecting lead time
	Distribution network characteristics and typology
	Transportation times per segment and possible frequency
	Total lead time door-to-door, with relative variability
	Transportation cost per segment
Inventory and flow data	Warehouse inventory level
	Pipeline (transit) inventory level
	Incoming orders, backlog or outstanding
Market Data	Customer demand forecast
	SLA
	Working calendar
	Calendar promotional events
	Possible substitute products

*Table 2 Simplified Version of the Model input*

Considering the distribution planning focus adopted by the model, it has been necessary to specifically develop a baseline list of relevant data, relying on expert opinions in the field. The most suitable level of data aggregation was chosen, in order to ensure the model

applicability: the list items were grouped according to the possible area of belonging to make the evaluation more intuitive.

The pilot cases have been performed to refine and revise the model. These enabled to realize the critical issues of the first version of the model leading to the construction of the final one. Firstly, the pilot cases have confirmed that some dimensions of the theoretical model are strictly correlated and consequently assessable with common maturity constructs. Therefore, the concept of Authorization & Accessibility, Accuracy & Credibility and lastly Presentation Quality & Consistency were merged together. These cases lead also to the maturity models reformulation, since in most of the cases, they were found to be too academic and strict, being not meaningful to represent the current companies' state. Moreover, the numeric scale of each maturity model for each dimension was revised and calibrated, leveraging on interviewed managers' feedback. The model input - Interesting data - was considered by both companies understandable and comprehensive, encompassing all the collectable data that may support the distribution planning process. Finally, thanks to the suggestions received by the interviewees it was possible to make the spreadsheet more manageable and intuitive to be filled.

Company	Sector	Availability	Quality	DSS Performance	DPI 2.0
A	Home care retailers	25,5%	47,1%	40,8%	4,9%
B	Pharmaceutical	66,2%	75,6%	80,4%	40,3%
C	Home and Personal care	76,6%	71,8%	71,9%	39,6%
D	Mass Distribution	78,0%	88,9%	83,0%	57,5%
E	Mass Distribution	64,8%	80,3%	77,9%	40,5%
F	Freight and logistics	67,7%	95,9%	77,5%	50,3%
G	Freight and logistics	51,5%	76,7%	71,7%	28,3%
H	Fashion	72,4%	88,2%	70,6%	45,1%

*Table 3 Case studies results*

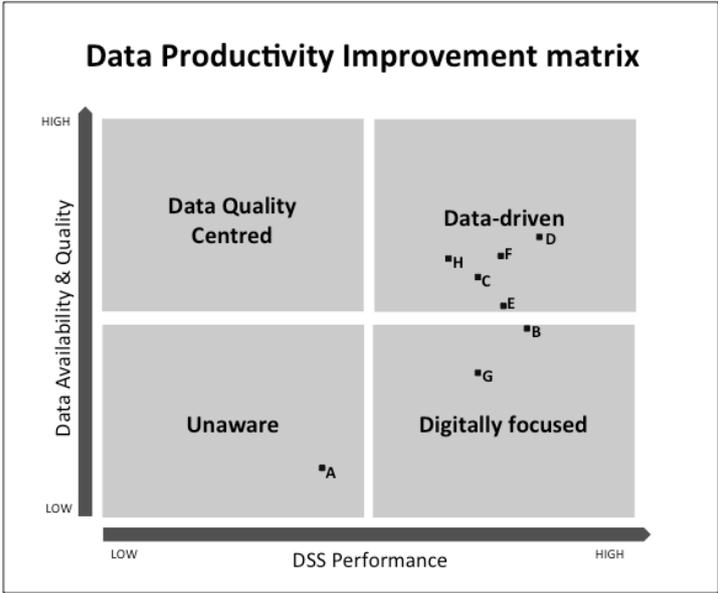
Globally, the model was validated through 8 different case studies. *Table 3* summarizes the Data Productivity Index 2.0 reached by each company interviewed and its related macro factors. From the overall results obtained, it is evident that Data Availability is the dimension

affecting the most the company results. According to the KPI design, existent data has predominant role, therefore leading to a bad score if a company does not collect data potentially useful in supporting a distribution planning process.

The interviews were performed according to a predefined protocol. Through the analysis of the overall results, it has been possible to highlight some important points. The relevance of the theme was confirmed by experts interviewed, since the company necessity to effectively measure data productivity in decision-making and the current lack of a suitable measurement tool. Indeed, companies realized the potential of implementing Industry 4.0 practices, and therefore the importance of collecting data from multiple sources to support decision making. The organisations showed a strong interest on the subject presented, in particular on the final result achieved and on the possible improvement actions. This is proven by the quite satisfying response rate gained and by their availability in providing insightful suggestions, comments and feedbacks. The model application confirmed the absence of difficulty encountered by the interviewee when filling the model. Case study reaffirmed the improved model applicability in the business context, adapting the model.

In the **Conclusion** section, several remarks are provided on the DPI 2.0 utilization and managerial impact. The Data Productivity model, in fact, should be considered as integral part of an improvement process and not only as absolute measure of performance. Through an evolutionary improvement path, our model suggests staged improvement and it could be considered a guide toward the achievement of a desired level of maturity. This facilitates the determination of key elements of current and potential capabilities, identifying the most critical issue. Two general priority strategies were identified with short-term and long-term perspective respectively. The Data-driven strategies aim at improving the data status, directly modifying their value, while the process-driven strategies aim at improving the data status identifying and deleting the root causes of errors and redesigning processes. A classification of the organisations' data exploitation in decision-making is proposed identifying two main dimensions - Data Availability & Quality and DSS Performance – according to the literature studies. From the intersection of the two dimensions, a matrix made up of quadrants is created: Data-Driven, Data Quality Centred, Digitally Focused and

Unaware. Each quadrant represents a class of companies requiring a specific improvement strategy. In *Figure 6* the resulting DPI 2.0 obtained by the companies interviewed were reported in the Data Productivity Improvement matrix.



*Figure 6 Data Productivity Improvement matrix*

Considered the novelty of the topic, the model presents limitations and possible directions for future research. As in the case of the OEE, the reference standard set for the DPI 2.0 is a value as close as possible to 100%. At the moment it is impossible to determine a rationale indicating a best-in-class reference value for this metrics, since the narrow sample on which the model was tested. It is necessary to keep on moving towards a standardisation direction, working with standardisation bodies to ensure the model transposition in business context, with the purpose to achieve a universal benchmark. Future research could be aimed at the identification of a wider set of dimensions for the model structure. Moreover, testing the model on a bigger sample through a survey, possible dependency among the different features could be detected through statistical analysis. Nevertheless, mitigation strategies have been deployed to mitigate the data assessment subjectivity, this is still an open question that could be better improved and investigated. In the transition from the qualitative to the quantitative approach, future research must strive for the maturity model calibration, understating if the value adopted by each maturity level properly represent the related sentences. Concerning the application, the model requires continuous maintenance,

so as to properly represent the current state of the business processes and to regularly update the input of the model. Eventually, this thesis work tested the model just in a specific decision-making process (distribution planning), and therefore a future direction is to enlarge the field of application, implementing the model in different processes.

## 2 Introduction

Over the years, knowledge and experience have become ever more important assets for manufacturing companies. Today, to survive in such a dynamic environment it is essential for organizations to react and adapt rapidly to changes, as well to recognise the role of knowledge in improving decision-making and performances (Shafiq, 2015).

For companies operating in turbulent environment, the quality of decision-making process is a fundamental key of management. Data-driven decision-making is based on evidence rather than intuition, by allowing the fulfilment of better results both in operational and financial terms. The application of Big Data enables enterprises to reach competitive advantage in the market. Data-driven decision is not always a simple proposition to realise for people and organisations. Inappropriate knowledge management could lead to high response times and lack of flexibility to adapt in dynamic environments.

Since the relevance of data for organizations, it becomes essential to find a way to measure the relative productivity – the most deployed efficiency measure in logistic-productive systems - in order to understand the current level of digitalization and digital information exploitation. There is a clear gap of knowledge highlighted in literature about this topic: the lack of a measurement tool to understand the effective support of data collected by organizations in the decision-making process.

In this last year, initial steps have developed addressing the topic. The first author approaching closely the theme is Strassman (2004), providing a quantitative measure to assess Information Productivity. Its main drawback is that it is built on aggregated financial indicators and accounting voices, limiting the productivity measure of each element. Two early academic studies appear to be relevant for this new measuring challenge. On one side, Cianca (2018) introduced a qualitative approach, based on maturity models, to measure data productivity in Industry 4.0. On the other side, Miragliotta et al. (2018) presented a metrics to measure how data support a specific decision-making process. The metrics was developed taking inspiration from the OEE index, breaking down the data assessment into three components: data availability, data quality and decision support system (DSS) performance.

# 3 Literature Review

## 3.1 Introduction

The starting point of this thesis work was the verification of the knowledge gap in literature, to justify the intention of investigating this topic. Indeed, the literature review was at first conducted with the main aim of exploring the Data Productivity theme. At the moment of research, it was identified an almost complete absence of papers concerning the measurement of the productivity related to the use of data collected by the organisations in decision making processes. The absence of papers focused on this specific issue established the novelty of the topic and conducted the research towards closer themes, such as the application of data productivity metrics in every possible area. This led to the confirmation of the need for further investigations.

Miragliotta et al. (2018) was considered the reference point and basis for the work. This enabled the clear identification of the main topics to examine thoroughly and the boundaries of the research to carry out the study.

Therefore, once conducted the first open research phase, the context of application was analysed in detail, in order to better understand the background environment. This led us to give an overview of the current Industry 4.0 scenario. Since traditional manufacturing is updating to the adoption of Industry 4.0 it was also considered necessary a focus on the advent of Big Data and on the strategic significance of data exploitation for organizations.

After having explained the reference framework, several macro arguments to analyse were identified:

- Data management and Decision Support System (DSS)
- Data Quality Assessment
- Maturity models
- Performance Measurement Systems and Manufacturing productivity
- Decision-making process

At first, the research focused the attention to the principal subject to study: the data. Deepening and widening the topic, this preliminary study moves from the concept of data to wisdom, describing the data-information-knowledge hierarchy. This point was then evolved through the concepts of knowledge management and business intelligence, in order to better explain the DSS, thus how data can be used to support companies' decisions.

Then the research headed to the exploration of the methods of data quality assessment, taking into account the rich body of literature existent about this theme. In particular the theme was also declined according to two key points for us: Big Data and the user's perspective, investigating also the possible strategies to adopt to enhance the data quality.

Furthermore, it was decided to investigate the theme of the development of an assessment instrument. For this case, the assessment method selected as reference was the maturity model framework. This kind of assessment instrument was chosen because of its growing importance and utilization, its wide range of application and its generally recognized validity. That's why a literature review about this topic was set up. Initially origins and overview of maturity models were presented and then the frameworks to develop maturity models were investigated.

Going back to the subject of study, the other fundamental topic to deepen is the measurement of productivity referred to the manufacturing context. Hence, a preliminary research on performance measurement systems in manufacturing was conducted, focusing then on the measurement methodologies of manufacturing productivity, such as the OEE (Overall Equipment Effectiveness).

The last point treated is the decision-making process. After having presented several classifications of decisions, the analysis submitted the structure of the decision-making process and the different kinds of decision within a supply chain.

## 3.2 Literature Methodology

In the previous part - Introduction to literature review - the core topics treated in order to sustain the successive empirical research were presented. Indeed, each topic identified has been inspected in depth with the purpose to consider all the elements necessary to fully describe it, even using the possible relationship and connections identifiable.

This study adopts a holistic approach by including all possible terms and similarly different keywords related to each macro argument. The set of these terms is used as keywords to search literature.

It is important to clarify that only journal papers published in English language are included. Pursuing Fahimnia et al. (2015) line of thinking, conference papers, papers in commercial magazines, and book chapters were almost completely excluded from the research to ensure quality, and essentially only journal papers, reviews, and papers in the press were considered.

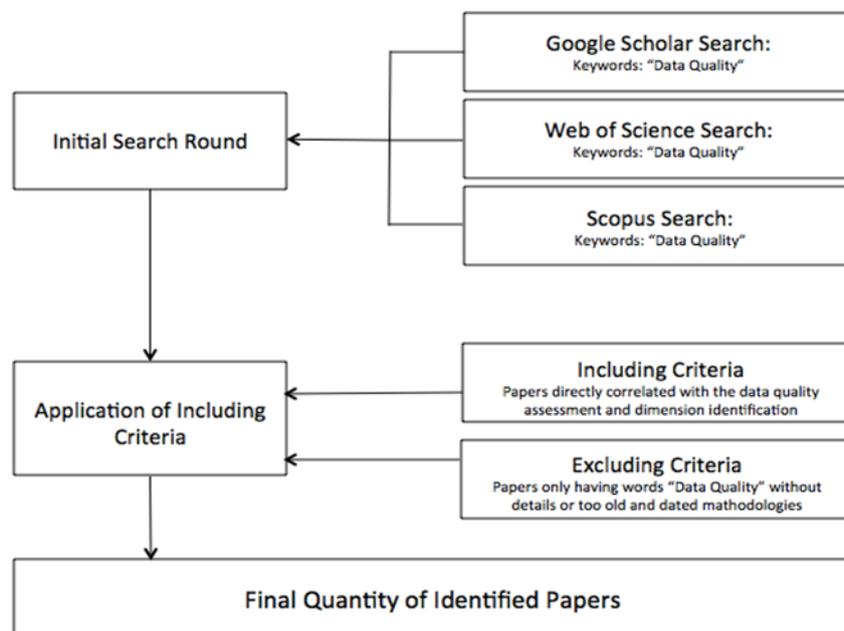


Figure 7 Example of literature selecting and identifying procedures

The approach used in order to select the papers and conduct the literature review, follows the two-state one initiated by Webster and Watson (2002).

For each keyword chosen, the first step was looking for published papers. The collection of material has been done using mainly three databases: Google Scholar, Scopus and Web of Science (WoS) databases. As second step, these papers were reviewed, and unrelated papers were carefully ruled out. Although the initial search was not restricted with a time limit, at the end the most recent papers published were privileged.

To recap, the papers selected according to macro arguments identified can be summarised and grouped into the following research categories along with the number of papers selected and analysed (*Table 4*).

Search themes	Papers selected
Industry 4.0 and Big Data	16
Data, Information & Knowledge	22
Data Quality	15
Decision Support Systems (DSS)	5
Decision-making process	16
Maturity Model	33
Productivity	11
<b><i>Total</i></b>	<b><i>118</i></b>

*Table 4 Papers selected for each macro arguments*

## 3.3 Literature Analysis

### Year of publication

The majority of the papers selected have been published in the last 10 years (50% on the total), in order to be updated on the topic analyzed. It was required to select also some old articles (from 1960 to 2000) since they provide the basic concepts, approaches and methodologies of some specific topics. For example, Simon (1960) was the first author introducing the classification between programmed and non-programmed decision; and then Gorry et al. (1971) distinguished among structured, semi structured and unstructured decisions. Another example is the work proposed by Paulk et al. (1993) – Capability maturity model version 1.1 - which is one of the first works introducing the maturity approach identifying the conventional five different levels of maturity: initial, repeatable, defined, managed and optimized.

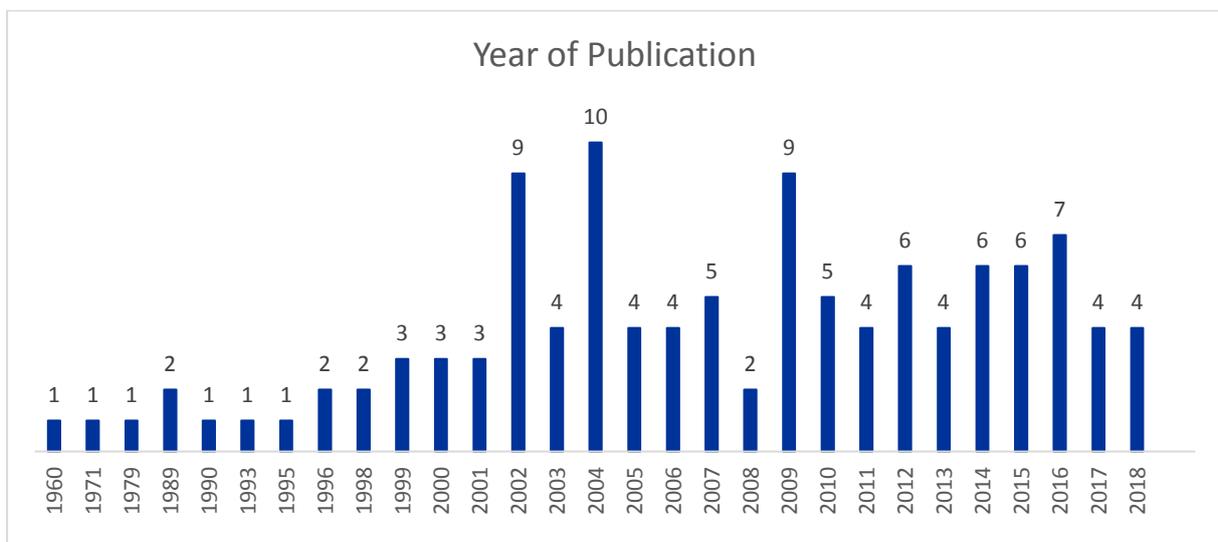


Figure 8 Year of Publication distribution

### Topic

As mentioned, the literature review was performed investigating mainly on six macro areas: data management and decision support systems, data quality assessment, decision making process, industry 4.0 and Big data, maturity models and performance measurement systems (Figure 9). It could be not found an honest-to-goodness balance among the different macro

areas. Indeed, a detailed and deep study was particularly required for the maturity model (28% on the total) and for the Data management (23%) fields since these the focal points of this study.

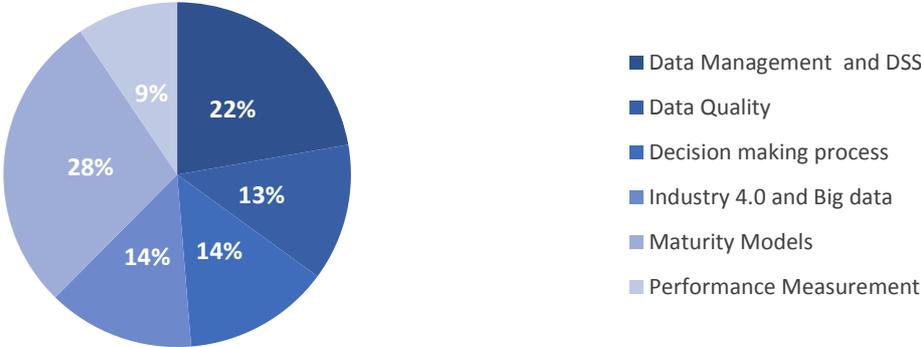


Figure 9 Main topics of the Literature review

In addition to the articles selected for these macro areas, it was made a literature review on the methodologies field. In particular, it was made to understand the approaches required to:

- how to execute a literature review
- how to implement case study methodology
- how to develop maturity models

## 3.4 Information and Data productivity - Current study

### 3.4.1 Information productivity – Quantitative financial approach

Strassmann (2004) introduces the Information Productivity Index as a metrics to rank enterprises, basing on the measurement of the “valued added to a corporation’s profitability by information centered task”. Therefore, the information productivity index objective is to evaluate the information investment, verifying how effectively companies manage their information.

The index is formally presented as follow:

$$IPI = \frac{Output}{Input} = \frac{Information\ Value - Added}{Transaction\ Costs}$$

The output represents the economic value of information resources defined as the profit after taxes minus all the economic costs; while inputs constitutes the economic cost of information resources including managing, accounting, marketing, planning, coordinating, training and communicating costs. This metrics focuses its attention on the effective use of the information technology rather than analyzing the information technology itself.

### 3.4.2 Data productivity - Qualitative approach

Cianca (2018) in his work proposed a qualitative approach to measure data productivity in Industry 4.0, implanting maturity models. The model is composed by the following main aspects: the data itself and the means and tools applied to exploit the data potential. Each aspect is then described by four independent dimensions, addressing the critical topics of each area.

Data		Models and Tools	
Completeness	Communication	Control	Elaboration
Access	Frequency	Knowledge Exploitation	Integration

Figure 10 Model Structure (Cianca 2018)

The parameters defined for the *Data* area are:

- Completeness: evaluation of the company digitalization
- Communication: concerning the data exchange among the different entities, departments and systems
- Access: assessment of the data accessibility level of a company
- Frequency: measurement of the level of data updated and acquisition

The parameter identified for the *Models and Tools* area are:

- Control: aiming at identifying the methodology applied to monitor the process output
- Elaboration: measurement of the company ability to extract knowledge through data analysis and processing
- Knowledge Exploitation: aiming at delineating how much operators relied on the extracted knowledge
- Integration: related to the degree of integration between data-coming from the different company departments.

To evaluate each dimension of the model the maturity approach was deployed, identifying five different level of maturity for each dimension.

### 3.4.3 Data productivity - Quantitative approach

Miragliotta et al. (2018) presented a method to measure the data productivity within manufacturing context, studying how data support a specific decision-making process. The

model was developed taking inspiration from the OEE index, breaking down the data productivity model into three components: data availability, data quality and decision support system (DSS) performance. As happen through the OEE application, each factor of the developed metrics highlights the process loses.

$$\text{Data Productivity Index} = \text{Data Availability} * \text{Data Quality} * \text{DSS performance}$$

**Data Availability:** the first factor is measured starting from a list of data identified as *Interesting data* for the specific decision-making process. Due to multiple losses such as lack of an appropriate data acquisition technology, only a part of the data included into list is actually generated and therefore available to the Decision maker. From the generated data, only that part of data, frequently updated are considered available and suitable to support the specific decision-making process.

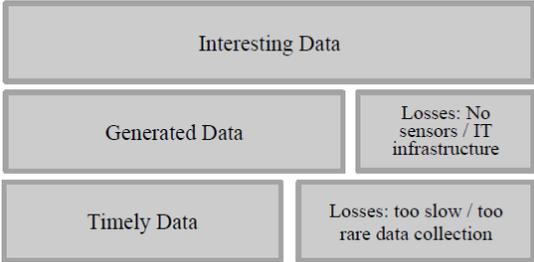


Figure 11 Data Availability factor (Miragliotta et al 2018)

**Data Quality:** the second factor aims to analyse if the data features – *completeness* and *correctness* - are suitable to back up the specific decision-making process. Within this factor the losses recognised are missing and wrong data.

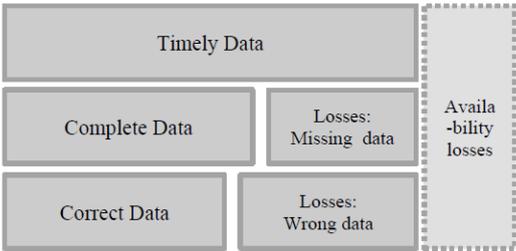


Figure 12 Data Quality factor (Miragliotta et al. 2018)

*DSS Performance*: the third factor evaluates, through four data features, how data are elaborated thanks to the technological layer. Firstly, it measures if data are *Valorised* by algorithms' implementation. The second feature considered is *On-time data* which aim is to analyse if data are processed in due time for the specific decision-making process. The third feature, *Aggregated data*, analyse if data are integrated and presented to the decision maker in a more structured and significant way, for example through the KPI deployment. Finally, the last factor asks to the decision maker which is the data's subset that is effectively *Used* to undertake the determinate decision.

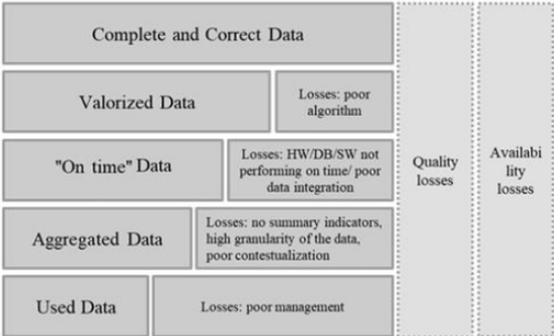


Figure 13 DSS Performance (Miragliotta et al. 2018)

In this study the Data productivity Index was tested on 3 company by means of an *Excel dashboard* to drive the features evaluation. To implement the model for each company, the interviewer must arbitrarily select a specific decision-making process and then evaluates each data of the provided detailed list. In case this detailed list of data is not considered comprehensive, it could be adapted by the interviewee, eliminating or adding new data.

## 3.5 Background: Industry 4.0

The concept of industrial revolution is a key point to effectively define the notion of Industry 4.0. The industrial revolution could be defined as the process of radical replacement of economic system main pillars, with a consequent shift in all the society principal aspects. Indeed, industrial revolutions characterized industrialization thanks to technological, organizational and conceptual developments.

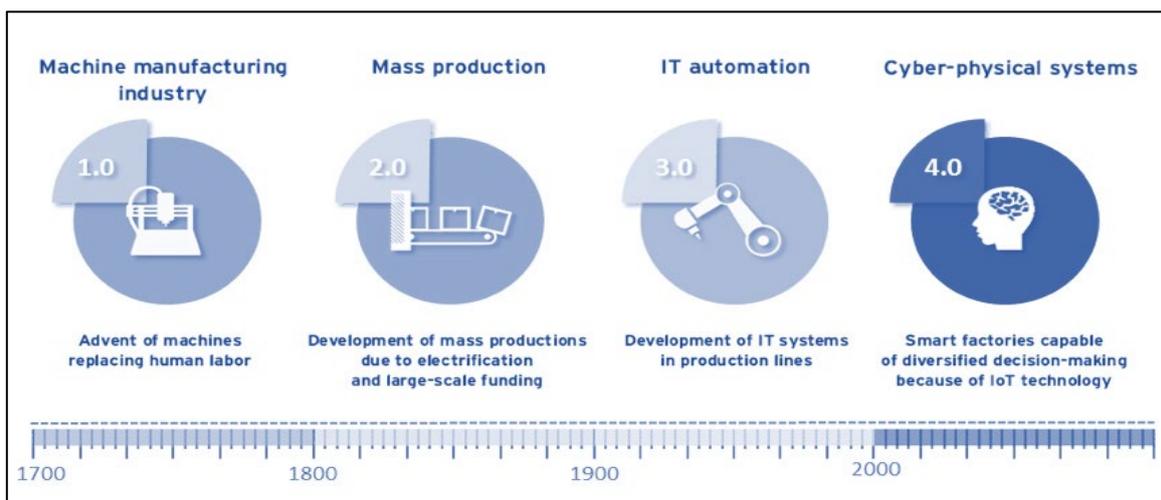


Figure 14 The fourth industrial revolution

Going through the history, it is possible to identify four industrial revolutions.

- First Industrial Revolution generally dated at the end of 18<sup>th</sup> century. It is characterized by the mechanization of the production and transportation sector, mainly thanks to the use of steam and waterpower.
- Second Industrial Revolution found its beginning in 1870, mainly driven by the exploitation of electricity as a source of power. This is the period of the mass production.
- Third Industrial Revolution starts conventionally in 1969, characterized by the spread of the digitalization. In fact, it set its foundations in the introduction of computers and automation in manufacturing processes.
- Fourth Industrial Revolution is a new paradigm characterized by a big novelty, where data are integrated into the equation of managing a manufacturing company.

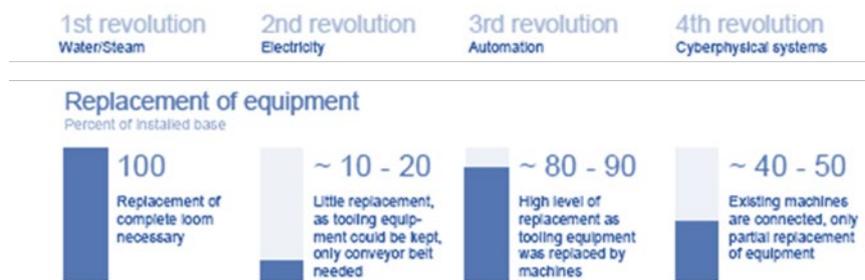


Figure 15 Replacement of equipment of the Industry 4.0

The concept *Industry 4.0* was introduced for the first time in 2011, at the Hannover fair in Germany. Industry 4.0 could be defined “a vision of the future of Industry and Manufacturing in which Information Technologies are going to boost competitiveness and efficiency by interconnecting every resource (data, people and machinery) in the Value Chain” (Politecnico di Milano, 2017). Indeed, the unprecedented advancement in the field of information and communication technologies (ICT) pushed organizations to utilize its features, thus the necessity of defining the new concept of Industry 4.0. According to McKinsey (2015), compared to the previous Industrial revolutions, the fourth one is not about replacing the existing assets (Figure 15).

The three *Main Pillars* of I4.0 can be identified: connection, digitalization and intelligence.

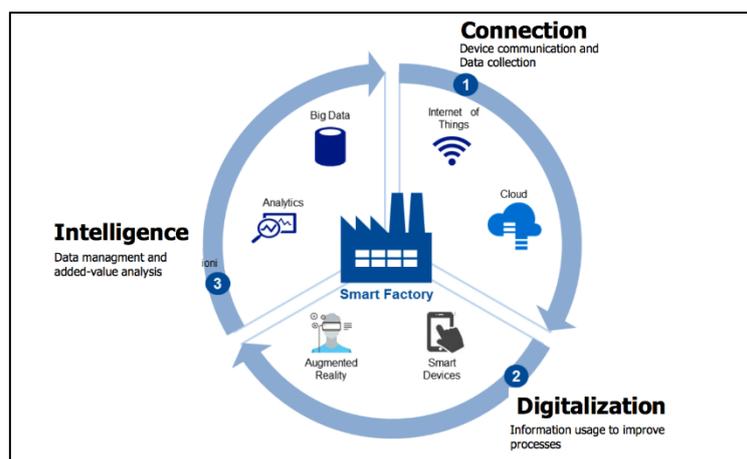


Figure 16 The three main pillars of Industry 4.0

The connection is enabled by the use of Internet of Things (IoT) and Cloud computing, making accessible shared remote computing facilities, such as cloud platform, web applications and online storage services. The second pillar is digitalization, which is strictly

connected to the concepts of smart devices and augmented reality. Intelligence instead is referring to data analytics and Big Data.

Focusing on the concepts and perspectives of I4.0, in his work Lu (2017) briefly described I4.0 as “an integrated, adapted, optimized, service-oriented, and interoperable, manufacturing process which is correlate with algorithms, big data, and high technologies.” Therefore, it is possible to identify the main guiding *Principles* of I4.0 (Shafiq, Sanin, Toro, Szczerbicki, 2015):

- Interoperability: Collaboration between all the components of the ecosystem
- Virtualization: Work in a scenario simulation logic
- Decentralization: Decision taken collectively by the smart network, which are reflected in quicker decision-making and problem solving
- Real-Time Capability: Improve efficiency thanks to real-time information acquisition and analysis
- Service Orientation model: Services leveraged via Internet of Services
- Modularity: Specific of flexible manufacturing system, based on the subdivision of the manufacturing process in simpler tasks

## Connecting physical and digital world

Industry 4.0 has the aim of connecting data into the equation of productivity in the company in a structured way. To do this, it is essential to merge two different worlds, thus attempting to create a synergy between operational technologies (OT) and information technologies (IT). Operational technologies are the central part of manufacturing and industrial environment, in which people deal with technology connected with processes, partially exploiting digitalization and their main scope is the monitoring of industrial systems performances. Whereas Information technologies are related to the use of physical devices, to create, manage and exchange all forms of electronic data. The basic idea is to merge these two worlds, seeking to turn the industrial automation work similar to an IT system.

I4.0 promotes the concept of Smart manufacturing which is mainly based on the technological concepts of Cyber-Physical Systems (CPS) and Internet of Things (IoT) (Böhler 2012; Hermann, Pentek, and Otto 2015).

- CPS are industrial automation systems of collaborating computational entities, enabling the intensive connection of the physical reality operations with computing and communication technologies. It consists of microcontrollers controlling sensors and actuators, enabling the data exchange (Lu, 2017).
- IoT represents the multidirectional integration between digital and physical world. Indeed, it is the inter-networking of physical objects with software, sensors, actuators and electronics that enables the collection and exchange of data, thanks to embedding computing and communication capabilities of objects. IoT can be considered as source of Big Data and Cloud computing as an enabler of IoT data processing capabilities always available.

Considering that IT technologies are moving toward the OT and vice versa, it is necessary to highlight this trend the incremented use of operational and physical devices – smart objects – endowed with computational capacity.

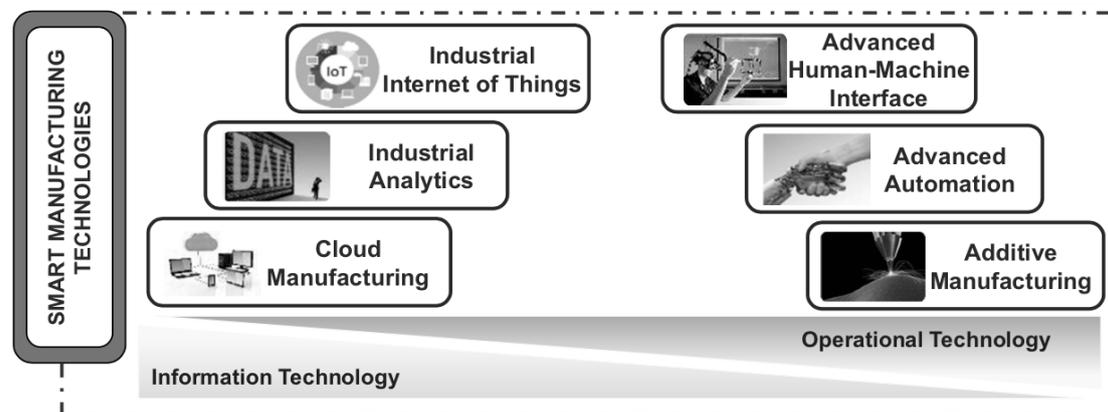


Figure 17 Smart manufacturing technologies (Politecnico di Milano)

Industry 4.0 brings together a broad array of new technologies to create value. According to the two different clusters (IT, OT) the main technology drivers identifiable are:

- Industrial Internet of Things, declination of IoT to the manufacturing world.

- Industrial Analytics focus on the methodologies and tools enabling the dataset examination and the conclusion drawing. Moreover, it allows also Big Data management, treatment and elaboration.
- Cloud Manufacturing is the application in the manufacturing field of cloud computing paradigm, it entitles using the network the diffused access to a set of manufacturing resources
- Advanced Human-Machine Interface used to acquire and transmit tactile, auditory and visual information (e.g. augmented reality viewers)
- Advanced Automation refers to the use of advanced manufacturing solutions (e.g. interconnected and collaborative robots)
- Additive Manufacturing is a kind of production process, where the material is joined through 3D printers connected to digital development systems

Concerning the Smart manufacturing field, digital technologies acts principally in three main areas:

- Smart Lifecycle Management, embracing the entire new product development process
- Smart Supply Chain, planning of physical and financial flows in the extended supply chain
- Smart Factory is connected to overall infrastructure and services (production, internal and external logistics, maintenance, quality, safety and compliance with rules)

Considering the processes involved in the deployment of the new technologies, the Smart Factory could be considered as the heart of digital transformations, with the major number of applications, approximately 80%. In the Smart Lifecycle field, the relative number of applications is around the 13%. The less widespread area is the Smart Supply Chain one (around 7%), even though the growing interest of organisations in achieving a better integration and collaboration in the Supply Chain (Politecnico di Milano, 2018).

### 3.5.1 Big Data

Leveraging Big Data as information is the heart of competition and growth for industrial businesses. Data-driven strategies and decisions based on real-time and historical process information are a fundamental point in performances optimization.

The term Big Data was coined in 1997 by two NASA researchers referring to the visualization challenge of large datasets (Cox, Ellsworth, 1997). Gartner (2014) provided a generally accepted definition of Big Data: “high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”. Big Data refers to the acquisition, processing and use of large amount of data, made possible by the advent of automated recording of data. Their exploitation represents a huge opportunity for businesses to improve their performances. In general, Big Data are able to potentially create value creating transparency, enabling the discovery of needs, segmenting populations, innovating business models, products and services, supporting human decision making (McKinsey, 2011). For example, they can be employed as an input for new marketing and advertising campaigns, as a detector to spot emerging trends in the market and as input for customer relationship management and to analyse customer group behaviours.

Big Data covers 5 dimensions (5 V’s of Big Data), representing the principal challenges faced by companies contemplating to make full use of them (Benabdellah, Benghabrit & Bouhaddou, 2016).

Aspects	Definition
Volume	Massive amount of data generated endlessly
Velocity	Speed of data produced and collected in real time and in streaming
Variety	Different data sources and heterogeneous formats: structured, semi-structured and unstructured data
Veracity	Orderliness, messiness and trustworthiness of available data
Value	Subjective aspect related to the non-exploitation of these data

Table 5 5V's Big data

Several new issues arise from the advent of Big Data. In fact, the storing and processing of massive amounts of data is problematic, such as the selection of interesting data and information to use among the available data. Another issue not to be underestimated is dealing with data policies related to ownership, confidentiality and security.

### 3.5.2 Impact on the organisations

I4.0 requires organizations a deep transition to handle the overall complexity. Indeed, it is designed to make manufacturing more efficient and forces companies to rethink the way they do business (McKinsey, 2015). The disruption of the traditional value chain creates the foundation for new business model for the industry, unlocking new value. The final scope is the enhancement of efficiency and competitiveness through a better cooperation and interconnection along the value chain, focusing on the end-to-end digitalisation.

Thanks to Big Data, managers are capable to measure and acquire more information about their businesses and directly translate this knowledge into improved decision-making and performance. In particular, companies, characterizing themselves as data-driven, reach better performances in financial and operational terms. Companies are beginning to turn these large-scale quantities of data into competitive advantage, enabling the resolution of complex problems in all the different Supply Chain Management (SCM) processes (Benabdellah, Benghabrit & Bouhaddou, 2016).

Several are the challenges to consider in the use of Big Data, from the technical to the managerial ones. The management challenges are five: leadership, talent management, technology, decision-making and company culture (McAfee and Brynjolfsson, 2012).

## 3.6 Data management and Decision support system

Decision support systems are defined as an “interactive computer-based systems that help people use computer communications, data, documents, knowledge, and models to solve problems and make decisions” (Power 2002). These systems must be viewed as auxiliary tools, not replacing the decision makers, but supporting and helping them in finding the optimal decision. Multiple types of DSS have been developed to support decision makers; particularly two macro-classes could be distinguished: on one side systems that are able to provide immediately structured data and information to the decision maker, on the other side systems capable to collect knowledge and to make it available when required to the user.

Even though DSS is a broad concept, three main features may be described:

- These systems are projected to improve decision-making process;
- The major aim of decision support systems is to support the decision makers, not to create a completely automatic process;
- These systems should be able to answer to the ever-changing requirements of the decision makers.

### 3.6.1 Hierarchical view of Data Information Knowledge & Wisdom

The conventional hierarchy from data, information, knowledge and finally to wisdom, commonly known as ‘Knowledge Hierarchy’, ‘Information Hierarchy’ or ‘Knowledge Pyramid’ is one of the basic models in the knowledge literature, firstly appearing in Eliot’s article (The Rock”, 1934). In particular, the three concepts of data, information and knowledge have been addressed by multiple authors. These ones are fundamental in the context of information science becoming the basic build block of the fields (Zins, C., 2007). Across the years, different meanings have been identified for each one, pointing out the nature of the relationship linking them. In general, the majority of the authors defined data as the raw number and facts, information as the processed data and knowledge as the authenticated

information (Alavi, M., & Leidner, D. E., 2001). This underlines that many authors' view is that data information and knowledge are in sequential order identifying the data as the raw material of information and information as the raw material of knowledge.

In traditional pyramid these four concepts are described as distinct categories among themselves:

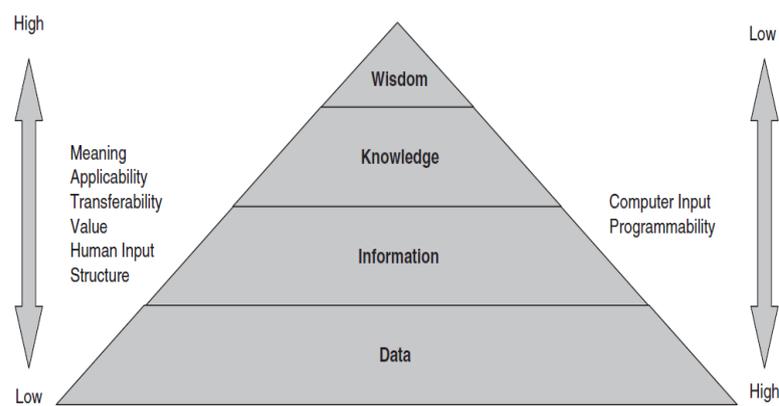


Figure 18. DIKW Hierarchy (Rowley, J., 2007)

**Data** have been seen as a fact that could exist in any form, usable or not, that needs to be structured to become information; it simply exists without having a significance beyond its existence. Data need to be processed in order to have a value and then become usable information. A more practical definition of data is given by Vercellis in 2009 that states: “Generally, data represent a structured codification of single primary entities, as well as of transactions involving two or more primary entities. For example, for a retailer data refer to primary entities such as customers, points of sale and items, while sales receipts represent the commercial transactions”.

**Information** comes from the data generation and processing, containing a real description of the events. Information are able to answer to questions that begin with question-words who, when and where (Ackoff, R. L., 1989). Therefore, the information are those data that have been extracted, processed and aggregated in order to describe a particular context, condition or situation. Relational connections are detected assigning a meaning and a value to these data.

*Knowledge* is the personalized information, processed in the mind of individuals related to fact procedures, concepts, interpretations, ideas, observation and judgment (Alavi, M., & Leidner, D. E., 2001). It could be defined as a product of a synthesis in the knowing person that could exist only in his mind. Information becomes knowledge when it is possible to transform it into instruction that makes possible to control a system. In other word information is transformed into knowledge when it is used to make decisions and develop the corresponding actions (C.Vercellis, 2009). Knowledge answer to the question how. It is possible to distinguish between two types of knowledge: *explicit* and *tacit*. The first one, the explicit knowledge is contained and represented by tangible media, such as word, images and audio recordings. It is easily representable and communicable by languages, symbols, rules and objects. On the other hand, the tacit knowledge is difficult to articulate, to text or drawing since it “tends to reside within the heads of knowers” (Dalkir, K. 2013). It is the results of the subjective experience of an individual, including mental model, beliefs and perspectives. As stated by Dalkir, K. (2013) in his work typically, the more tacit knowledge is, the more valuable it tends to be.

On the top of the pyramid the concept of *Wisdom* is defined as the collective and individual experience of applying knowledge to the solution of problems. It is an extrapolative and non-deterministic process, that implies the function of judgment that is related to ethical and aesthetic personal values. The main difference between wisdom and the other levels is that the answer obtained is not easily achievable, requiring the judgment between what is wrong and what is right (Bellinger, Castro & Mills, 2004). In simple words, wisdom could be defined as knowledge that has been processed in some meaningful way based on the personal experience (Faucher, Everett & Lawson, 2008).

In the literature different view of this classification are provided. Ackoff (1989) proposed a classification made of five different categories distinguish between the following five concepts: data, information, knowledge, understanding and wisdom. The additional stage, understanding, is defined as the process by which a person makes a synthesis of the knowledge previously held in order to create a new one, learning from something that is previously known. Thanks to the understanding people own the ability to undertake useful action basing on something that was understood beforehand. Tuomi (1999) in his work

“Data is more than knowledge” propose a totally different view of the traditional hierarchy, inverting the order and the relationship among the concepts of data information and knowledge. In fact, its methodology assumes that “knowledge must exist before information could be formulated and before data can be measured to form information” (Alavi & Leidner, 2001). The author state that a rethink on the relationship among these three concepts is required in order to develop information system support for knowledge management. The author states data are the once that emerges last only after knowledge and information are available; these are created starting from the knowledge of someone as a solution to practical problem.

### 3.6.2 Knowledge management vs Business Intelligence

Knowledge management is a widely used terms which refers to collecting, identifying and leveraging the knowledge within an organization in order to support and help an organization in competing (Alavi, M., & Leidner, D. E. 2001). It could be defined as the process of capturing, structuring and managing the knowledge throughout an organization in order to develop and reuse best practice, to work faster and less costly. One of the main issues concerning this discipline is to capture, codify and share the knowledge owned by each person within the organization (Dalkir, K. 2013). In fact, the new knowledge is created through the relationship between tacit and explicit one.

According to Nemati et al. (2002) the shift from tacit to explicit knowledge is possible through four main steps: socialization, articulation, integration/combination and understanding - internalization. Thought the first step, the socialization there is the exchange of experiences, skills and all the possible kind of tacit knowledge among all the individuals. In the externalization step there is the passage from tacit to explicit knowledge through for example brainstorming, determination of casual map and knowledge extraction expert systems. Then new explicit knowledge is created in the combination step, finding new pattern and relationship among the already existing knowledge. Finally, in the last step the new relationship found are proved and tested and therefore turning into new tacit knowledge.

The aim of knowledge management is to support the workers in the decision-making process, making the knowledge visible within the organization, creating a knowledge sharing culture and building a knowledge infrastructure that encourage people to interact and collaborate. It requires different set of activities, in particular it is based on the following four basic process: creating, storing/retrieving, transferring and applying knowledge (Alavi & Leidner, 2001).

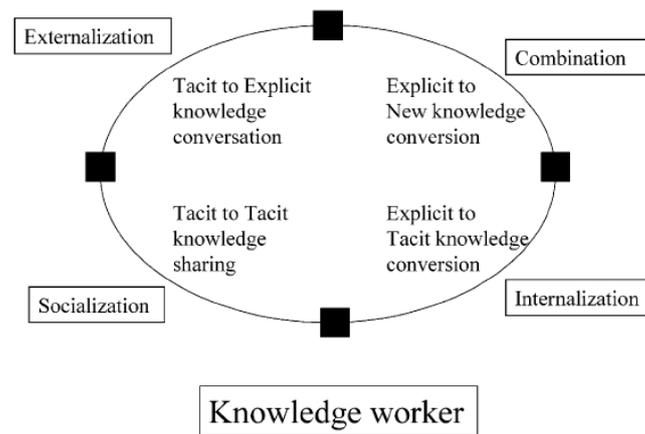


Figure 19 From Tacit to Explicit Knowledge (Nemati et al., 2002)

The concept of knowledge management could be compared to the Business Intelligence one. C.Vercellis (2009) in his work “Business Intelligence – Data Mining and Optimization for decision making” defines Business intelligence as “a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making process.” From this definition knowledge management and business intelligence could be presented as similar concept and disciplines owing similar objective. In fact, their main aim is to support people in the decision-making process developing a knowledge information environment. Nowadays in the organizations, the decision-making process are becoming more complex and dynamic. In order to make effective and timely decisions, they should be based on analytical and mathematical method and not only on intuitive methodologies, experience and personal knowledge. The ability of the individuals to make decisions is one of the major factors that influence the organization competitiveness.

According to C. Vercellis (2009) is possible to define a distinction between the two disciplines:

- *Knowledge Management* is focused on the processing and treatment of unstructured and implicit information mainly founded in documents and past experience.
- *Business intelligence* is focused on the processing and treatment of more structured information and organized in databases.

Notwithstanding this distinction is not that strict and strong leading to a intertwine of these two concepts. For example, there are different data analysis techniques such as text mining and web mining that give to business intelligence the ability to analyse also unstructured information.

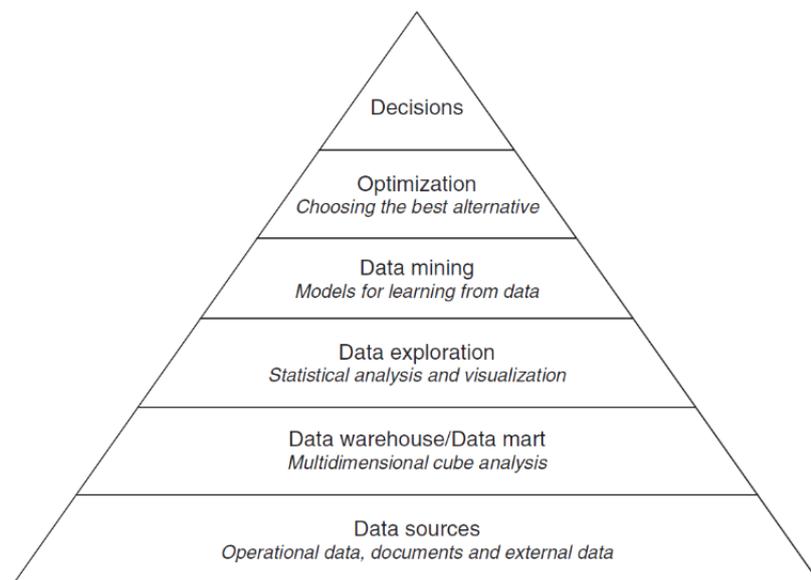


Figure 20 Components of a Business Intelligence systems (C.Vercellis, 2009)

A business intelligence system provide support to the decision makers thanks to knowledge and information captured for the data through the use of mathematical model and analytical application. The business intelligence process is composed mainly by two activities: getting data in and getting data out (Watson & Wixom, 2007). It describes how the data are collected, stored, treated and processed to extract firstly information, then knowledge in order to select the best option among different alternatives in a decision-making process. It

is possible to define six main components of a business intelligence system: data sources, data warehouse, data analysis, data mining, optimization and decisions.

## Data sources

Data sources are the first components highlighted on the bottom of the pyramid. Indeed, managing data, the initial stage is to collect and integrate data coming from primary and secondary sources. A distinction could be made between structured data mainly coming from internal sources and operational databases and unstructured data mostly obtained from external sources providing unstructured documents such as email. The literature offers distinct data classification whose aim is to explore in detail the data nature and its origin. Different authors distinguish between three main type of data: internal data, external data and personal data. The aim of knowledge management systems is the integration of structured data coming from external and internal sources with the structured ones coming from the personal sources.

The first category, *Internal data*, refers to the everyday operation and transaction involving the main entities within company processes mostly customers, suppliers, products and sales. They are stored in different systems characterizing the everyday company management such as administration, accounting, production and logistics. Internal data are generated and collected by the following components of the information system:

- *back-office systems*, that manage basic transactional records such as orders, invoices, inventories, production and logistics data;
- *front-office systems*, that deal with data from call-center activities, customer assistance, execution of marketing campaigns;
- *web-based systems*, that regard sales transactions on e-commerce websites, visits to websites, data available on forms filled out by existing and prospective customers.

These data are gathered through transactional applications and the whole set of the application enabling the internal data storing is called Enterprise Resource Planning (ERP).

The *External data* refer to those ones coming from company external sources and these are mostly used to determine and enhance the wealth of data stored in internal databases. This data category is able to provide also additional and complementary information collected by

external operators, such as agencies, related to sales, future trend predictions for specific business industries, economic and financial indicators, customers opinions.

Nowadays in Industry 4.0 context data acquisition and collection is strongly increasing, enabling companies to collect massive amount of data. Multiple data sources allow to acquire both internal and external data from different technologies such as machinery, sensors, CPS, IoT devices and communication network (Khan, Wu, Xu & Dou 2017).

Finally, *Personal data* refer to those data, usually unstructured, coming from the personal skill and knowledge of the decision makers. Those data are used in the decision-making process and are stored inside worksheets or local databases in the decision maker personal computer or mobile.

A further distinction between data is provided by Cleven & Wortmann (2010) in their article distinguishing between two main categories: data domain-subject area and metadata. The first one is related to the business domain at hand, instead metadata is connected to the domain data.

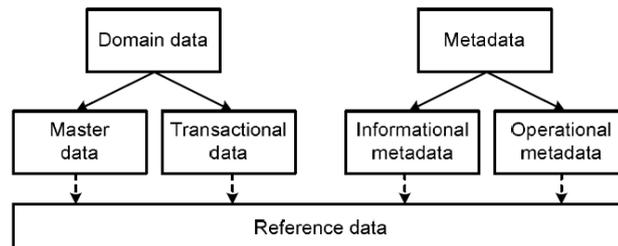


Figure 21 Data taxonomy (Cleven & Wortmann, 2010)

The *Data domain* category could be divided into master and transactional data. A *Master data* definition is provided by Loshin (2009): “Master data objects, are those core business objects used in the different applications across the organisation, along with their associated attributes, definitions, roles, connections, and taxonomies.” Typical master data entities are customers, employees, vendors, suppliers, products, locations, profiles, contracts representing the building blocks of the companies’ databases. *Transactional data* constitute all business transactions, typically sales orders, production requests or invoices. *Meta data* is represented by two classes: on one hand there the *operational metadata* which aim is to facilitate design and technical operation of information system; on the other hand,

the *informational metadata* whose objective is to enable and facilitate the understanding and access of domain data.

## Data warehouse & Data marts

All the type of data coming from different sources are stored in databases known as Data Warehouses and Data Marts through the three main steps: extraction, transformation and loading (ETL). During the first phase, *extraction*, the selected data are extracted from the internal and external sources. Through the second phase the data quality is enhanced through data correction and cleaning. This phase is required in order to ensure homogeneity and integration among all the data coming from the different sources. The last step is to load the extracted and transformed data into the data warehouse making them available for the decision-making process.

Data marts could be considered a departmental data warehouse with smaller size, containing data of a specific company department. Therefore, data stored in the data marts are a sub-category of the data stored in the company data warehouse, related to a specific domain/department. The aim of data warehouse is to provide an input to the tools performing business intelligence analysis and supporting decision making process, known as on-line analytical processing applications (OLAP).

Data warehouses provide a support to the decision-making process through the on-line analytical processing tools (OLAP) and analytical database. OLAP was firstly defined in 1933 by Codd et. al. as “the dynamic synthesis, analysis, and consolidation of large volumes of multidimensional data”. These tools are able to process data in different way analysing data through various dimension, performing what-if analyses and discovering hidden pattern (Gorla, N., 2003). The main capabilities are to view the data through different level of detail, slicing, dicing, aggregating and disaggregated data.

## Data exploration & Data Mining

After data have been collected and stored within data warehouses and data marts. These are available to be processed by “*passive business intelligence analysis*” such as query, reporting system and statistical methods. According to Vercellis (2009) these methodologies

are defined as passive ones, since the decision maker, before applying analysis tools, needs to set hypothesis and define data extraction criteria. The aim of data exploration is to identify the relationships within data set through three main steps: univariate analysis, where each attribute of a dataset is analysed as a single entity; bivariate analysis, where the attributes are analysed and treated in pairs to identify a possible relationship between them; and finally multivariate analysis, where multiple attributes are analysed to evaluate existing relationships within the dataset. These statistical analyses mainly use graphical methods - such as scatter plots and box plots - and summary statistics - such as variance and mean in the univariate analysis or covariance and correlation in the bivariate analysis.

The third level of the pyramid includes data mining and machine learning techniques whose purpose is to learn from the data, extracting information and knowledge. The aim is to find hidden patterns and relationships in a dataset and infer rules to predict future behaviour, performing both descriptive, predictive and prescriptive analysis.

- *Descriptive analysis* aims to summarize data in meaningful graphs and to obtain standardized and customized reports to identify possible patterns and trends in data. Example of descriptive models are clustering, summarization, association rules and sequence discovery (Deshpande & Thakare, 2010).
- Through *Predictive analysis* the past performances and historical data are analysed to predict future behaviour, trends, risks and to find relationships not commonly identified with traditional statistical techniques (Evans & Lindner, 2012). It predicts unknown values exploring and using the known values, applying classification, regression, time series analysis and prediction models (Deshpande & Thakare, 2010). These analyses answer the question "what could happen".
- *Prescriptive analysis* objective is to give suggestions concerning the best possible action that an individual could undertake. Feedback systems capable of tracing the results produced by the undertaken action are required by these types of studies. Prescriptive analysis is able to answer the question "what should happen".

Data mining techniques help organizations to make proactive knowledge decisions, going beyond the analysis provided by traditional tools.

The term data mining is mostly used by statisticians, data base researchers and management information systems communities. It is frequently used as a synonymous of *Knowledge Discovery in Databases (KDD)* but in reality, data mining is a part of the KDD process (Deshpande & Thakare, 2010). KDD is defined as the process to extract knowledge from high volume of data, identifying “valid, novel, potentially useful, and ultimately understandable patterns in data” (Fayyad et al., 1996). The process starts learning and creating a target dataset which is cleaned and processed to be easily handled. Subsequently, useful tools and features are required to represent and transform the cleaned data, depending on the task’s goal. A data mining algorithm, such as classification rules or tree, regression, clustering, sequence modelling, dependency and line analysis must be selected to search pattern of interest. The identified pattern and rules must be interpreted and evaluated: on one hand removing redundant or insignificant pattern and on the other hand, the significant one must be made understandable for the users/decision makers.

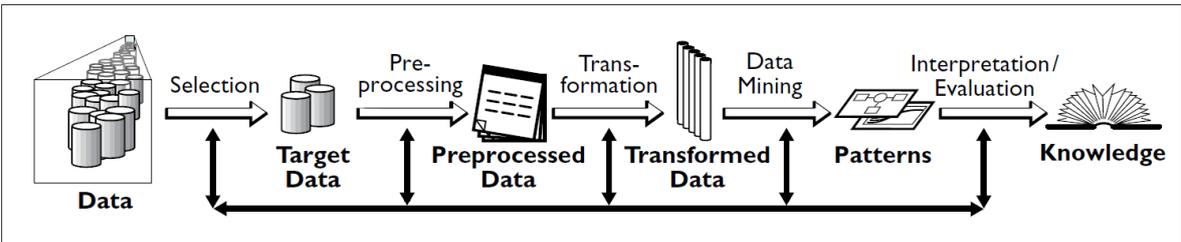


Figure 22 KDD process (Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P., 1996)

### Optimization & Decisions

Optimization models are adopted in order to select the best solution among a set of alternatives, analysing the problem from different aspects. These are useful but time-consuming models in supporting a decision-making process. On the top of the pyramid the natural conclusion of a decision-making process is the selection of the identified best solution. Necessary, the selection is driven also by the unstructured information owned by the decision makers, even if a business intelligence methodology is adopted. In fact, the aim of a business intelligence system is to make the decision more structured and objective as much as possible.

## 3.7 Data Quality Assessment

Data quality plays a crucial role in all organizations, business and government application. It is a significant performance issue affecting both the operating activities and the decision-making process. Different studies state that quality could be defined both as a multi-dimensional and hierarchical concept where each data quality dimension is related to a particular aspect of data (Wang, Reddy, & Kon, 1995; Liu & Chi, 2002; Capiello, Francalanci, & Pernici, 2004). The skill to assess data plays a crucial role within organizations since it enables to monitor and to evaluate the as is situation and then to detect the possible area of improvement.

In literature it has different definition and understanding. According to quality management “quality is appropriate for use or to meet user needs or it is quality of data to meet customer needs” (Sidi, Panahy, Affendey, Jabar, Ibrahim, & Mustapha, 2012). A common definition of data quality proposed by several authors is “fitness to use” (Pipino, Lee, & Wang, 2002; Capiello, Francalanci, & Pernici, 2004) which means the capability of data to satisfy the user requests and specifications.

The data quality issue is constantly increasing its importance due to several factors affecting the organization. Information systems are continuously changing their structure going from a hierarchical one to a network-based structure increasing exponentially the number of potential data sources within a company. Another important aspect concerning the quality issue is the advent of the big data, as deeply explained by Cai & Zhu (2015) in their work. Big data makes easier and faster the generation of huge volume of heterogenous data acquiring them from various sources. Nowadays organization are becoming more conscious about the benefits achievable through this massive amount of data, such as improve the service quality provided, better understanding of the customer needs and desires. In this context, data quality issue plays a crucial role, becoming a necessary condition to generate and create the value form big data analysis.

### 3.7.1 Data quality assessment methodologies

Assessing data quality is still an open issue since a rigorous method has not been defined yet, as instead happened to monitor quality in the manufacturer field. The literature offers a wide variety of techniques to assess data quality identifying a different set of dimensions and attributes, but it is still open the request to have an objective measure and a comprehensive range of metrics to assess the quality of data (Cappiello, Francalanci, & Pernici, 2004).

Wang et. Al. (1995) in his work “Toward quality data: An attribute-based approach” present a hierarchal view of data quality, identifying the following four macro dimensions accessibility, interpretability, usefulness and believability. According to the author, firstly it is required to verify if the data user owns the means to get data access. The second dimension taken into consideration is the understandability, since data should be easy to interpret looking their syntax and semantics. In order to support the decision-making process, data must be considered useful for a determinate task. Therefore, two different data aspects should be assessed: on one hand the data must be relevant for the task at hand, and on the other hand data must be timely. The author describes the timely dimension through two characteristics; the data currency referring to “when the data item was stored in the database” and data volatility referring to “how long the items remain valid”. Lastly is required to verify the data believability through the assessment of four different dimensions: completeness, consistency, credibility and accuracy.

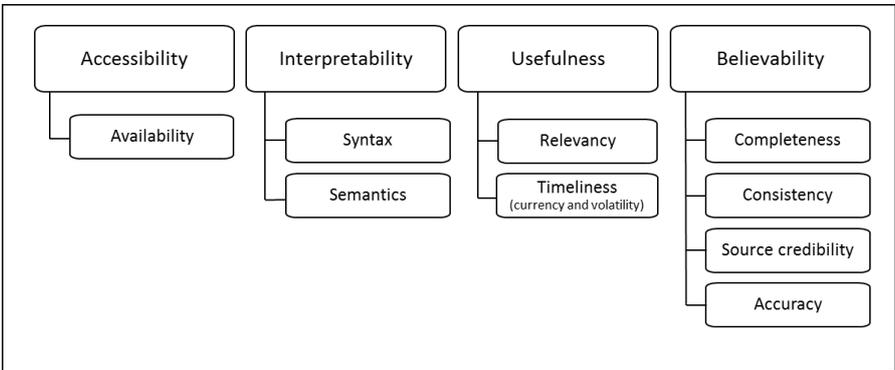


Figure 23 Data Quality Hierarchy (Wang et al. 1995)

Liu & Chi (2002) introduce a new perspective of Data Quality, proposing a concept of theory-specific data quality which states that “Data Quality is defined and measured as the extent to which data meet the needs and the specifications of a theory”. The starting point of this methodology takes into consideration the data evolution during the lifecycle distinguish among four consequential stages:

- *Data Collection* where data are observed and catch from the real world;
- *Data Organization* where data are orchestrated in different data bases;
- *Data Presentation* where data that have been already collected and organized are then process and treated in order to be properly presented to different users;
- *Data Application* where data are used and applied to solve a certain task.

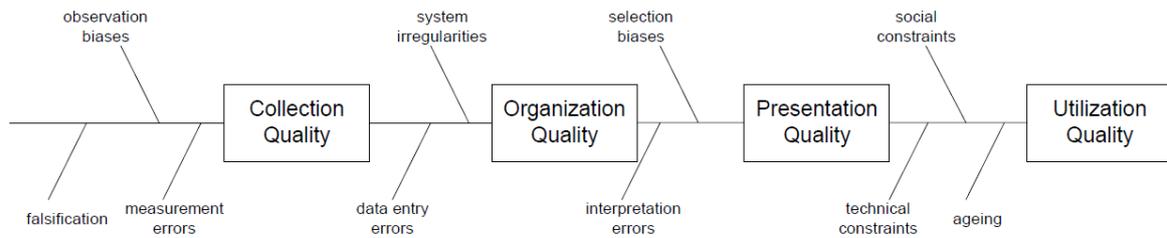


Figure 24 Fishbone of the Evolutional data Quality (Liu & Chi 2002)

Basing on the data evolutionary stages the models defines four different hierarchal views of data quality: collection quality, organization quality, presentation quality and lastly utilization quality. The aim of each view is to measure the quality of the collected data, stored data, presented data and utilized data. In order to view data quality to be both theory-specific and evolutionary, two ancillary methodologies are presented by the authors. The first one is the Fishbone diagram, applied to analyse which are the different causes of data inaccuracy during its lifecycle. The fishbone diagram is a useful managerial tool to resume which are all the factors that lead to poor data quality. The second methodology instead analyses how at each stage of the *Data evolutionary life cycle*, a particular data matches with the requirements of a theory. The model proposed put its attention on the fact that data quality could not be an objective assessment, without considering the theory outside.

In data and information quality field a noteworthy model is the AIMQ - A methodology for information quality assessment (Lee, Strong, Kahn, & Wang 2002). The proposed model is the only one in literature that evaluate the quality of information concentrated on benchmarking, assessing information objectively. The model classifies the quality dimensions identified by Wang & Strong (1996), into a matrix basing on the importance perceived by the data users. The axes of the matrix are “conformity to specifications” and “conformity to users’ expectations”, identifying four classes of information: dependable, usable, sound and useful.

Following, each dimension is evaluated through two questionnaires: the goal of the first one is to understand which the most significant quality dimensions are; instead through the second one each quality dimension is evaluated. The last step of the model is to interpret the results with benchmarks stressed in literature.

	Conforms to specifications	Meets or exceeds consumer expectations
Product Quality	Sound information	Useful information
	Free-of-error	Appropriate amount
	Concise representation	Relevancy
	Completeness	Understandability
	Consistent representation	Interpretability
		Objectivity
Service Quality	Dependable information	Usable information
	Timeliness	Believability
	Security	Accessibility
		Ease of operation
		Reputation

Figure 25 The information quality model

The Data quality assessment methodology proposed by Pipino et al (2002) aims to provide a guideline to define the data quality metrics, independently from the specific context and scenario. Three steps could be identified in the model application:

- First step: an objective and subjective evaluation of the quality dimension is done; the author identifies multiple quality dimensions, without providing a hierarchal view of them.
- Second step: the two different assessments must be compared and evaluated in order to identify possible inconsistencies.

- Third step: the aim is firstly to underline which are the causes of inconsistencies between the objective and subjective assessments and then to identify the possible areas of improvement.

Bovee et al. (2003) proposed a new framework to evaluate information quality, highlighting that this is a subjective evaluation that should be treated taking into consideration the user perceptions. The model presented has a hierarchical structure made by four macro quality dimensions. Firstly, the information accessibility should be analysed, since if information is not accessible all the other qualities and factors becomes non-significant. Secondly the information must be meaningful and easily understandable, in order to be managed and exploited by the users. The third sub-attribute is information relevancy; in order to evaluate this dimension, the information timeliness has a considerable importance since a too dated information is considered not useful. Finally, the information integrity is presented through four sub-attributes: accuracy, completeness, consistency and existence. In this work firstly appear the sub attribute existence deemed necessary by the authors to lead to obtain valid conclusion.

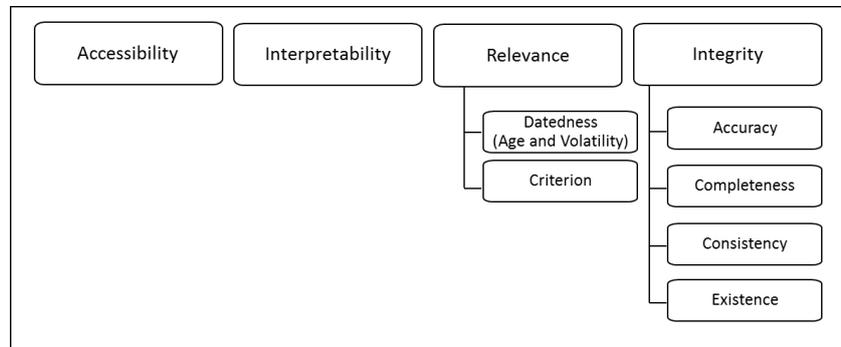


Figure 26 Data Quality Hierarchy (Bovee et al. 2003)

## Data quality assessment and Big data

The big data era is changing organizations which are becoming more skilled at collect and analyse unprecedented scale of data. In this context, data quality field is becoming more important than ever, marking the transition from a close-world assumptions of database systems, where all the data semantics could be specified beforehand, to an open world view where rules are defined ex post, learning from data (Saha & Srivastava 2014). Cai & Zhu (2015) discussed data quality issue under this new perspective, noting the new challenges

triggered by the big data introduction. The new issue is to extract and collect high quality data from the multiple sources available in big data context. Back in time organizations mostly generate and collect data from their internal databases making easier the data quality evolution. Nowadays data within enterprises are massively increasing and constantly changing, requiring higher effort to assess data quality.

The data quality framework proposed is defined exploiting the traditional approaches presented in literature and additionally taking into consideration the Big Data advent. The hierarchal model is made up of five different dimensions of data quality – availability, usability, reliability, relevance and presentation quality. While the first four dimensions are viewed as fundamental to assess data quality, the last dimension may be treated as a plus to improve customer satisfaction. Then, each dimension is described from one to five elements as shown in *Figure 27*. Some elements presented within the model are particularly suitable to the Big Data era, for example:

- Structure is defined as “the level of difficulty in transforming semi-structured or unstructured data to structured data through technology”. This element plays a crucial role in Big Data context since among the huge amount of data collected more than 80% is usually unstructured.
- Metadata are crucial in big data era since with the growth of data types and source, the users might easier twist the meaning and concept of data. Thus, metadata are necessary to decrease inconstancy and misunderstanding, describing the different features of the various datasets.

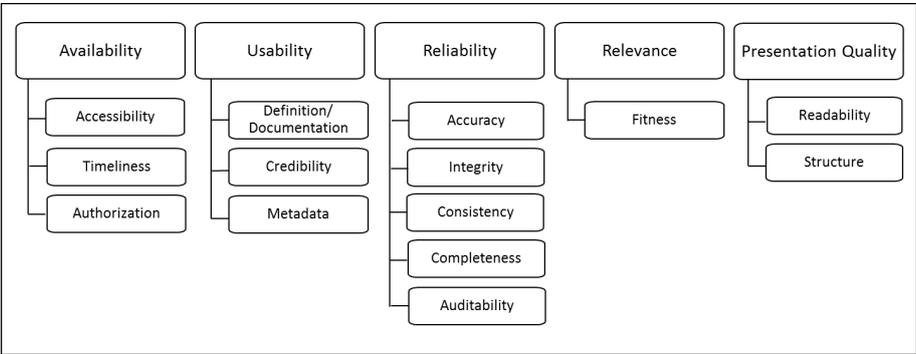


Figure 27 Data Quality Hierarchy (Cai & Zhu 2015)

Analysing several approaches to assess data quality proposed in literature, the majority of the authors considers mainly the following dimensions: accessibility, accuracy, completeness, interpretability and timeliness. *Table 6* summarises the data/information quality model analysed in literature, underlining the different hierarchical perspective of each model and the different dimensions identified.

References	Object	Quality dimensions
<i>Lee et al. 2002</i>	Information	Accessibility, appropriate amount, believability, completeness, concise representation, consistent representation ease or operation, free of error, interpretability, objectivity, relevancy, reputation, security, timeliness, understandability
<i>Pipino et al. 2002</i>	Data	Accessibility, appropriate amount of data, believability, completeness, concise representation, consistent representation, ease of manipulation, free of error, interpretability, objectivity, relevancy, reputation, security, timeliness, understandability, value-added
<i>Cai et al. 2015</i>	Big data	Availability (accessibility, timeliness, authorization), usability (definition, credibility, metadata); reliability (accuracy, integrity, consistency, completeness, auditability); relevance (fitness); presentation quality (readability, structure)
<i>Bovee et al. 2003</i>	Information	Integrity (accuracy, completeness, consistency, existence); accessibility; interpretability; relevance (datedness – age and volatility, criterion)
<i>Wang et al. 1995</i>	Data	Accessibility (availability); interpretability (syntax, semantics); usefulness (relevancy, timeliness – current, non-volatile); believability (completeness, consistency, source credibility, accuracy)
<i>Liu et al. 2002</i>	Information	Collection quality (accuracy, objectivity, trustworthiness of the collector, completeness, clarity) Organization Quality (collection quality, reliability of data clerk, consistency, storage efficiency, retrieval efficiency, navigability); Presentation Quality (organization quality, faithfulness, neutrality, interpretability, formality, semantic stability) ; Application quality (presentation quality, ease of manipulation, timeliness, privacy, security, relevancy, appropriate amount of information)

*Table 6 Data Quality Methodologies*

### 3.7.2 Application context and User’s perspective

Different authors agreed to the fact that quality is not related just to its feature, but it is strictly connected to the application context and to the users (Cappiello, Francalanci, &

Pernici, 2004). Data quality could not be considered as an objective valuation since individuals using same data for different scope could ask for different quality requirements. Moreover, the models proposed in literature tend to assess quality mostly putting their attention on the data quality dimensions, without taking into consideration the application context and real use of data. According to Cappiello et al. (2004) in order to obtain a real evaluation of data quality, a user's perspective must be considered, analysing which are the user expectations of a particular database within a specific process.

In fact, also according to Pipino et al (2002), the data quality assessment could be both objective and subjective and it could be made performing two different paths:

- *task-dependent assessment* where the data evaluation is contextualized to a specific decision-situation;
- *task-independent assessment* where the data are evaluated without consider the contextual application.

In literature the objective evaluation of data quality dimensions through algorithm is presented just for limited case. For example, Naumann & Freytag (2005) in their work presented an approach to evaluate the information sources completeness, defined as the ratio between the information amount owned and the total information presented in the real world. Generally, in order to perform an objective data quality assessment three functional form could be identified (Pipino, Lee, & Wang, 2002).

*Simple ratio.* The simple ratio is defined as the ratio between the total outcomes and the desired outcomes, below is shown the most used form:

$$\text{Simple ratio} = \frac{\#undesirable\ outcomes}{Total\ outcomes - 1}$$

This measure implies that 1 presents the best case and the most desired situation, on the other hand 0 present the worst-case scenario. Different quality dimension could be assessed applied this form such as completeness, correctness, consistency, relevancy concise representation and ease of manipulation.

*Min or max.* The minimum or maximum form is applied when it is required to consider together different quality indicators. This operator is also used to adjudicate an aggregate

value of data quality with a unique dimension and for a set of data. Through this methodology the maximum or the minimum value among a different range of values is chosen.

*Weighted average.* The weighted average is another form applied when it is required to consider together a set of quality dimensions. This dimension is suitable when the organizations are able to recognize the significance of each quality dimensions. The weight of each variable should have values from 0 to 1 in order to get normalized results.

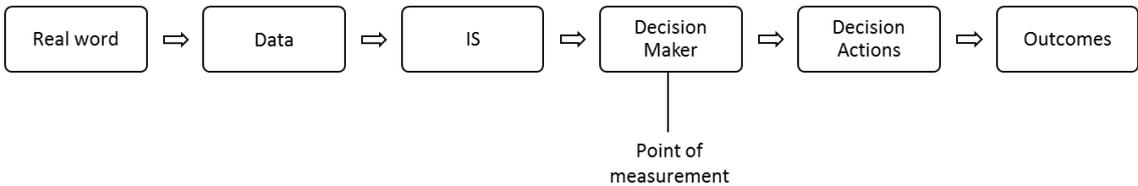


Figure 28 Measuring the perceived value of information (Ahituv, 1989)

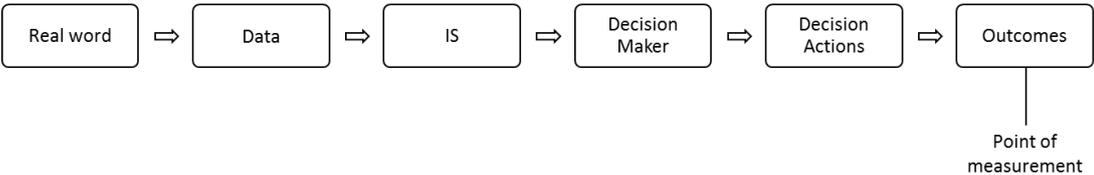


Figure 29 Measuring the realistic value of information (Ahituv, 1989)

According to Ahituv, N. (1989) the value of data changes according to perceptions of the user which should own the capability to recognize the benefits achievable from the exploitation of an information systems and to transform them into an objective value. In order to assess the perceived value of information the author proposed different methods including the following two:

- the first one, known as “*Monetary Equivalent*”, asks to the user to indicate how much he is willing to pay to keep receipting and preserving data coming from the information system. This method could be useful to compare evaluations given to different information systems and their features.

- The second one asks to the user to indicate how he perceives a particular feature of an information system, given a rank usually from one to seven. Analysing the global scores given to the features of an information systems it is possible to compare different systems and their features.

Even if these methods are subjective, they turn out to be useful and determinate in evaluating the performance of information systems especially when an information system is applied to take unstructured decisions. As shown in *Figure 28* and *Figure 29* the perceived evaluation is far from the “real outcome” of the information system and it could lead to conflicting results. In fact, it could happen that a user evaluates an information system better and more suitable for a specific task than another one, but the real outcomes of the two information systems do not reflect these perceptions.

Cappiello et al. (2004) propose a framework to support the data users in the evaluation of data quality since different users could have different perception of the same data. In the model proposed the users are categorised in classes according to similar features and all the users falling in the same class must have the same service accessibility. The model proposed is represented as follow:

#### *System characteristics*

$$U = \{u_1, u_2, \dots, u_y\}$$

$$UC = \{uc_1, uc_2, \dots, uc_i\} \quad \forall u_y \in U \quad \exists uc_i : u_y \in uc_i$$

$$S = \{s_1, s_2, \dots, s_j\}$$

$$DB = \{db_1, db_2, \dots, db_k\}$$

$$QD = \{qd_1, qd_2, \dots, qd_n\}$$

#### *Evaluation Functions*

$$CQ(uc_i, s_j) = \begin{bmatrix} CQ_{11} & \dots & CQ_{1n} \\ \dots & & \dots \\ CQ_{k1} & \dots & CQ_{kn} \end{bmatrix}$$

$$CQ_{kn} = f_{kn}(db_k, qd_n)$$

$$\text{where } f_{kn} = f(uc_i, s_j, db_k, qd_n)$$

#### *Class Level constrains*

$$CQLevel(uc_i, s_j) = \begin{bmatrix} CQLevel_{11} & CQLevel_{1n} \\ \dots & \dots \\ CQLevel_{k1} & CQLevel_{kn} \end{bmatrix}$$

Where:

- $u_Y$  = user
- $uc_i$  = user class
- $s_j$  = service
- $db_k$  = data source from which information is extracted
- $qd_Y$  = data quality dimension that must be evaluated
- $f_{kn}$  = user assessment function with class requirements
- $CQ_{kn}$  = quality function with the n-th data quality dimension for the i-th user class and the j-th service
- $CQLevel_{kn}$  = data quality level that expresses the user class requirements

Each user specified its Subjective Quality Level that could differ from the Class Quality level (CQLevel). Therefore, each user could have its own requirements differing from the belonging class. This proposed framework aim is to personalize the data quality assessment, merging two fundamental concepts in data quality field: the data relevance together with the user perceptions.

### 3.7.3 Strategy to enhance data quality

According to Batini et al. (2009), in order to improve data quality two strategies could be identified. The first one is known as Data-driven which aim is to enhance the quality of the data within an organization directly modifying the values. For example, to obtain timeliness data, a database could be updated replacing the existing data with a more current database. Different set of techniques could be helpful to enhance data quality applying a data driven strategy. The most common techniques are:

- *Data standardization or normalization* which aim is to replace or modify “nonstandard data value with corresponding values that complies with standard” (Batini et al., 2009);
- *Record linkage* which purpose is to find if the same data is represented into different tables through different representation;

- *Error localization and correction* mainly based on statistical studies; their aim is to identify those data that are not respecting a predefined quality requirement;
- *Acquisition of new data* and *Source trustworthiness* where the data quality factor drives the selection of a new data sources;
- *Data integration* where data coming from different sources are integrated and unified identifying a unique representation and format.

The second strategy is known as Process-driven, which scope is to enhance the data quality reconstructing the process applied to produce and modify data. To apply this strategy two subsequent steps are required: through the first one, namely *Process control*, the data are checked and analysed within the process; then through the second step, *Process redesign* the causes of low quality are identified and eliminated creating a new process. Comparing these two different approaches, the Process-driven strategy led to more efficient results in long time both on performances and costs, since it focuses on the root cause elimination of the quality issue. On the other hand, In the short period the process driven strategy required a significant investment and therefore, the Data-driven strategy turns out to be more efficient and suitable in the short time.

## 3.8 Maturity Model

The concepts of process or capability maturity are increasingly being applied in many domains, in order to measure competency, capability and level of sophistication. In general, maturity could be defined as “the state of being complete, perfect or ready” (Mettler, 2009; Karkkainen & Silventoinen, 2016).

The most popular way for evaluating maturity is through the use of maturity models. They are used as an instrument to conceptualize and measure the organization maturity, basing on a comprehensive set of criteria and characteristics that need to be fulfilled in accordance with the maturity level. Maturity models were developed to assist organization in their endeavour of gain and maintain competitive advantage, improving quality, reducing costs and time to market, etc. These models are also used as an evaluative and comparative basis for improvement (Fisher 2004; Harmon 2004).

Significant characteristics of maturity models are detectable, such as the maturity concept, the dimensions, the levels and the assessment approach (Lahrman, Marx, Winter, & Wortmann, 2011). More specifically, it is possible to identify three different maturity concepts. Curtis et al., 2010 defines people capability as “the level of knowledge, skills, and process abilities available for performing an organization’s business activities.” Whereas the process maturity determines the extent to which a specific process is explicitly defined, managed, measured, controlled, and effective.” (Paulk et al., 1993). According to Gericke et al., (2006) the object (or technology) maturity indicates the development level of the specific object. These models can be applied in distinct focus area structured in accordance with several dimensions, like the specific capability area, the process area and the design object. Maturity models consist of a sequence of maturity levels for a class of objects. Levels are archetypal states of maturity, distinguished by a descriptor and the relative detailed description. Maturity can be measured both qualitatively (using descriptions) or quantitatively, in a discrete or continuous manner (Kohlegger, Maier, & Thalmann, 2009).

Moreover, it is important to point out the difference between readiness and maturity, considering that the readiness assessment is conducted before engaging in the maturity

process, while the maturity assessment wants to capture the as-is situation (Schumacher et al., 2016).

A further point concerns the collocation of maturity models in-between the two concepts of *model* and *method*. Models are meant to reflect and give the exact description of the current state of a particular application domain. Thus, models generally represent a formal description of specific aspect of the reality, with the scope of better understanding and communicating. They can be either descriptive, explanatory or predictive. In short, models are aimed at delivering state descriptions (what), whereas methods are focused on the specification of activities. In fact, methods provide rules to solve issues and reach specific goals (how) (Mettler, 2009).

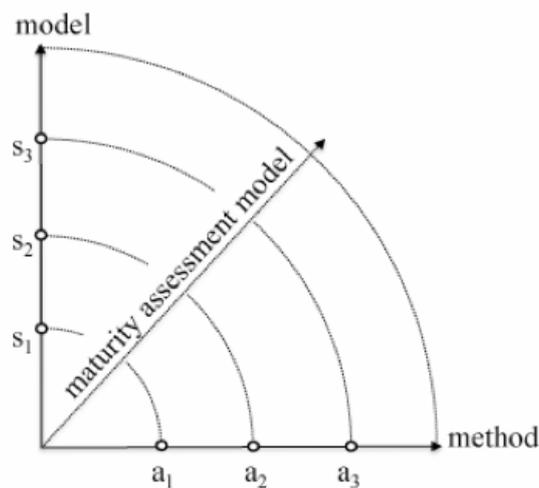


Figure 30 Positioning of maturity models' in-between models and methods (Mettler, 2011)

Indeed, maturity models combine the state descriptions (maturity levels) with the improvement activities to meet specific goals (Mettler, 2011).

### 3.8.1 Origins and overview of maturity models

Maturity approaches were initially developed in the field of quality management. Crosby's Quality Management Maturity Grid was one of the earliest examples, describing the firm behaviour at the different five levels of 'maturity' - Uncertainty, Awakening, Enlightenment, Wisdom and Certainty – for the following aspects of quality management: management

understanding an attitude, quality organizational status, handling problems, cost of quality (coq) and quality improvement activities (Fraser, Moultrie & Gregory, 2002).

Stage V: Certainty	“We know why we do not have problems with quality”
Stage IV: Wisdom	“Defect prevention is a routine part of our operation”
Stage III: Enlightenment	“Through management commitment and quality improvement we are identifying and resolving our problems”
Stage II: Awakening	“Is it absolutely necessary to always have problems with quality?”
Stage I: Uncertainty	“We don’t know why we have problems with quality”

Table 7 Part of the Quality Management Maturity Grid (Fraser, Moultrie & Gregory, 2002)

	Management concepts	Definition	System	Performance standard	Measurement
Certainty	“No reason for not doing things right”	Conform the requirements	Prevention	Zero defects	The price of non-conformance
Enlightenment	“Get serious about quality”	Satisfy customer	What do we really need to know?	Six Sigma	Complete transaction rating
Awakening	“We need to get better”	Continuous improvement	ISO 9000 MI-Q-9858	Continuous improvement	Customer complaints
Regression	“Let’s apply for the award”	Delight the customer	Buy some guru tapes and show them	Acceptable Quality Level	Benchmarking
Uncertainty	“Let’s get certified”	Goodness	Award criteria	What traffic will bear	Opinion

Table 8 The Quality Management Process Maturity Grid (Fraser, Moultrie & Gregory, 2002)

One of the main derivatives of Crosby work is the Capability Maturity Model (CMM) for software formalized by the Software Engineering Institute (SEI) – Carnegie Mellon, (Paulk, Curtis, Chrissis, & Weber, 1993). The model is a standard for the evaluation of companies’ quality level in Information and Technology sector. According to CMM, the software process maturity is defined as “the extent to which a specific process is explicitly defined, managed,

measured, controlled and effective.” The development of the model was done with the purpose of providing improvement road maps to the company in the software industry. This with the aim to enhance to a higher maturity level their processes, resulting consequently in a better service quality and savings both in terms of time and costs; thus, evolving toward a culture of software-engineering and management excellence. The range 1-5 was defined for the levels, that were labelled as ‘Initial’, ‘Repeatable’, ‘Defined’, ‘Managed’ and ‘Optimising’. The levels were further explained thanks to generic description provided for each one (Table 9).

Optimising	Continuous process improvement is enabled by quantitative feedback from the process and from piloting innovative ideas and technologies.
Managed	Detailed measures of the software process and product quality are collected. Both the software process and products are quantitatively, understood and controlled.
Defined	The software process for both management and engineering activities is documented, standardised and integrated into a standard software process for the organisation. All projects use an approved, tailored version of the organisation’s standard software process for developing and maintaining software.
Repeatable	Basic project management processes are established to track cost, schedule, and functionality. The necessary process discipline is in place to repeat earlier successes on projects with similar applications.
Initial	The software process is characterised as ad hoc, and occasionally even chaotic. Few processes are defined, and success depends on individual effort and heroics.

*Table 9 Maturity levels of the software CMM (Fraser, Moultrie & Gregory, 2002)*

The CMM staged structure is based on product quality principles presented by Walter Shewart, Edwards Deming, Joseph Juran and Philip Crosby. This kind of structure makes possible a continuous process improvement, based on several small evolutionary steps, instead of revolutionary innovations. The defined levels help the organization in the prioritization of improvements efforts (Paulk, Curtis, Chrissis, & Weber, 1993).

Since 1991, the success of the CMM determined the spread of many new CMMs in a myriad of disciplines. Some of the most remarkable include models for systems engineering, software engineering, software acquisition, workforce management and development, and Integrated Product and Process Development. However, the use of multiple models has been problematic, since applying several models not integrated within and across the

organization limits the organizations in focusing and achieving improvements successfully. To consolidate the different models was developed a single extensible framework pursuing enterprise-wide process improvement, the Capability Maturity Model Integration (CMMI, 2000).

### Types of Maturity Models

Fraser et al. (2002) build a classification of the maturity models basing on the degree of complexity, according to the mutual components present in the model. In this classification, the maturity models could be identified according with three basic groups: maturity grids, hybrids and Likert-like questionnaires, CMM-like models.

- Maturity grid: The first developed frameworks for maturity models. They consist of a matrix or a grid containing different descriptions of the performed activities for each maturity level. This typology is of moderate complexity.
- Likert-like questionnaires and Hybrids: The first ones are considered a simple form of maturity model. They are constituted of a set of questions representing a statement of 'good practice' to which a score should be assigned. The respondent selects a score to the relative performance respecting a scale from 1 to n. Hybrids model are the ones combining Likert-like questionnaire approach with definitions of the maturity levels.
- Capability Maturity Models (CMM): The architecture of this kind of models usually follow a standard format and is more formal and complex. They can be used for certifications (Maier et al., 2012). For each process area, key practices addressing to a series of goals and measures of maturity are identified.

### 3.8.2 Maturity model development framework

It is possible to highlight a certain degree of arbitrariness in the maturity models' development, taking into consideration the ever-growing number of maturity models. De Bruin et al. (2005) stated the importance of a basic standard development framework. This lies in the ability to draft a model that is highly generalizable and enables standardization. The importance is emphasised according with the aim for which the model is applied and

considering if the resulting maturity assessment is descriptive, prescriptive or comparative. Considering a model purely descriptive, there is a single point application of the model, with no action for improving maturity. This model is the best one to assess the as-is situation. Instead a prescriptive model enables the development of a road-map for improvement, in order to increase the business value, comparing similar practices across the organization identifying maturity levels within disparate industries. While for comparative models, it is necessary to get a sufficient amount of data to allow a valid comparison. Thanks to the development framework proposed, initially it is possible to guide the evolution of the model from a descriptive phase to a prescriptive and comparative phase. In the *Figure 31* the generic phases are presented.



*Figure 31 Model Development Phases (De Bruin et al. 2005)*

- *Phase 1 – Scope*

This phase will influence the following phases. The definition of the scope determines also boundaries for the model application and the main focus of the model, thus the establishment of the referring domain to which the model would be targeted and applied. The specific spotlight on a domain influences the model specificity and extensibility.

- *Phase 2 – Design*

In this step the needs of the reference audience were taken into consideration, looking for a proper balance between the complexity of the reality and the model simplicity. It is important to define *why* they want to apply the model, *how* the model can be applied in reality according to the different organizational structures, *who* should be involved in the model application and *what* can be achieved in applying the model. The number of stages could vary according to the different model and how they are reported to the audience.

- *Phase 3 – Populate*

At this point it is fundamental to identify domain components and possible measurements. In case of mature domain, the relative components could be identified through an exhaustive literature review. If the domain is characterized by complexity, the identification of related component could be difficult. For this reason, the main aim is to attain components that are mutually exclusive and collectively exhaustive, also looking for sub-components, supporting the development of assessment questions. This is possible resorting to exploratory research methods such as Delphi technique, Nominal group technique, case study interviews and focus groups. While in a relatively new domain, the existing literature review could represent only a starting point, not providing sufficient evidence, it will be necessary to proceed with other means of identification. The key point when populating the model is the selection of the combination of the proper research methods for the respective development.

Therefore, the maturity measurements have to be determined. The application of the maturity model can be supported by instrument such as a survey - quantitative method, enabling statistical analysis and improving results comparison, it is advisable.

- *Phase 4 – Test*

This is the moment to test the relevance and rigor of the model, together with the completeness and accuracy of the same, compared with the previously identified scope. Furthermore, it is necessary to evaluate also the assessment instruments for validity, checking they are measuring the right object and the reliability, to be sure that the results obtained are accurate and repeatable.

- *Phase 5 – Deploy*

To deepen the critical issue of model generalizability, first there is an initial organizational application where design collaborators are considered as primary

respondents. This can lead to acceptance of the model. Therefore, it has to be applied within entities independent from the model development. These entities were chosen in accordance with the original scope of the model application.

- *Phase 6 – Maintain*

The resources necessary to maintain the model are dependent on the maturity model goal. The evolution of the model is required in case the specific knowledge of the domain broadens.

Considering maturity models as artefacts, aiming at solving problems of determining the company's current status regarding their capabilities, it is possible to assume that the maturity model development is close to the application area for the design science guidelines defined by Hevner et al. (2004). Becker et al. (2009) translated these guidelines postulating eight requirements for the development of maturity models

- Comparison with existing maturity models: to motivate the development of a new maturity model, it is necessary a comparison with the already existent ones, since the new model could be an improved version of pre-existing one;
- Iterative procedure: the development of maturity models is step by step;
- Evaluation: iterative evaluation of usefulness, quality and effectiveness of the artefact;
- Multi-methodological Procedure: a variety of research methods is deployed in the development;
- Identification of Problem Relevance: to demonstrate the problem importance for researchers and/or practitioners;
- Problem Definition: to exactly define the problem, the prospective application domain and the relative conditions;
- Targeted Presentation of Results: the focus in the presentation of the maturity model must be on the specific needs of the users;
- Scientific documentation: each step of the design process, the parties involved, the methods applied, and the consequent results should be documented in detail.

Against this backdrop, also Becker et al. (2009) introduced a procedure model for the development of maturity model (see flow chart in *Figure 32*). The elements of this model find their roots in the requirements above identified. In particular, the starting point of the procedure model is the problem definition.

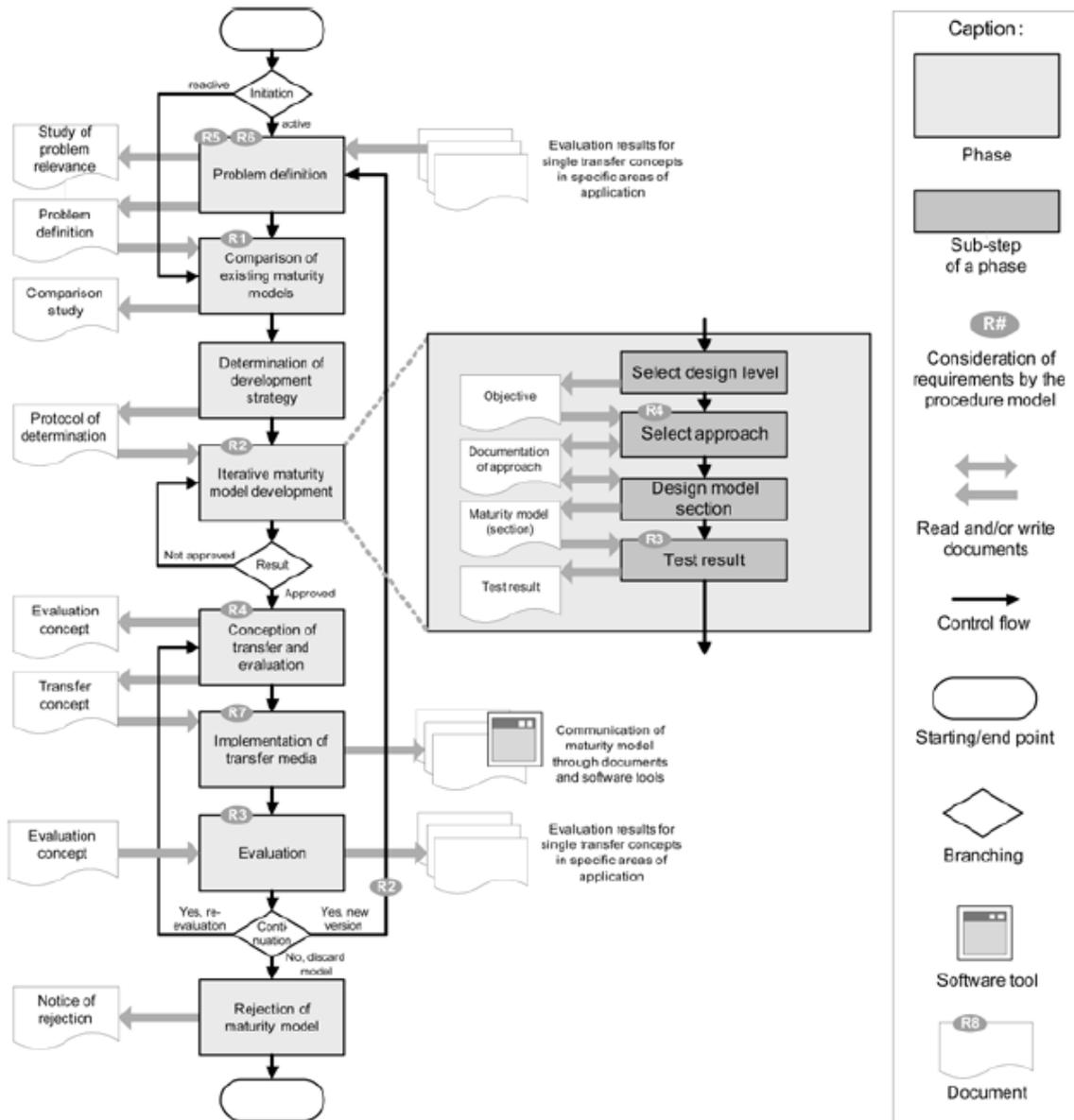


Figure 32 Procedure model for developing maturity models (Becker et al., 2009)

## 3.9 Performance measurement in manufacturing

The companies' adoption of a formalized Performance Measurements System within their IS is driven by multiple reasons; specifically, according to Sianesi (2014) three main objectives could be pinpointed.

*First objective.* A performance measurement system is a supportive tool for managers during decision making processes, enhancing their quality and timeliness. Information must be collected and presented to the users systematically in order to properly support managers who should obtain signal related to company results, current status, environmental variables and customers' requirements. A proper set of indicators supports future decisions also learning from company history performance measurements. In fact, analysing and finding causal relationship between operative conditions and performances, the decision maker is facilitated when a certain phenomenon recurs.

*Second objective.* A performance measurement system that provides a wide set of information, makes possible to trigger the motivation process of a company's resources. Indicators are the foundations of improvements processes, indeed according to both Deming and Drucker "you can't manage what you don't measure". The absence of metrics makes subjective the improvements' assessments, while a structured performance measurements system enables to isolate the contribution of each department/resources in order to detect where causes of poor performance are located. Therefore, metrics could be seen also as a supportive tool to motivate and incentivise employees within companies and organizations.

*Third objective.* The last aim of a performance measurements system is to analyse two distinct aspects: the prospect of results and the prospect of trend. Prospect of result identifies metrics that are able to understand if an organization owns the skills to manage the critical variables, which are the foundations of the competitive advantage. While, prospect of trend focuses its attention on the future management condition required for the performances' maintenance and improvements.

According to the author, performances of a production system could be broken down two classes (Figure 33): internal performances (efficiency) and external performances (effectiveness).

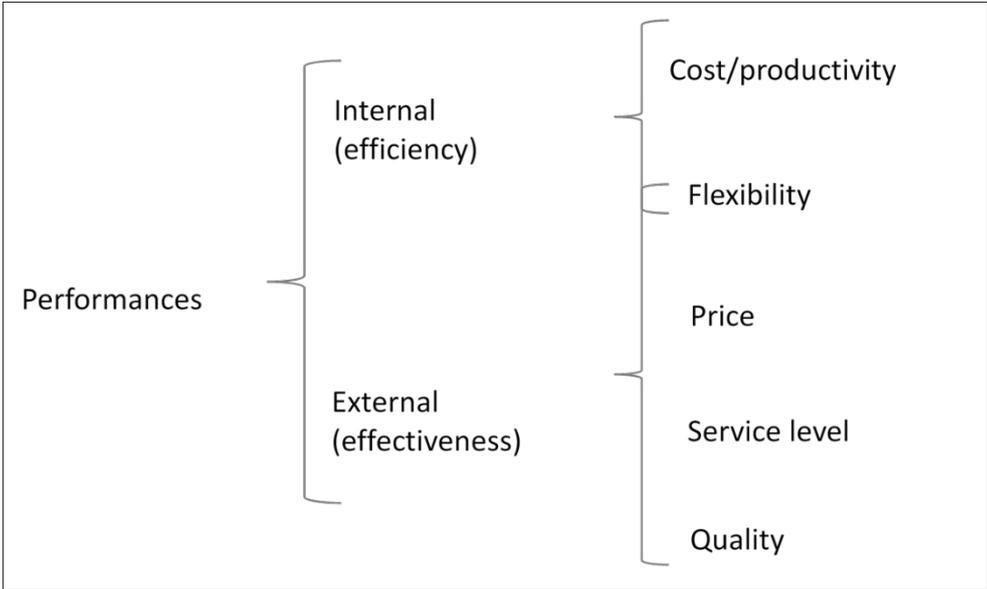


Figure 33 Performances tree (Sianesi 2014)

*Internal Performances.* The first category measures the efficiency of production factors, determined by the operating condition of the production system such as the machineries’ productivity capacity, resources’ reliability and workforce’s characteristics. These metrics are not directly connected with final customers, but they are still essential for the economic surveillance of companies. The efficiency metrics measure the resources consumption required to achieve a certain service level defined by the final costumers. Usually these metrics are expressed by the ratio between input and output. It is possible to distinguish among productivity and flexibility related to the cost performances.

*External Performances.* The second category, measurable by external stakeholders, aims to measure the degree of customers’ satisfaction compared to the company objectives. Normally these metrics are expressed by the ratio between successful events and total events, where an event could be an order, a delivered product and a successful event might

be the order's delivery in compliance with the customer's requirements. It is possible to distinguish among quality, service level and flexibility related to time performances.

According to Brandolese (1990), flexibility could not be defined *ex ante* as an efficiency or an effectiveness performance. Indeed, on one side it analyses the costs (internal performance) required by a company to respond to possible variation such as the introduction of a new product, increase or decrease the production volume and changes in the delivery plan. On the other side flexibility could be seen as an external performance measuring the timeliness whereby a company reacts to the same potential changes.

Both external and internal performances often are presented with the term Key performance indicators (KPI), which are a set of quantitative indicators applied to monitor companies' results. The definition of indicators capable to monitor the activity and process performances, enables to present uniquely and clearly the company ability to achieve its goals.

### 3.9.1 Manufacturing Productivity

Productivity is the most deployed efficiency measure in case of logistic-productive systems being applicable to multiple production factors (e.g. materials, workforces, capital, energy). It enables managers to understand the degree of efficiency whereby the different inputs are deployed. Broadly speaking, the productivity of a production factor is defined as a ratio between the output volume and the volume of inputs. For example, the workforce productivity could be defined as the ratio between the volume produced and the number of workers. The application of these kind of indicators, even if it does not provide any explanation why a loss happens, is extremely easy and provides a synthetic judgement of the system under observation.

In literature, the setting of a single metrics for productivity is not presented, although all the authors agree with general definition. The Australian Bureau of Statistics (ABS) states: "In a very general sense, the best way to think about productivity is by thinking of production. You can have increased production from an increase in inputs, you can have increased

production due to a more efficient use of those inputs or a combination of both of those things. In a growth accounting framework, you can in simple terms measure productivity by looking at the ratio of output to one or more inputs. When you decompose it, in a sense, productivity is actually the residual of that calculation” (Australian Bureau of Statistics 2009).

For the further analysis, the manufacturing productivity must be broken on two indicators: utilization rate and efficiency. The utilization rate defines “how much” a specific resource is utilised, for example the machinery utilization rate might be seen as the ratio between the plant opening hours and the effective production time. Instead the efficiency defines “how” a specific resource works; for instance, for the machinery is the relationship between the production effectively detected and stored and the effective productive hours, which is a proxy of the real production time (Sianesi 2014).

According to Huang et al. (2003) these traditional metrics present the following drawbacks:

- A comprehensive evaluation of the manufacturing equipment performances is not provided
- The application of these metrics does not make possible the identification of the problems and the possible area of improvement

The constant struggle of improvement and optimization of company productivity is mainly driven by the intense global competition. This led to the need for a more rigorous method to trace productivity within the manufacturing companies. Nakajima (1975) is the first author that launch the concept of Total productive maintenance introducing a completely new metrics called OEE – Overall Equipment effectiveness.

## OEE

The OEE represents a global standard to measure the production unit efficiency, even if its definition refers to the word *effectiveness*. It is an aggregated metrics that could be applied at three different levels in a manufacturing environment (Bulent, Tugwell e Greatbanks 2000):

- to measure the global performances of a plant;

- to compare the OEE value of different lines within the same factory to detect where are located the poor line performances;
- to identify which is the worst machine performance in case the machine process works individually.

This concept has been widely applied as a quantitative metrics, at first in the semiconductor manufacturing operations (Huang, et al. 2003). Currently, the OEE is a standard measure employed throughout companies and sectors, particularly into the ones involved in World Class Manufacturing or Lean Manufacturing projects.

OEE is a function of three mutually exclusive components, availability, performance and quality, enabling companies to examine and analyse each factor separately. The foundation of this metrics is to detect which are the chronic and sporadic disturbances impacting negatively the manufacturing process. Nakajima (1988) identifies “six big losses” which consume resources without adding any value to the final product: breakdown losses/equipment failure, set-up and adjustment, minor/idling stoppage, reduced speed losses, defect/rework in process losses and start-up/reduced yield. These losses could be grouped into three macro classes: downtime losses, performance losses and quality losses.

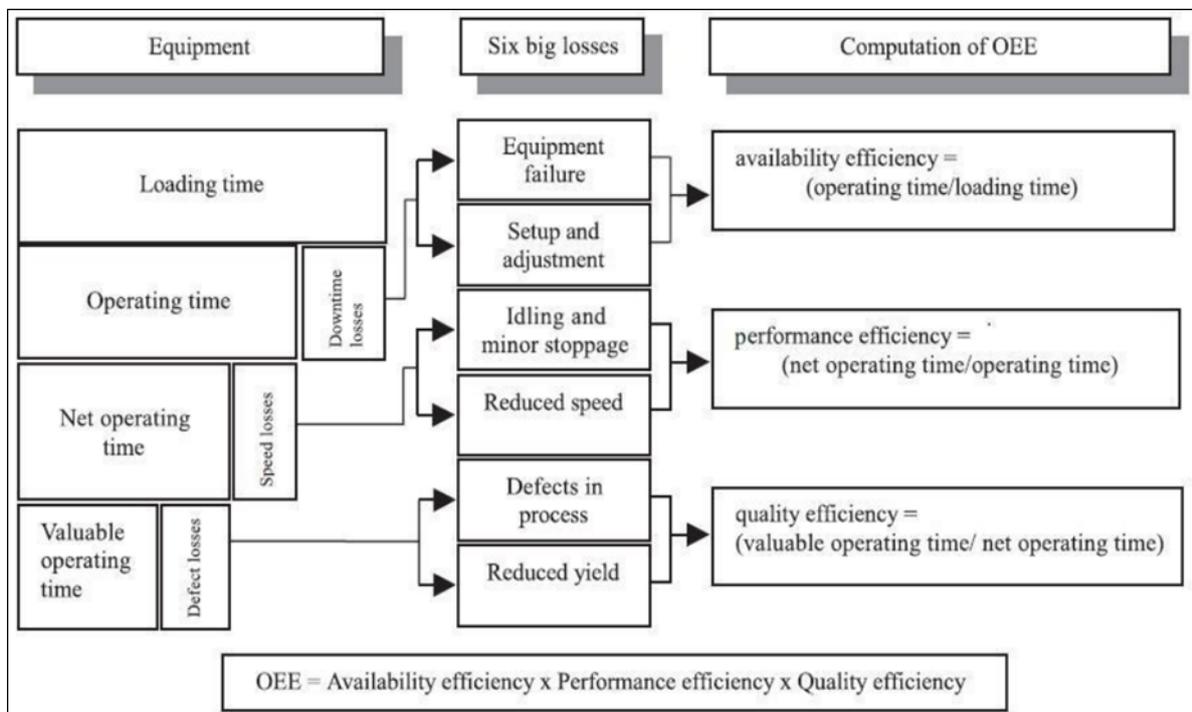


Figure 34 OEE (Nalajima 1988)

*Downtime losses* are distinguished among the following losses:

- *Breakdown losses* are due to equipment failures, breakdowns and other unplanned downtime that could occur within a manufacturing process. These losses could be classified on one side as a quantity loss due to defective production, on the other side as a time loss occurred when the machinery productivity is decreased. Moreover, unplanned downtime might be broken down into the following elements: Mechanical downtime, Electrical downtime, Electronic downtime, Operator shutdown, Downstream operation downtime, Upstream operation downtime, External to plant downtime, Planned downtime overruns (Lichtenberg, 1996).
- *Set-up and adjustment losses* are the result of downtime required when production of one item ends and the facilities need to be well-equipped for another item. Set-up losses include also the warm-up time and the run-in time, required to ensure high-quality production. Finally, also the time related to production part changes needs to be taken into account.

*Performance losses* determine the performance efficiency of a specific machine and are related to the real production operating condition.

- *Idling and minor stoppage losses* are due to production interruption for transitory incorrect operations and when a machine is idling. These losses could significantly impact the company productivity although their causes are easily identifiable and fixable. The minor stoppage losses are often known as “machine hiccups” since apparently are not recorded and tracked by companies.
- *Reduced speed losses* occur when exists a distinction between the machinery design speed and the actual operating speed. In fact, in reality machinery could run with a lower speed than the designed one: this happen mainly when a manufacturer is operating with obsolescent machinery and components or when intentionally machineries are run with lower speed by employees to avoid problem and reduce breakdowns (Smith & Hawkins 2004).

*Quality losses* must be managed carefully since “*If equipment is available 24 hours a day, 7 days a week ... and if it’s performing at its highest design cycle rate ... then if it’s not producing the highest level of quality, it’s just producing scrap at full capacity*”. These last two losses are related to the number of defects with the productions.

- *Defect in process losses* are due to malfunctioning production equipment, where the production time is deployed to manufacture unusable items.
- *Reduced yield losses* occur specially when the production starts, and the machinery requires a certain amount of time to get stable and to produce quality pieces.

Indeed, Nakajima (1988) defined the OEE as “a bottom-up approach where an integrated workforce strives to achieve overall equipment effectiveness by eliminating the six big losses”. Within its work the author proposes ideal value for each component: availability 90%, performance 95% and quality 99%.

Each element of the OEE is formulated as follows:

The *Availability* element is defined as the ratio between the actual production time and the potential production time. The actual production time is connected to the total stoppage time coming from unscheduled breakdowns, waiting time, changeovers, no supply or transport and in general all the unplanned stoppages. Instead the planned stoppages such as waiting time required to complete orders, no workforce available due to breaks, planned maintenance and cleaning activities, are already included within the potential production time (Dal, Tugwell, & Greatbanks 2000). Summing up the availability factor analysed the theoretical production time counter to unplanned breakdown and stoppages event.

<b>AVAILABILITY</b>	<b>A</b>	<b>Potential production time</b>	
	<b>B</b>	<b>Actual production time</b>	Availability losses: - Breakdowns - Waiting/changeover - No supply or transport

Figure 35 OEE - Availability (OEE Foundation website)

The *Performance rate* of the OEE could be computed in a variety of ways. According to Nakajima (1988) this factor is defined from the comparison between the production time and the ideal cycle time required to produce a fixed amount of output. Instead according to De Groote (1995) the performance rate is computed comparing the ideal and real production achievable from a defined fixed time.

Whatever else the measure applied, the performance rate detects the percentage of the actual output (real pieces made by a machine). Formally it could be presented as follows (Dal, Tugwell, & Greatbanks 2000):

Performance efficiency % = (Net operating rate \* Operating speed rate) \* 100; where:

- Net operating time = (No. produced \* Actual cycle time)/Operation time
- Operating speed rate = Theoretical cycle time/ Actual cycle time

Mainly the performance losses are the results of minor stoppages for obstructions, feeding errors and machinery running with slower speed than the potential one.

<b>PERFORMANCE</b>	<b>C</b>	<b>Theoretical output</b>	<i>Availability loss</i>
	<b>D</b>	<b>Actual output</b>	

Figure 36 OEE - Performance rate (OEE Foundation website)

The third factor, the *Quality rate* (Figure 37) is deployed to detect which is the portion of defective production on the total one. It is important to point out that this rate implies the number of defects related only to a specific stage of production (Dal, Tugwell, & Greatbanks 2000). The quality losses are related mostly to two factors: the scrap produced during the production and the related rework required within a process.

According to Dal et al. (2000) some issues arise from the OEE definition. Firstly, multiple opinions concerning the acceptable level of availability, performance and quality are provided in literature, highlighting the differences among the various fields of application.

Secondly, its deployment required a high data sources credibility, in order to be conceivable and believable to production management. In fact, a key concern of the OEE computation is the high amount of detailed performance data required, compared to the other performance measurement metrics. The simplification of the required data should be the main goal of this metrics since “the most important objective of OEE is not to get an optimum measure, but to get a simpler measure that indicates the areas for improvement” (Ljungberg 1998). It provides a perspective on production improvement, but it should be applied in comparison with the traditional metrics in order to obtain a more comprehensive overview of the manufacturing process.

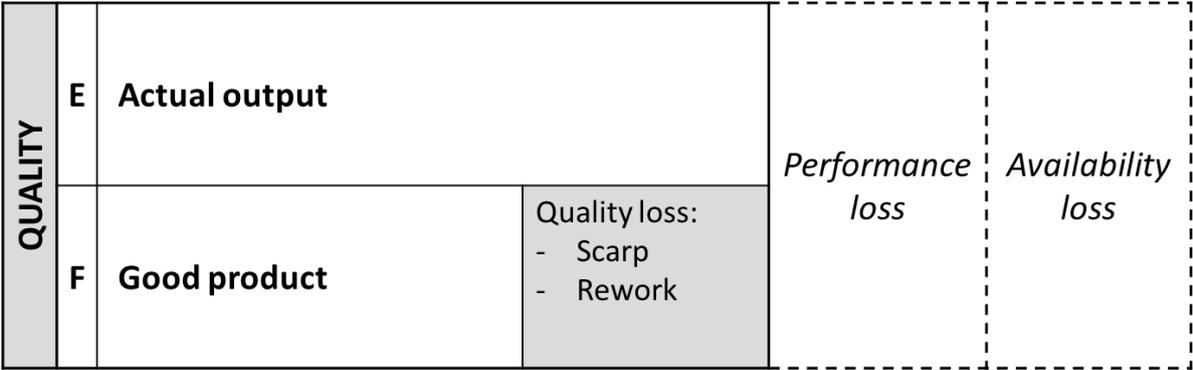


Figure 37 OEE - Quality (OEE Foundation website)

Although the emphasised issues, the OEE metrics is a well know approach currently applied within different companies and sectors to monitor and stressed possible improvement within the machinery and operational resources performances.

## 3.10 Decision-making process

Decision-making process is an integral part of all managerial activities, affecting companies' performances and in general their success or failure (Kaner & Karni 2004). These processes are always becoming faster and more complex due to the global competition and the technology innovation. In fact, decision makers must undertake decisions under a higher time pressure in order to ensure the company success, boosting competition, rapidly reacting to competitors' action and new market conditions (Spina 2012).

Previously within companies, decisions were taken mainly relying on the employees' knowledge and intuition, without following a structured and well-defined process. Nowadays, the context where companies are operating is changing: the decision-making processes are becoming more complex and dynamic, requiring a more rigorous method and approach, such as analytical methodologies and mathematical models, to deal and cope with them (Vercellis 2009).

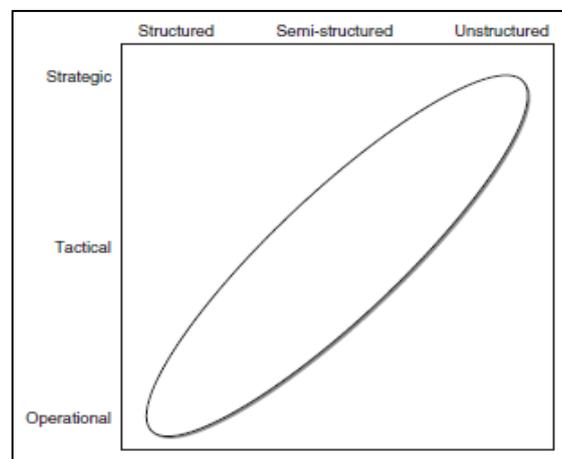


Figure 38 Decisions taxonomy

Simon (1960) in his work distinguishes between “programmed decisions” defined as repetitive and routine activities whereby a predefined procedure exists to deal with them; and “non-programmed” decisions identified as new and unstructured activities, not driven by specific procedure, that should be solved relying on decision makers ability and intelligent. Subsequently other authors elaborate this classification distinguishing between

three different type of decisions: structured, semi structured and unstructured (Gorry & Scott Morton 1971).

*Structured decision is a “decision based on a well-defined and recurring decision-making procedure.”* Structured decisions are those ones suitable for automation since they could be modelled by an algorithm. Most of the case even when decisions makers are coping with a structured decision their intervention is required to cope with unexpected values or events. *Unstructured decision is a “decision that cannot be depicted through a predefined sequence of steps or that is composed by an element (input, output, transformation process) that cannot be detailed described”.* Unstructured decisions occur typically with unexpected events, where decision maker needs to cope the situation without a predefine structure and process. To timely support this kind of decisions the knowledge of the decision maker is essential. *Semi-structured decision is “a decision composed by some structured phases and other unstructured ones”.* This type of decision is very common within companies. Decision makers can take advantage of DSSs by having access to information to better face the unstructured part and using algorithms to automate the structured part of the decision-making process.

Decisions could be classified looking at their scope identifying strategic, tactical and operational decisions (Kaner & Karni, 2004). *“Strategic decision is a decision which has effects on the entire or an important part of the organization for a long period of time”.* These decisions, characterized by a long-time horizon, fall under the responsibility of the company’s top management experiencing a great impact on company’s objectives and policies. *“Tactical decision is a decision which produces an impact on only part of an enterprise and is usually restricted to a single department”.* The time horizon characterizing these decisions is a medium-term horizon, usually up to a year. Tactical decisions are usually taken by middle managers, such as the heads of the company departments. *Operational decisions refer to ““handles the day-to-day activities of producing services or products and works directly with customers and suppliers”.* These decisions are usually taken by lower organizational level, responsible for a single task such as sub-department heads, workshop foremen, back-office heads.

Decisions within companies are taken at a certain level of risk and uncertainty. Generally, these two factors are reasonably reduced at the operational level where the nature of decisions are predominantly programmed. While at higher managerial level, decisions mainly being unprogrammed ones, are mostly taken under high level of risk and uncertainty. For example, decisions for restricted operational tasks are programmed and low-risk, the ones dealing with middle managers are mainly characterized by low-medium risk and finally the ones taken by top-manager are mostly affected by a high risk and uncertainty (Spina 2012).

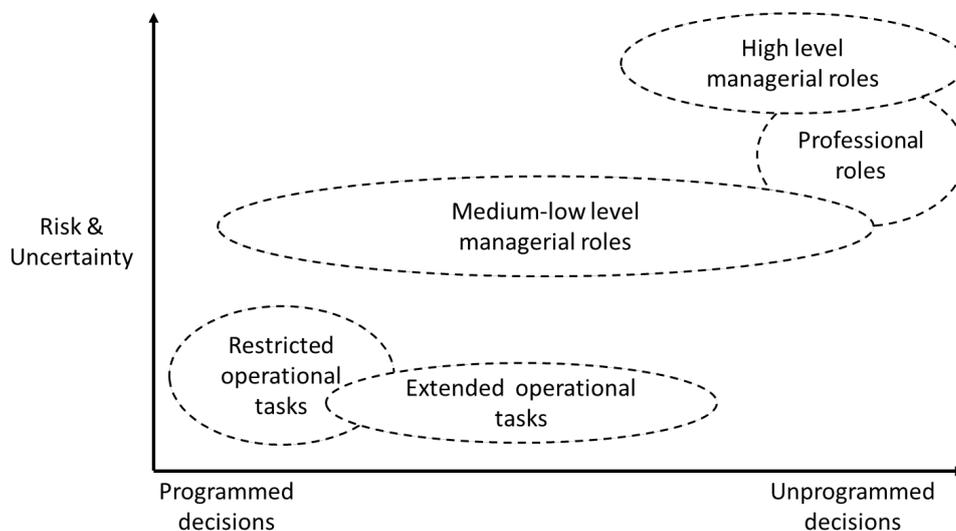


Figure 39 Organizational roles and decisions (Spina 2012)

The concept of *limited rationality* presented by Herbert (1947) is a key element in the decision-making process field. It stresses that decision makers do not undertake decisions under optimal circumstances since in reality within companies the objectives are not well-defined, the entirety of information required is not available, or a considerable period of time is required to get the information and resources needed. Therefore, it is not taken the optimal decision defined by an objective function and analyzing all the possible alternatives, but the most satisfactory and favorable solution is chosen.

The decision-making process is composed by two main subsequent step the problem identification and its solution. The first step is termed *Problem setting* including the problem perception, objective definition and the analysis of the relevant factors for the alternative

selection. The second step, known as *Problem solving*, involves the alternative identification and the assessment, analyzing their capability to achieve the objectives, the decision maker choice and the performance monitoring. In reality, the distinction between the problem setting and the problem solving is not that clear and defined since decision makers while are analyzing a problem are already assessing its solutions and alternatives (Spina 2012).

A managerial decision is an organization process which involves analysis, interpretation assessment, intuition and action which lead managers to select an alternative following a structured process. Three different approaches could be identified resuming the multiple situation that could be faced by a decision maker (Mintzberg 1990):

- *Thinking first* is an analytical approach, according to which problem setting play a crucial role; the problem is deeply analyzed looking for all the possible effect and consequences of each alternatives identified.
- *Doing first* is a more practical approach, whereby the focus is on the problem solving; the solution is identified experimenting with the different alternatives.
- *Seeing first* focus its attention on problem setting, a solution is identified from the beginning and attempts are made to implement it.

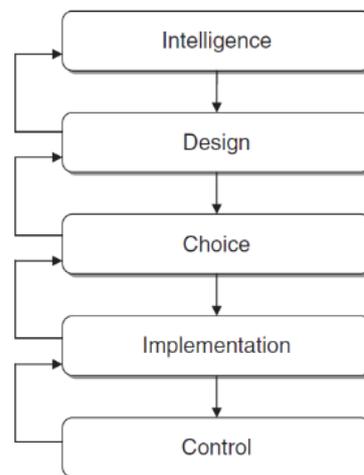


Figure 40 Phases of the decision-making process (Vercellis, 2009)

According to multiple authors (Vercellis 2009; Spina 2014) the decision-making process might be described by the following steps: intelligence, design, choice, implementation, review/control.

1. *Intelligence*: firstly, there is the perception of the problem and the possible threats and opportunities that could arise. Different individual and department within the same company could perceive the problem in different and opposite way. In this step the decision maker identifies the objective, the possible constraints, and collect all the information assessed to be useful to cope with the problem identified.
2. *Design*: once defined objectives and constraints, a reference system must be designed building a model of reality on the basis of which decision will be taken. The identification of a reference model is a preliminary step, required to generate the different alternatives. This generation is a sequential process which aims to identifies a set of alternatives meeting the requirements. The development of the alternatives is affected by the concept of limited rationality, that limits the generation of a perfect and exhaustive list of alternatives. The factors affecting the alternatives' generation and assessment are the actors' different perception of the problem, multiple skills, impossibility to obtain complete and comprehensive information, and the consequent limited analytical capacity of the decision makers. Summing up, this second step focuses on the alternatives' generation and assessment.
3. *Choice*: once the design face is done and the alternatives have been assessed there is the final choice. The alternatives which best satisfy the objectives and respect the requirements should be selected. In real decision-making process, multiple objectives in trade off might be present, making harder the identification of the best alternative. The decision maker selection is subjective, based on which parameter he considers the most appropriate.
4. *Implementation*: the identified decisions are implemented planning the required actions through the definition of activities, timelines, means, roles and responsibility of the individual involved. Subsequently all the planning activities are effectively realized, balancing and coordinating decision makers with the executors.
5. *Review/control*: finally, the result must be monitored and controlled through indicators defined into the intelligence step. This final phase could reveal slippages between the results and the initial objective, leading to the generation of new threats and opportunities and triggering a new decision-making process.

### 3.10.1 Decision-making process within supply chain

To identify which are the main decision-making process within companies the supply chain operation references (SCOR) model could be implemented (Lockamy & McCormack, 2004). The SCOR model was evolved by the Supply-Chain Council (SCC) to assist companies in increasing the performance within their supply chains, providing a process-based approach. Its aim is to describe and analyze the logistic and informative flow, relationships among the focal company, customers and suppliers. This to improve the alignment between the marketplace and the strategic response of a supply chain. The model facilitates communication, providing a standard format useful to reach the advocated performances, designing and reconfiguring the supply chain of a company (Huan, Sheoran & Wang 2004).

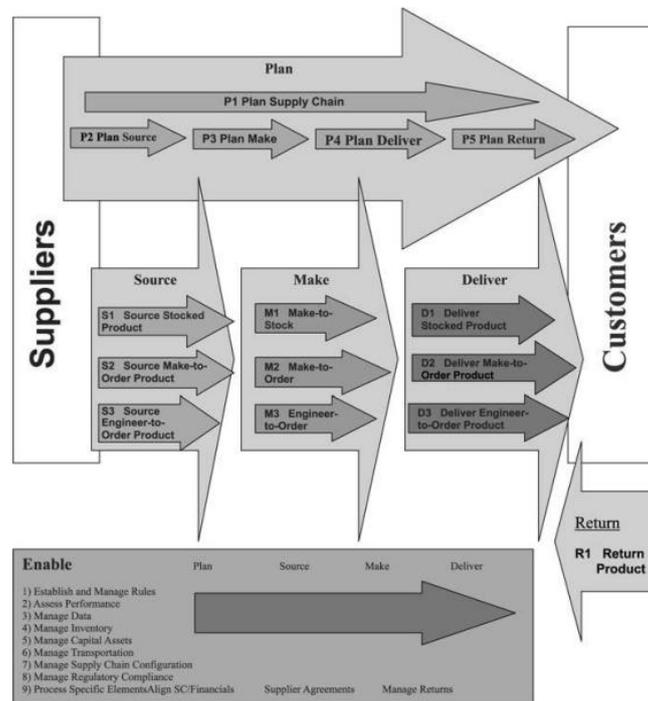


Figure 41 SCOR version 4.0 (Supply-Chain Council 2000)

The model has a hierarchal structure broken down into three level (Figure 41): moving from the level one to the subsequent ones a higher level of detail is achieved, describing systematically all the business activities. The *first level* defines five management process/decision areas: plan, source, make, delivery and return (Sianesi 2014).

*Plan.* This process objective is to balance the demand with the available resources; it develops the actions lines designed to manage the acquisition of the resources need, and to optimize the purchasing, production and distribution activities.

*Source.* This process involves all the execution activities of goods and services procurement, required to satisfy the demand. The main activities are suppliers' identification and selection, negotiation, purchasing, materials, product and service retrieving (Romano & Danese 2010).

*Make.* This process consists on components and materials transformation into finished products to satisfy the current or predicted demand. The main activities involved are production monitoring, warehousing, quality control, packaging and distribution phases.

*Deliver.* This process concerns the delivery of finished goods in order to meet demand; it typically involves the order, transportation and distribution management.

*Return.* This process, also known as reverse logistics, describes all the activities related to returned goods due to failure/malfunction. Therefore, it contains the following process: authorization, post-sales, the shipment of replacement parts or substitute products.

At the *second level*, each process of the first level could be further described through three categories' process (Sianesi 2014):

- *Decision Planning processes* are defined as all those processes require to plan the activities to align available or potential resources with the future predicted demand. In this context decisions related to performance objectives, information management, stock and warehouse objectives, investment choice, supply chain risk management are made.
- *Execution processes* belong to operational management, activated by the planning processes and demand. The activities involved are material transformation, goods movements and scheduling.
- *Enable processes* are those ones empowering both execution and planning processes, in order to prepare, maintain and manage all the information required by the planning and the execution processes.

According to Lockamy & McCormack (2004) the planning process have a significant importance in all SCOR supply chain planning decision areas. The author identified seven supply chain management planning decision categories related to the SCOR model which are: operations strategy planning, demand management, production planning and scheduling, procurement, promise delivery, balancing change and distribution management.

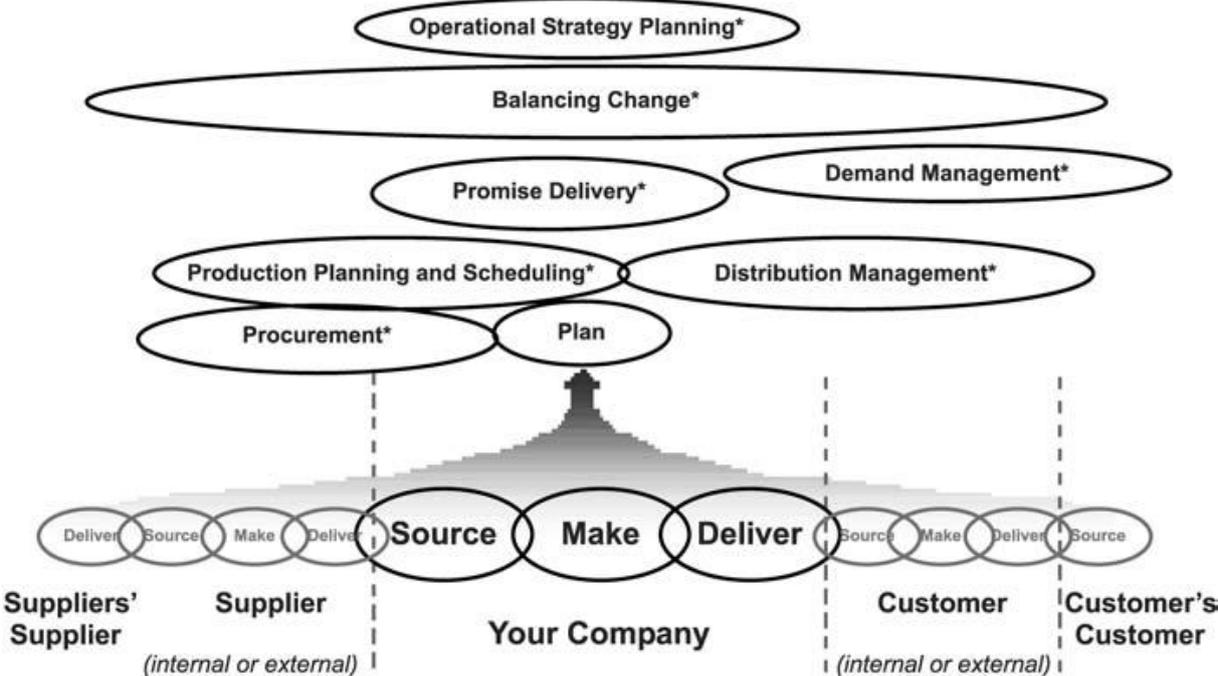


Figure 42 Supply chain decision categories mapped to the SCOR model ()

As shown in the Figure 42: promise delivery and operational strategy planning decisions should be aligned with the internal decision area; balancing change must be aligned across the entire supply chain; the procurement and production planning decision should be lined up both across internal and suppliers decisions area; finally the demand and the distribution management decisions should be spanned across both internal and customer decision areas.

# 4 Objectives

## 4.1 Literature gaps

To clearly represent in a synthetic way the main themes discussed in the literature, it is useful to resort to a Cartesian chart, providing a structured representation of all themes object of the research. Each theme is presented along with the detail of the number of papers considered.

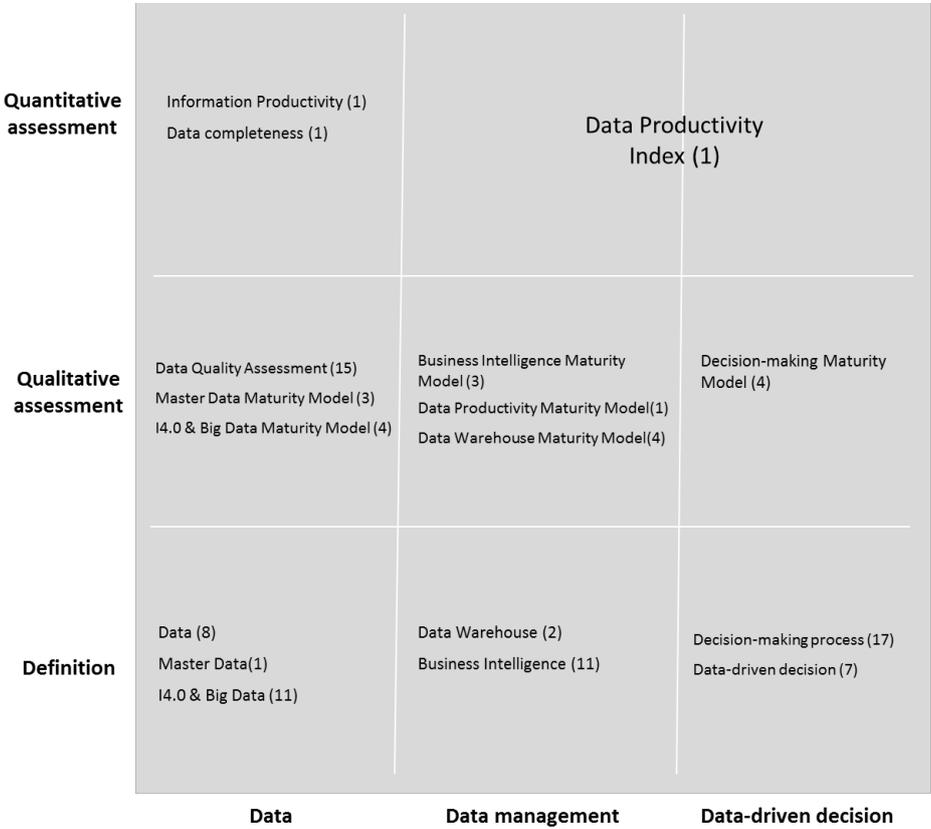


Figure 43 Papers' chart

The chart proposed is composed by two axes. The horizontal axis describes the objects of the researches, thus the topics of Data, Data Management and Data-driven decision. While the vertical one indicates the assessment methods explored, starting from the simple definition, proceeding with the qualitative assessment then to come to the definition of a quantitative measure. Moreover, in the *Figure 43* each intersection between the two axes

represent a specific research theme, the so-called building blocks of literature review. Their main aim was to detect the unavailability of sound and suitable literature studies and researches on a specific topic.

The review of the extent knowledge allowed to highlight two gaps in the literature, mainly concerning the lack of academic studies related to the possibility to identify a quantitative measure of data management and data-driven decision. For the first time, these topics were covered by the Miragliotta et al. (2018) work. Considered the relevance of these topics for academia and as well as for the industry, and also the great potential of the thinking behind, it was decided to focus on this work, taking it as starting point and to provide a possible development. To do this, it was thought to start in particular from the potential limitations identifiable.

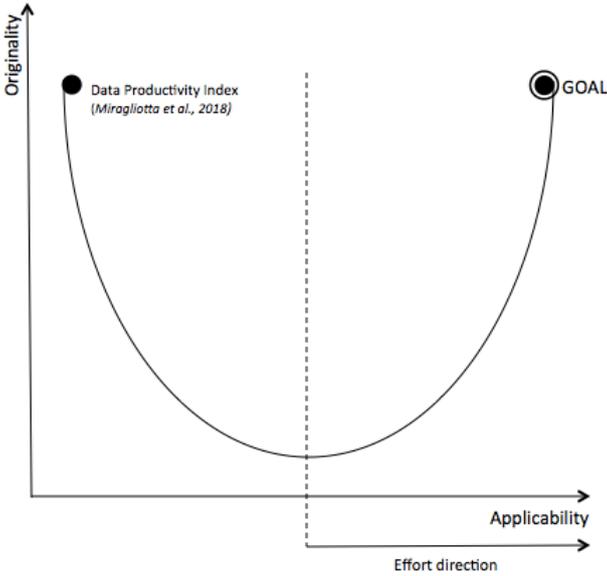


Figure 44 Thinking behind our research objectives

The previous work is largely marked by a great originality at the expense of the business applicability. This is a completely new approach, considered the purpose to not focusing only on data, but also on the transformation into information and knowledge. Most researchers restricted their work on data, without facing the rising complexity encountered with information and knowledge. Indeed, the implementation of this measurement tool in the organisations was characterised by high complexity, requiring significant investment in terms of time and resources. Indeed, to make the data productivity computation a detailed

list of data must be defined each time for the specific decision-making process under analysis. Moreover, the interviewee could adapt the input in case it is not considered comprehensive, eliminating or adding new data.

The proposal of this thesis work is to overcome this main issue, trying to figure out a new compromise. A compromise based on an enhanced applicability, without sacrificing the starting originality. Thus, the effort to solve this issue must be made moving towards a greater applicability in the business context, which means making the measurement least onerous to conceptualize, to understand and to transmit to companies. Furthermore, to reach that goal, it is essential to focus especially on the standardization, with the purpose to make this index comparable between different companies working in different sectors and reality.

## 4.2 Research Questions

To sum up, the output of this study is to provide a measurement tool able to quantify in a synthetic way the exploitation of data in organisations decision-making process, starting from the limitations of the previous work on the topic. Particularly this index is intended to sustain businesses in the optimization of data usage with the aim to support decision-making and problem solving. Therefore, the objectives of this dissertation are of both an academic and a practical nature.

### 4.2.1 Academic Objectives

#### RQ1: Which are the most appropriate features to assess data productivity?

The first limitation identifiable in the Data Productivity model developed by Miragliotta et al. (2018) consists in the definition of the data assessment dimensions. The overall three factors - Data Availability, Data Quality, DSS Performance - presented in model and inspired by the OEE structure, were not considered sufficiently comprehensive. Thus, their relative sub-dimensions were defined without taking exhaustively into account the rich body of literature already existent on the topic. However, the factors proposed could be seen as a good starting point.

#### RQ2: Which is the most suitable approach to evaluate in a structured way the data productivity dimensions?

After the identification and definition of a reference theoretical framework for the data assessment, the next objective is to find a more applicable way to evaluate the dimensions identified. Moreover, the assessment of each dimension has to become more objective. Indeed, the evaluation of each sub-dimension in the previous work was too much depending on the specific user perspective. Therefore, the need of an evaluation approach capable to determine each data feature in a more understandable, structured and guided way, ensuring to overcome the subjectivity issue. Moreover, it is necessary to find an approach which is suitable for the achievement of an overall numeric index computation.

## 4.2.2 Practical Objectives

RQ3: How is it possible to enhance the model applicability in the business context?

Two are main limits identified to the standardization. The first one concerns the input of the model: the list of data defined as object for the evaluation. This list is too detailed, it varies from one decision to another and each interviewee can modify and add to the list elements depending on the specific circumstances. This is the first constraint making impossible the universal comparability of the index. The second limitation is the subjectivity of the data assessment. Different evaluations of the single data could be provided by people belonging to the same organisation, since the evaluation of the specific data taken into consideration depends on the decision-making process under consideration and on the specific user's perspective. Moreover, in the Data Productivity model the evaluation given is highly subjective, since the score assigned has not been accompanied by the relative description, but rather depends to the interviewer yardstick. Indeed, the value obtained massively depends on the application context and it can vary a lot, especially between different users. Therefore, the precautions taken to mitigate the subjectivity of the approach are considered not enough. A basic method should be developed within this framework with which quality measurement can be objectively designed. It is necessary to strengthen the approach.

# 5 Research Methodology

The model development was done in accordance with guidelines and frameworks from academia, with the purpose to maintain the work consistent with the academic requirements. The design and development methodology used globally follows the development phases of the framework proposed by García-Mireles et al. (2012) to guarantee the validity for an academic design science research.

This methodology was integrated resorting also to the design science approach described by Hevner et al. (2004) and revised by Becker et al. (2009). Becker in his work, related to maturity models development procedure for IT management, relied on the guidelines defined by Hevner et al. (2004) to build the maturity model requirements. This was made possible considering the maturity model as an artifact whose objective is to define the current status of organisation's capabilities and means for improvements.

Paper	Inception	Elaboration	Construction	Deployment	Maintenance
Hevner et al. (2004)	Guideline 1: Design as an Artifact	Guideline 6: Design as a Search Process	Guideline 6: Design as a Search Process	Guideline 7: Communication of Research	-
	Guideline 4: Research Contributions	Guideline 5: Research Rigor	Guideline 5: Research Rigor	Guideline 6: Design ad a Search Process	
	Guideline 2: Problem Relevance			Guideline 3: Design Evaluation	
Becker et al. (2009)	R6 (Problem Definition)	R4 (Multi-methodological Procedure)	R4 (Multi-methodological Procedure)	R7 (Target Presentation of Results)	-
	R5 (Identification of Problem Relevance)	R2 (Iterative Procedure)	R2 (Iterative Procedure)	R8 (Scientific Documentation)	
	R1 (Comparison with existing maturity models)			R3 (Evaluation)	

*Table 10 Methodologies followed*

According to the methodology developed by García-Mireles et al. (2012), the procedure to develop scientifically valid models consists of 5 main activities: inception, elaboration,

construction, deployment and maintenance. These phases are going to be better described following the two above-mentioned methodologies (*Table 10*).

### First phase: Inception

In this phase the main decisions to take are the identification of the problem, the participants and the final model target. But first of all, taking as reference Hevner's 'Guideline 1: Design as an artifact' and 'Guideline 4: Research Contributions', it is necessary to conduct previously a literature study. The literature study was conducted principally to find the presence of existing researches on the topic intended to investigate. Then a structured research on close themes was carried out, with the goal of building a model to measure the productivity of data in the decision-making process. Therefore, the study was directed to the possible methodological assessment approaches, focusing in particular on the maturity one. An important point to help the development of the model was considered the comparison with existing maturity models and the attainment of a general overview. All the maturity models used as a source of inspiration to create this new model are going to be presented in the relative section.

Concerning the identification of the problem, it could be defined with Hevner's second guideline: 'Problem Relevance'. Recognising the relevance means also defining precisely the problem, the application domain and conditions. The target of analysis was limited to a specific decision-making process using data as support. In the application, the model was tested in the distribution planning. Consequently, the manufacturing and distribution companies were identified as target to apply the model. The object of the study treated by the maturity model is a theme considered innovative and also relevant to researchers: the development of a tool for the evaluation of data-driven decision-making process. This was even stated by one reviewer of Miragliotta et al. (2018) paper that recognized the relevance of the topic under analysis both for the academia and for the industry. Moreover, the knowledge and level of interest later was also better established by interviewing the potential users of the model.

## Second phase: Elaboration

This phase mainly consists in defining and designing the architecture of the model and the relative parameters. In particular taking as reference the OEE factors, three main dimensions of assessment were identified: Data Availability, Data Quality and DSS Performance. For each one, several sub-dimensions were designed in accordance with literature studies. In the section “Model Structure” the model architecture will be presented detailed in every aspect.

As stated by Becker et al. (2009) the aim of this phase is also the determination of development strategy and consequently the iterative model development. This could be done following the Hevner’s ‘Guideline 6: Design as a Search Process’. In fact, the elaboration of the model features should be done following an iterative procedure and a multi-methodological approach. The use of Delphi studies together with expert interviews helped in the definition and iterative consolidation of model sub-components. The methods selected need to be attuned in a rigorous way.

## Third Phase: Construction

The main scope of this phase is the practical conception and creation of the model. Therefore, the following point submitted concerns the maturity evaluation. According to the different sub-dimension under analysis, five maturity levels were identified. To each level was given a quantitative evaluation, depending on the way it has been made the relative sentence. Thanks to the quantitative evaluation given, it is possible to link to the formulation of an index capable to return a number - precisely a percentage – giving a clear indication of the current state to the users of the model.

The central phase of the process was the definition of maturity models, done in accordance with an iterative procedure. The first version presented of the model was developed in accordance with the literature studies conducted. Whereas the final version of the maturity model later proposed was obtained thanks to modifications made possible on the basis of users interviewed resorting to case studies and expert opinions. Also, in the model construction, it was considered essential to follow a multi-methodological procedure, as just explained in the Elaboration process. This always ensuring that the fifth guideline ‘Research Rigor’ was respected by reference to appropriate data collection and analysis techniques.

The artifact was developed trying to maintain applicability and generalizability, thus always looking for rigor.

#### Fourth phase: Deployment

The deployment was carried out for the theoretical validation of the model thanks to the application of case studies. Always taking as reference Hevner et al., (2004), this phase could be better explained according to the ‘Guideline 3: Design Evaluation’. The author states, “utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods. Evaluation is a crucial component of the research process”. Indeed, considering the design as an iterative and incremental activity, the evaluation phase represents the opportunity to receive essential feedback to the construction phase. In order to deal with this issue, to design the maturity model in a complete and effective way, it was used an observational methodology, ensuring the compliance with the requirements and the constraints. This means an in-depth study of the artifact in the business environment, through the use of case studies. The whole process to validate the model and provide practical information was conducted following the Yin methodology (2003), better explained in the related section.

According to Hevner et al. (2004), the last rule is ‘Communication of Research’ emphasizing the importance of targeting the results at the specific user group. According Becker et al. (2009) this guideline can be explained through two principles. The first one is the ‘Targeted Presentation of Results’ meaning that the model should be targeted to the users’ needs. In this case, since the work has both academic and practical aim, a dissertation and an Excel dashboard were developed. While the last one is the ‘Scientific Documentation’ indicating and recommending the necessity of a detailed documentation of the design process, considering for each step of the process, the parties involved, all the methods implemented, and the results obtained.

#### Fifth phase: Maintenance

This phase is not present both in Hevner et al. (2004) guidelines and in Becker et al. (2009) requirements. Referring to literature, the aim of this step could be addressed to the process of assuring and preserving the model updated and reviewed over time. As concerns the work

done in this thesis, the maintenance was done after the model application in case studies, through the implementation of modifications in order to preserve the scope of the model and clarify its application.

# 6 Model building

At this stage the inception, elaboration and construction of the model are described. The first point faced is the presentation of the theoretical model structure, key dimensions and the relative sub-dimensions of analysis. After that the maturity levels are defined and better explained in their sentences and quantitative assessment. Finally, the transition from the quantitative assessment of maturity levels to the drafting of an index was done and better explained with the help of the formula descriptions.

## 6.1 Model Structure

The theoretical model's structure takes inspiration from the OEE metrics that is an aggregated metrics, representing a global standard to measure the production unit efficiency. Its application could be done at three different levels: to measure the global performance of a plant, to compare the different lines performance and to identify the worst performing machine. OEE is a function of three mutually exclusive components - availability, performance and quality - enabling companies to examine and analyse each factor separately.

The Availability refers to the machine actual production time, which takes under consideration the total stoppage time – due to unscheduled breakdowns, waiting time, changeovers - differing in this way from the potential production time. The Performance factors aim to detect the percentage of the actual output made by a machine, which differs from theoretical output because of minor stoppage for obstructions, feeding errors and machinery running with lower speed. As concern the quality rate, it indicates the portion of defective product due to scrap produced and related rework.

Starting from these three factors, it was possible to identify a measurement for the productivity of data in decision making process. This was done thanks to the formulation of a

metrics made by three major components: Data Availability, Data Quality and DSS Performance.



Figure 45 Data Productivity Factors

Generally, these three main components could be defined as follows:

- *Data Availability* aims to detect if data are accessible and available to the user when required, supporting the decision maker.
- *Data Quality* aims to evaluate if data available to the user are of proper quality to support the specific decision.
- *DSS Performance* aims to evaluate the decision support system, assessing the information system capability to integrate and process data, in order to properly support the user within a decision-making process.

To assess each component, a set of subfactors have been identified, leading to the final construction of the theoretical model (Figure 47). The subfactors selection of the components of Data availability and Data Quality was mainly driven by the data assessment approaches proposed in literature (Wang et al. 1995; Wand et al. 1996; Lee et al. 2002; Liu et al. 2002; Pipino et al. 2002; Bovee et al. 2003; Cai et al. 2015). In particular the model structure is inspired by the one introduced by Cai et al. (2015), since it discusses the data quality assessment under the Big Data advent, taking into consideration the new related challenges triggered. In fact, nowadays the new issue and challenges is to extract data having a good quality from the multiple sources available to enterprises (e.g. sensors, multiple databases and so on). Concerning the last component, DSS performance, the literature does not provide a predefined sub-set of factors to assess the decision support system of an enterprise. Therefore, the identification of the dimensions applied to evaluate this component was the results of the literature on the data management, knowledge

management, business intelligence and DSS field. To sum up, it was conducted a literature study firstly to better understand the topic and the related themes and consequently on the data assessment dimensions. In this way the model structure was defined together with the model dimensions' definitions.



Figure 46 Elaboration of model dimensions

The final model structure is: Data availability described by data existence, authorization, accessibility and timeliness; Data Quality defined by data accuracy, credibility, relevance, completeness, presentation quality and consistency; DSS Performance identified by data integration, data processing, supporting infrastructure and data usage.

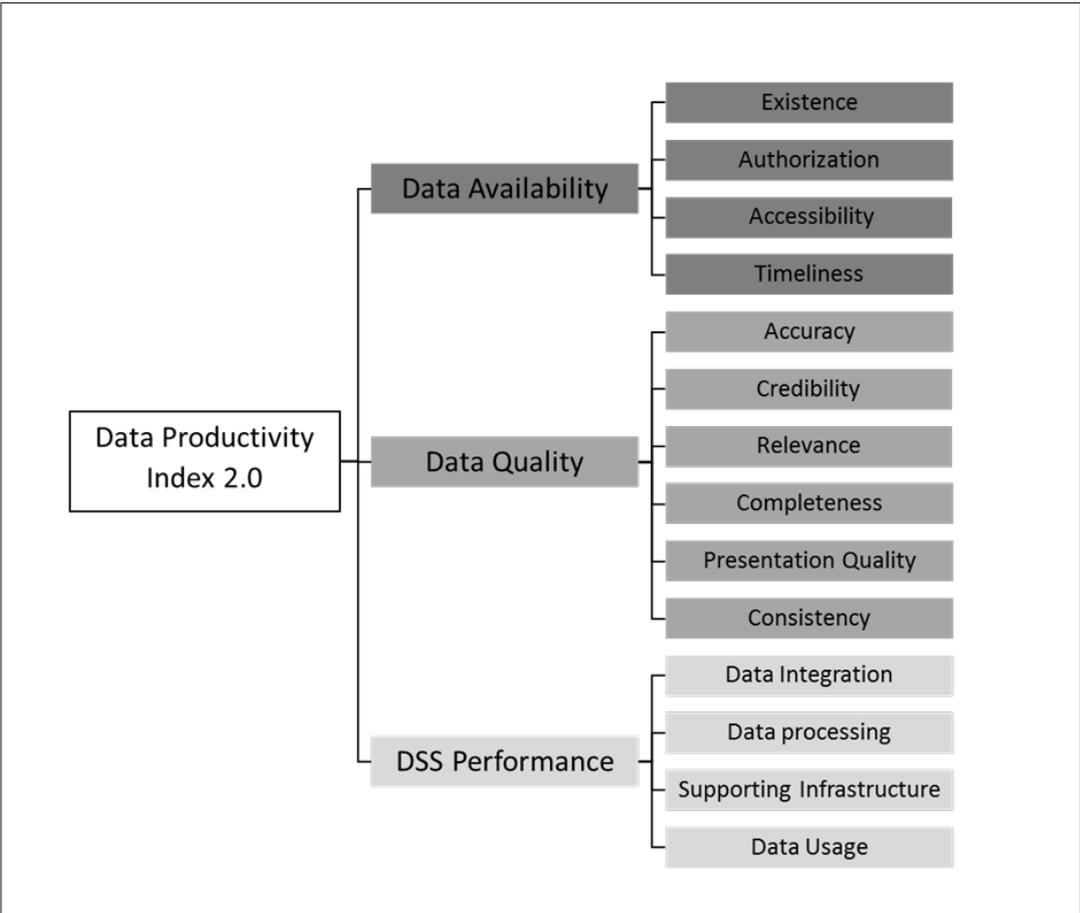


Figure 47 Data productivity factors and sub-factors

## 6.1.1 User perspective and task dependence

Before analysing each factors of the model introduced, some observation concerning the data assessment must be done.

Indeed, data evaluation is something strictly connected to the user and to the application context. Wang et al. (1995) discussed that multiple users could have different quality requirements, and even for the same individuals, various data applications might need different data quality features. Data assessment, for its nature, is subjective and strictly related to the user perceptions. Therefore, in order to obtain a real data evaluation, the user's perspective must be taken into account, understanding which the user expectations for the data under analysis are.

As mentioned, the data assessment is also strictly linked to the application context; in fact, Pipino et al. (2002) discusses that data assessment could be made performing two paths: task depended and task independent. Also, Wand and Wang (1996) notice that data quality depends upon the use of data: what could be defined as good and high-quality data for one circumstance, it may not be adequate for another one.

In this study, data assessment is performed taken into account both the data user and the application context:

- Context: the proposed model is applicable to a specific decision-making process.
- User: the dimension of analysis must be assessed by an individual which is strictly related with the decision-making process selected.

In this way it is pursued the main objective of the proposed model: assessing how data support the user within a specific decision-making process.

In the following paragraphs each component will be deeply described through the subfactors definitions.

## 6.1.2 Data Availability

As mentioned before, in the OEE metrics the availability factor refers to the actual production time. Thanks to a parallelism, the Data availability factor relates to those data that are available to support the user in a specific decision-making process.

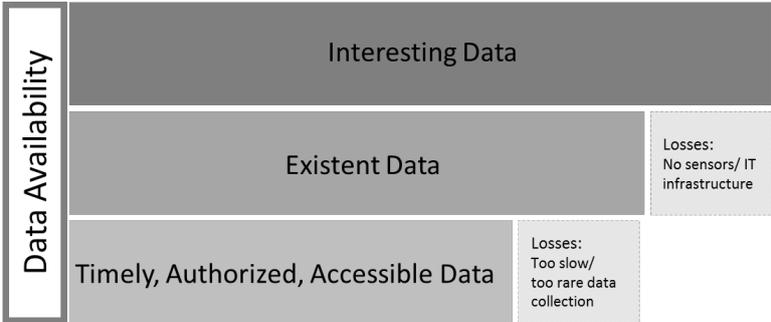


Figure 48 Data Availability

As shown in *Figure 48*, the starting point of the Data Availability component is the definition of a set of *Interesting data*. The interesting data could be defined as a list of data deemed necessary for the decision-making process under analysis. This list is the input of the model and therefore, its definition is considered a crucial point in model application and implementation. Then the Existent Data are identified, as a subset of the Interesting data, selecting those ones that are actually generated and collected by the company. Finally assessing the data authorization, accessibility, and timeliness, the data availability parameter is obtained. All the features, just shortly presented, are going to be better explained in the following part.

In literature the concepts of data availability, timeliness and accessibility are considered, under different perspective, as data quality dimensions (Wang et al. 1995; Wand et al. 1996; Lee et al. 2002; Liu et al. 2002; Pipino et al. 2002; Bovee et al. 2003; Cai et al. 2015). With the aim to follow the OEE structure, taken as a reference into the model construction, it was taken the decision to broke up the broad concept of Data quality introduced by the different authors, distinguishing between the two-macro factors: Data Availability and Data Quality.

## Data existence

The data existence stresses which is that part of the interesting data lost due to poor data collection. This loss is mainly due to the difficulty of a company to collect a particular set of a data coming from external or internal sources. In particular the losses of internal sources are mainly due to the lack of sensors able to capture data across company processes and activities, for example sensor able to trace the components position within company warehouses or distribution centres. Moreover, both the lack of external and internal sources may be explained by the inability of the IT infrastructure to collect and memorize all the interesting data required to properly support a specific decision-making process. Generally, these losses could be traced back to enterprises managerial choices and IT infrastructures capability. Summarizing, the concise definition of Data Existence proposed in this work is:

*“Data Existence refers to the extent to which data are effectively generated and collected in a company”*

Once the subset of existent data is identified, the other dimensions must be assessed.

## Data Authorization

The authorization feature appears in the work of Chai & Zhu (2015) as one of the sub factors defined to asses data availability. Among the methodologies analysed, this work is the only one including this dimension. It is aim is to verify if the users own the right to access at the existent data, in order to exploit them during the decision-making process. For example, it could happen that a particular dataset belongs to a specific individual/department and therefore, an authorisation is required to the decision maker to deploy it. The authorization process may necessitate a particular effort, being resources and time consuming.

Definition	References
Authorization refers to the extent to which an individual or organization has the right to use the data.	Cai & Zhu (2015)

*Table 11 Data Authorization definition*

Due to the lack of literature, the Data authorization definition is mostly driven by the one proposed in Cai & Zhu (2015) work:

*“Data Authorization refers to the extent to which an individual or organization has the right to access the data”*

### Data Accessibility

The dimension Data Accessibility is proposed into multiple works addressing the data availability field and more in general the data quality assessment issue. This feature aims to assess the process required to obtain a specific data, evaluating if this is easily and quickly retrievable by the users. In order to be accessible data should own a high level of openness, being available across the whole organization and a user interface must be provided. Data accessibility is a process that may require effort, human resources and time.

Table 12 shows the different accessibility definitions proposed by the authors.

Definition	References
Accessibility refers to the extent to which is difficult for users to obtain data; whether a data access interface is provided, data can be easily made public or easy to purchase. Accessibility is closely linked with data openness, the higher the data openness degree, the more data types obtained, and the higher the degree of accessibility.	Cai & Zhu (2015)
Accessibility refers to the extent to which the information retrieve is fast and easy. Information retrieval may require a certain amount of time or have an associated measure of cost to the user.	Bovee et al (2003)
Accessibility refers to the extent to which data is available, or easily and quickly retrievable.	Pipino et al. (2002)
Accessibility refers to the extent to which data are available, existing in some form that can be accessed.	Wang et al. (1995)

Table 12 Data Accessibility definitions

Summarizing, Data Accessibility in this work is defined as follow:

*“Accessibility refers to the extent to which is difficult for users to obtain data: if they are easily and quickly retrievable and if a data access interface is provided”*

The boundary between the concepts of data authorization and data accessibility is fuzzy: the accessibility focuses on the process need to reach a certain data, while authorization checks if the user owns effectively the right to get this specific data.

### Data Timeliness

Data Timeliness notion is introduced by almost all the authors dealing with data assessment. Timeliness is related to the concept of real time, defined as the system capability to answer as quickly as needed to the user. Decision makers should rely on data that are constantly and regularly updated, in order to exploit their potential within the specific decision-making process. To define data timely, the time frame between data collection, processing and utilisation should be reduced to the minimum. This concept becomes particular crucial, with the advent of big data; in fact, nowadays data contents are constantly changing, increasing data and information volatility.

Definition	References
Timeliness refers to the extent to which within a given time, data arrive on time, data are regularly updated and whether the time interval from data collection and processing to release meets requirements.	Cai & Zhu (2015)
Timeliness refers to the extent to which data are aged and volatile. Age is a measure of how old the information is, based on how long ago it was recorded. Volatility is a measure of information instability, the frequency of change of the value for an entity attribute.	Bovee et al. (2003)
Timeliness refers to the extent to which data is sufficiently up-to-date for the task at hand.	Pipino et al. (2002)
Timeliness refers to the extent to which data are sufficiently up-to-date for a task.	Liu & Chi (2002)
Timelines refers to the extent to which the delay between a change of the real-world state and the resulting modification of the information system state.	Wand et al. (1996)
Timeliness is characterized by currency – when the data item was stored in the database - and volatility – how long the item remains valid.	Wang et al. (1995)

Table 13 Data Timeliness definitions

A not proper data updating leads to data losses; the cause of these losses must be found on the IT infrastructure capability to keep update the data coming from both internal and external sources. Enterprises may adopt the so-called real time infrastructure that enables a constant and automatic data and information sharing across all the functions and departments. For instance, if data related to the current state of a machinery are collected every three months, it could be difficult to properly plan the maintenance and to avoid machine breakdown. The wrong identification of data collection frequency and/or too slow data transitions could be considered as losses. *Table 13* exhibits the detailed definitions of timeliness given by each author.

Taking inspiration from the definitions presented, in this work timeliness is basically defined as follows:

*“Timeliness refers to the extent to which data are regularly updated for the task at hand and whether the data arrive on time”*

### 6.1.3 Data Quality

In the OEE metrics the quality factor measures the percentage of actual production obtainable taking into consideration the losses due to rework and scraps. Continuing with OEE parallelism, the Data quality factor aim is to detect which are those data owing poor quality, by limiting their utilisation and exploitation. These losses are mainly due to inaccuracy, non-credibility of the sources, missing values, format inconsistency across time, and meaning ambiguity.

Data quality issue is deeply discussed in literature due to its increasing importance within organization. Notably, with the advent of big data always more sources are available, leading to a massive data collection. In this context, high data-quality should be ensured, in order to exploit and extract value from data. In literature, data quality is commonly defined both as hierarchal and multidimensional concept where each data quality dimension is related to a particular data aspect. The methodologies analysed propose multiple perspective to assess data quality, defining different set of dimensions.

Therefore, data quality dimensions proposed in the model were selected and defined taking as a reference the rich body of literature addressing the theme. This selection was mainly driven by the work proposed by Cai & Zhu (2015), since its aim is to review the previous data quality assessment methodologies under the big data perspective. The resulting dimensions are accuracy, credibility, completeness, presentation quality and consistency.

### Data Accuracy

In Data quality theme, accuracy plays a primary role in most models of both data and information quality, defined as an intrinsic characteristic. This dimension is present in all the methodologies analysed under the following different names: free-of-error, accuracy or correctness. Assessing this factor is challenging requiring standard roles and procedures, to make it objective. Although, the methodologies proposed do not provide a unique accuracy definition, most of the authors agree with the general definition which establishes that accurate data are those one free of error. Unprecise data are not able to describe the real state of the world, becoming inappropriate to support decision makers within decision-making process.

Table 14 sets out the accuracy definitions introduced in literature.

Definition	References
Accuracy refers to the extent to which data representation (or value) well reflects the true state of the source information and if information (data) representation will not cause ambiguity.	Cai & Zhu (2015)
Accuracy refers to the extent to which information are true or error free with respect to some known, designated, or measured value.	Bovee (2003)
Accuracy refers to the extent to which data is correct and reliable. This dimension represents the data correctness.	Pipino, et al. (2002)
Accuracy refers to the extent to which collected data are free of measurement errors.	Liu & Chi (2002).
Accuracy refers to the extent to which information system represent the real word.	Wand et al. (1996)

Table 14 Data Accuracy definitions

To ensure the accuracy of a certain dataset, targets must be defined as a comparable and reference values. In certain case data inaccuracy are easily traceable and detectable, such as

the gender which is describe by two distinct values; while in other cases the context of data application becomes crucial to assess the data correctness. Data inaccuracy may result from human mistake, data manipulation, manual data collection, and generally form measurement errors. Inaccuracy could also arise for intentional falsification and distortions, creating data that do not match anymore with the real word objects.

Within this work, the accuracy definition resulting from the literature analysed, is as follows:

*“Accuracy refers to the extent to which data are correct and free-of-error”*

### Data Credibility

In literature, the concept of data credibility is not always introduced explicitly, being presented under different perspectives. The dimension credibility is explicit presented by Cai & Zhu (2015) referring to the believability and reliability of the data collected and generated by enterprises. Other authors proposed different definition that could be traced back to the credibility definitions such as reputation and objectivity. Therefore, within this work it is considered as a broad concept involving the definitions of reputation, objectivity and trustworthiness of the collector (Table 15).

Definition	References
<i>Credibility</i> refers to the extent to which a source or message is believable. Credibility of data has three key factors: reliability of data sources, data normalization and the time when the data are produced.	Cai & Zhu (2015)
<i>Reputation</i> refers to the extent to which data is highly regarded in terms of its source or content. <i>Objectivity</i> refers to the extent to which data is unbiased, unprejudiced, and impartial.	Pipino et al. (2002)
<i>Trustworthiness of the collector</i> refers to the extent to which the collector has integrity of not committing falsification. <i>Objectivity</i> refers to the extent to which the sample selected for observation is representative of a population.	Liu & Chi (2002)

Table 15 Data Credibility definitions

The definitions embedded in credibility concepts are presents just in few methodologies found in literature. Notwithstanding this, it was taken the decision to include this dimension in the model presented for two consequential reasons:

- It is acknowledged the influence of data sources credibility on the overall data quality assessment.
- it is recognised the negative impact that incredible data sources could have when deployed to support a decision-making process.

The aim of this factor is to verify the believability of the data sources, detecting when data come from specialized company of the specific field, industry or sector. To ensure data credibility practitioners should periodically monitor and control data sources deployed by companies. Usually data sources may be distinguished among three different classes: external, internal and personal. Companies should adopt only certified and specialized sources in order to obtain data characterized by high reliability.

In this work, the definition of this factor is summarised as follows:

*“Credibility refers to the extent to which data sources are checked, certified and trustworthy.  
But also, to the extent to which data is unbiased, unprejudiced and impartial”*

## Data Completeness

The completeness dimension, together with accuracy and timeliness, is included across all the quality assessment methodologies investigated to perform this study. Its assessment is made analysing the presence of all values of a single component. A certain data is defined complete if all the values and parts required to obtain a comprehensive description are present.

The definition provided by Wang et al. (1995) goes beyond the general one that views a set of data as complete if “all values for a certain variable are recorded”. Indeed, the authors define data completeness as the capability of an information system to provide every remarkable state of real word. This definition is not only based on the missing data values, but rather its aim is to analyse the possible data combination. This broader concept of data

analysis, combination and representation across various dimensions and meanings will be further included in the model with the DSS performance factors. The lack of values within a dataset may stem from data manual collection, employee manipulation, inappropriate maintenance of databases and information systems. *Table 16* sums up the data completeness definitions found in literature.

Definition	References
Completeness refers to the extent to which the values of all components of a single datum are valid.	Cai & Zhu (2015)
Completeness refers to the extent to which all required parts of an entity's information are present.	Bovee et al (2003)
Completeness refers to the extent to which data is not missing and is of sufficient breadth and depth for the task at hand.	Pipino, et al. (2002)
Completeness implies that all values that are supposed to be collected as per the collection theory should be collected.	Liu & Chi (2002)
Completeness refers to the extent to which information has a sufficient breadth and depth for the task.	Lee et al. (2002)
Completeness refers to the extent to which data are of sufficient breadth, depth, and scope for the task at hand.	Wand et al. (1996)
Completeness refers to the extent to which an information system is able to represent every meaningful state of the represented real word system.	Wang et al. (1995)

*Table 16 Data Completeness definitions*

On this work the completeness definition presented, emphasises that the assessment of completeness must be related to the context in which data are deployed.

*“Completeness refers to the extent to which data is not missing and of sufficient breadth, depth, scope for the task at hand”*

### Data Consistency

In literature, consistency is related to multiple data aspects including their value, representation and physical appearance. A consistent data must keep its value domain, meaning, format and representation across time and space. Moreover, according to Cai & Zhu (2015) data are consistent if they are logically correlated, owing the same meaning and

representation when located in different storage areas. To ensure consistency, keeping the same or almost similar data representation across time and space, a process of data synchronization must be performed. To guarantee this feature, the information system should have the ability to keep constant the format also after data treatment and processing.

Inconsistency and irregularity within a database may be due several reasons such as:

- lack of a uniform and standardized rules to represent data;
- lack of data updating monitoring – for example two individuals may update at the same time the same data;
- lack of an automatic mechanism to update data – for examples there is not an automatic updating of ID-codes when there are changing across time.

Table 17 exhibits the consistency definitions introduced in literature.

Definition	References
Consistency refers to the extent to which after data have been processed, their concepts, value domains, and formats still match as before processing.	Cai & Zhu (2015)
Consistency refers to the extent to which multiple recordings of the values for an entity’s attributes are the same or closely similar across time or space.	Bovee et al. (2003)
Consistency refers to the extent to which data are presented in the same format.	Pipino et al. (2002)
Consistency refers to the extent to which different data in a database are logically compatible.	Liu & Chi (2002)
Consistency refers to the extent to which data values are expected to be the same for the same situation.	Wand et al. (1996)

Table 17 Data Consistency definitions

In term of this model, consistency is defined as follows:

*“Consistency refers to the extent to which data are: (1) presented in the same format, being the same or closely similar across time or space, (2) after processing their concepts, value domains and formats still match as before.”*

## Data Presentation Quality

Presentation quality is an overall concept, formally introduced only by Liu & Chi (2002) and Cai & Zhu (2015). Its aim is to assess if data are clearly presented, codified and easily understandable in order to match the required needs. As suggest by Liu & Chi (2002) It encloses different dimensions defined by the other authors addressing the theme such as: understandability, concise representation and interpretability (Pipino et al. 2002); interpretability, concise representation (Lee et al. 2002); faithfulness, neutrality and interpretability (Liu & Chi, 2002); interpretability (Bovee et al 2003; Wang et al., 1995).

Presentation quality refers to those data capable of being understood by the user and that transmit a meaningful or significative message. Data must be able to “speaks for itself” (Wang et al. 1995), being easily interpretable through the use of a concise and appropriate syntax and semantics. Presentation quality is one of the dimensions mostly affective by the user’s perceptions; something could appear highly meaningful and clear for one user and not interpretable and understandable for another one.

Definition	References
Presentation quality – Readability refers to the extent to which data content are correctly explained according to known or well-defined terms, attributes, unit, codes, abbreviation or other information.	Cai & Zhu (2015)
Interpretability refers to the extent to which information retrieved are understandable, deriving a meaning form it.	Bovee et al (2003)
<i>Understandability</i> refers to the extent to which data is easily comprehended. <i>Concise representation</i> refers to the extent to which data is compactly represented. <i>Interpretability</i> refers to the extent to which data is an appropriate languages, symbols, and units, and the definitions are clear.	Pipino, et al. (2002)
<i>Faithfulness</i> refers to the extent to which the presented data are identical to the origin in meaning and precision. <i>Neutrality</i> refers to the extent to which data selected for presentation are not in favour of any particular opinion or purpose. <i>Interpretability</i> refers to the extent to which data have clear meaning.	Liu & Chi (2002)
Interpretability refers to the extent to which the user is able to understand the syntax and semantics o data.	Wang et al. (1995)

Table 18 Data Presentation Quality definitions

In this work Presentation quality is defined as follows:

*“Presentation quality refers to the extent to which data are clear and understandable (understandability) and compactly presented (concise representation)”*

## Data Relevance

Lastly to ensure quality, data must be considered relevant for the specific domain and purpose of interest within the specific application context. This dimension is presented across the various data quality assessment hierarchy. The major part of the authors agrees with the general definition stating that data to be relevant must match the theme and the user requirements. Relevant data are helpful and supportive tools within a determinate decision-making process, driving the user’s evaluation and choices.

According to Cai & Zhu (2015) the relevance dimension involves also the assessment of the amount of available data effectively deployed by the user. In this work the data usage is included later, as a specific dimension required to evaluate the DSS performance.

Definition	References
Relevance – Fitness refers to the extent to which data produced matches users’ needs in the aspect of indicator definition, elements, classification.	Cai & Zhu (2015)
Relevance refers to the extent to which the domain and purpose of interest are specific to the user and determine the desired information pieces.	Bovee et al. (2003)
Relevance refers to the extent to which data is applicable and helpful for the task at hand.	Pipino, et al. (2002)
Relevance refers to the extent to which data are applicable and useful for a specific theory.	Liu & Chi (2002)
Relevance refers to the extent to which data fit requirement for making the decision.	Wang et al. (1995)

*Table 19 Data Relevance definitions*

Into this model, relevance defined as follows:

*“Relevance refers to the extent to which data match the theme and are helpful for the task at hand.”*

## 6.1.4 DSS Performance

The third factor of the model is related to the performance of the DSS. The aim of this dimension is to analyse if the underlying infrastructure is capable to support the decision makers, exploiting data and extracting value from them. After data have been generated and made available with a proper quality to the user, they must be adequately treated to effectively support the decision makers within a specific process. Therefore, also the infrastructure must be able to properly support the user during the decision-making process.

The performances of the decision support systems are becoming particularly crucial with the Big Data advent, where the amount of data processed and treated by organizations is constantly increasing. The information systems are continuously mutating going from a hierarchal structure to a network-base one, increasing the potential sources available to companies. Data generation is always becoming easier and faster, increasing the company capability to collect massive amount of heterogenous data. Nowadays, organizations must deal more and more with unstructured data; according to Cai & Zhu (2015) more than 80% of all data collected by companies is unstructured, increasing the need to own technologies able to transform semi-structured or unstructured data into structured one.

Therefore, organizations should be capable to deal with the large amount of data that may potentially be generated and collected. Especially in Big data context, company technology infrastructure should be able to:

- speed-up the storage processing large amount of data;
- to perform data analysis techniques such as data mining, simulation techniques in order to understand and exploit the value retrievable from the large amount of data available;

The DSS performance is broken into four components: Data Integration, Data Processing, Supporting Infrastructure and Data Usage.

## Data Integration

The first factor analyses the level of integration among data coming from different sources such as external and internal databases, sensors, and so on, therefore focusing on how data are stored and shared within a company.

To obtain high level of integration companies must deploy integrated databases (Zhou et al. 2015) to incorporate multiple sources of data coming from the various functional groups and departments; and integrated interfaces (Zhou et al. 2015; Hermann et al. 2016) to share and communicate data across the organization, avoiding discontinuity. Interfaces must be provided to enable communication among organization, human machine and user/product. In companies, data integration enables interoperability which is the capability to cooperate and exchange information and services through reliable systems, optimizing resources. Information integration and exchanges among the multiple players, allow the data user to obtain a more comprehensive and global view of the current company state. This higher visibility enhances the user ability to generate alternatives and therefore the capability to optimize the entire decision-making process.

In order to easily manage and integrate data coming from different sources and departments, companies should own databases and systems that are able to connect each other. Data integration is also one of the main objectives defined by the Industry 4.0. Standardized systems must be developed across companies, operating within the same supply chain, setting uniform standards in order to gain integration and connection (Zhou et al. 2015).

Organizations to be capable to integrate data coming from the different sources and actors of their supply chain, should be fully digitalised.

Formally, the definition given to Data Integration is:

*“Data Integration refers to the (1) interoperability, defined as the capability to cooperate and exchange information, and to the (2) level of integration among data coming from multiple sources (e.g. different databases, sensors).”*

## Data Processing

The Data Processing dimension examines how data are treated to properly support the user during a determinate decision-making process. Indeed, for organizations “to reach competitive advantages, voluminous data needs be managed, analysed and fed into the decision-making process” (Gorla, 2003).

Data analysis is made possible by the deployment of data warehouses, which provide a support to companies thanks to analytical database and online analytical processing tools (OLAP). These tools are able to make multidimensional analysis, obtaining a general view of the companies’ operations. Moreover, these tools are capable to perform data analysis discovering hidden pattern among multiple data set and what-if analysis. Data should be analysed and viewed under multiple perspectives performing aggregation, drill down and roll up data, in order to detect and get more details and information. The level of data aggregation is particularly significant in supporting a decision-making process. It enables to aggregate data into metrics and indicators (such as KPI), providing to the user comprehensive and general information.

Decision-making process could be fed up by the information retrieved by both structured and unstructured data. Data might be processed by highly innovative and advanced algorithms, deploying for example machine learning and artificial intelligence solutions. Indeed, machine learning are systems capable to perform algorithms that “learn from existing data without being exactly programming” (Hermann et al., 2016), creating value added and effectively supporting the decision makers.

Therefore, in this work Data processing is defined as follows:

*“Data-processing refers to the (1) data treatment by state-of-the-art data processes and algorithms with the aim to exploit their potential, and to (2) the ability to analyse data on various dimensions, aggregating, disaggregating and presenting data in form of indicators to better support the decision makers.”*

## Supporting Infrastructure

Supporting infrastructure aims to understand if the underlying company infrastructure, in term of hardware and software, is capable to deal and process massive amount of data. The company technologies must be able to extract information and knowledge from the huge amount of generated data.

Especially nowadays, in big data context companies own the ability to collect massive amount of data leading to unprecedented opportunities. The ability to treat and process data quickly, to extract valuable information from them, it is considered a real competitive advantage that leads companies to take better and faster decisions. As the potential available sources are increasing, also the amount of unstructured data is enhancing. Indeed, organizations are dealing with both structured and unstructured data, and the real value relies on the company ability to treat ad process unstructured data, changing them into structured one.

According to Power (2013) three main type of analysis may be performed by the information technology infrastructure of a company:

- Descriptive and retrospective data analysis which aim is to analyse historical data in order to detect pattern and understand why events have occurred.
- Predictive data analysis which objective is to simulate and generate different scenario through the historical data analysis, in order to predict and understand what could happen in the feature.
- Prescriptive data analysis which deploy quantitative analyses of the real time data in order to react to future events, planning recommend measures.

Descriptive and retrospective analysis are made possible by the classical algorithms and analysis performed by the data warehousing and business intelligence field. While in order to perform predictive and prescriptive analysis more innovative algorithms and technologies are required such as data mining, computer-based modelling, simulation or progressive data visualization techniques.

The performance of the underlying infrastructure is related also to its capability to process data fast enough to support timely the decision maker. The infrastructure should be able to elaborate and process data within a time frame considered reasonable by the user.

Below the definition of Supporting Infrastructure provided in this model.

*“Supporting infrastructure refers to the level of (1) adequacy of the underlying hardware and software infrastructure to run data analysis techniques and to (2) the capability to process data on time to properly support the decision maker.”*

## Data Usage

Finally, the last dimension introduced in the model is related to the effective use of the Data. Its aim is to understand which are those data that are effectively deployed by the decision makers during the decision-making process phases. This dimension is strictly related to the concept of limited rationality presented by Herbert (1947) which stressed that an individual makes decisions under limited cognitive capabilities.

Individuals do not make decisions under optimal circumstances since they are not fully aware about the context, objectives and about the potential value retrievable from the available sources. In an ideal situation, decisions are taken under full rationality than enables to select the best alternatives following a mathematical approach.

Therefore, Data Usage is defined as follows:

*“Data usage refers to extent to which data are actually used in a specific decision-making process by the decision maker”*

*Table 20* summarizes the definition of each dimension presented.

Dimension	Definition
Data Existence	Existence refers to the extent to which data are effectively generated and collected in a company.
Accessibility	Accessibility refers to the extent to which is difficult for users to obtain data: if they are easily and quickly retrievable and if a data access interface is provided.
Authorization	Authorization refers to the extent to which an individual or organization has the right to use the data.
Timeliness	Timeliness refers to the extent to which data are regularly updated for the task at hand and whether the data arrive on time.
Accuracy	Accuracy refers to the extent to which data are correct and free-of-error.
Credibility	Credibility refers to the extent to which data sources are checked, certified and trustworthy. But also, to the extent to which data is unbiased, unprejudiced and impartial.
Completeness	Completeness refers to the extent to which data is not missing and of sufficient breadth, depth, scope for the task at hand.
Presentation Quality	Presentation quality refers to the extent to which data are clear and understandable (understandability) and compactly presented (concise representation).
Consistency	Consistency refers to the extent to which data are: (1) presented in the same format, being the same or closely similar across time or space, (2) after processing their concepts, value domains and formats still match as before.
Relevancy	Relevance refers to the extent to which data match the theme and are helpful for the task at hand.
Data Integration	Data Integration refers to (1) the interoperability, defined as the capability to cooperate and exchange information, and to (2) the level of integration among data coming from multiple sources (e.g. different databases, sensors).
Data Processing	Data Processing refers to (1) the data treatment by state-of-the-art data processes and algorithms with the aim to exploit their potential, and to (2) the ability to analyse data on various dimensions, aggregating, disaggregating and presenting data in form of indicators to better support the decision makers.
Supporting Infrastructure	Supporting Infrastructure refers to (1) the level of adequacy of the underlying hardware and software infrastructure to run data analysis techniques and to (2) the capability to process data on time to properly support the decision maker.
Data Usage	Data usage refers to extent to which data are actually used in a specific decision-making process by the decision maker.

*Table 20 Model dimensions definitions*

## 6.2 Data Productivity Maturity Model - DPMM

In order to apply and measure the whole set of sub-dimensions identified concerning the data exploitation in decisions, it was decided to use the maturity model approach. This was considered a valid way to accomplish the evaluation of the data dimensions just presented and specifically to gather information regarding the current data-driven decision-making maturity level in the company. At this historic moment, the concept of maturity is increasingly being applied in very different domains and the most popular way to assess it is thanks to the use of maturity models.

### 6.2.1 Research contributions on the topic

Information and data are always a more important and crucial competitive factor, considering the huge potential they may represent in generating value for companies. The organisations' ability to process larger amounts of data makes possible to improve the decision-making process in a shorter time compared to competitors (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011).

The majority of companies do not hold enough sophisticated data management strategies. In particular large companies encounter difficulties in managing massive amount of data, since the lack of proper strategy to exploit them (Davenport and Prusak, 2000; Otto and Hüner, 2009). To assess the maturity level, it is possible to use maturity models (MMs) that are able both to describe an organizational context and prescribe the improvement actions to follow. Thus, they have a twofold objective: descriptive and prescriptive. From here on out, some maturity models on the topic – particularly the ones considered more useful according to us - are going to be presented.

### Business Intelligence and Data Warehousing Maturity Model

Focusing always more on the topic, in particular on Business Intelligence (BI), a high number of BI MMs has been developed. In the one proposed by Lahrmann et al., (2011) BI maturity is based on the concepts of “deployment”, “use” and “impact”. The impact of BI in the company is detailed into individual impact, related to the achievement of improved

decisions, and organizational impact, leading to better organizational performances. BI capabilities, BI practices, BI IT and organizational support together constitute the BI deployment system. In particular talking about BI IT, the model is considering BI applications, architecture, data and infrastructure. In the meanwhile, the concept of BI practices is connected to BI development, operations and management processes, but also to the strategic alignment and to BI services definition and measurement. BI capabilities relate exclusively to the skills and competences of the BI function itself.

In the field of data management, also the theme of data warehousing was studied following the maturity approach. Eckerson et al., (2014) presented their model consisting in six stages, as a benchmark for progress.

Stage	Major Characteristics
Prenatal: Management Reporting	Management reporting systems produce set of static report distributed on regular basis to the organisation employees
Infant: Spreadmarts	Unique set of data, metrics and rules are contained in not aligned spreadmarts (surrogate data marts)
Child: Data Marts	Knowledge workers tailor data marts according to their needs
Teenager: Data Warehouses	Departments have the need to standardize data marts, creating a central warehouse (data warehouse) with logical dependent data marts running in the same database
Adult: Enterprise Data Warehouse (EDW)	Integration Machine continuously consolidating all analytic structures, stewardship necessary to manage EDW and scorecard to align with the corporate strategy. In this phase investments in DW environment begin to pay off (positive effect on ROI)
Sage: BI Services	Additional opportunities offered by BI services: interactive extranets, web services, decision engines

Table 21 Research Contribution Maturity models - Business Intelligence

## Master Data Management Maturity Model

The main scope of the maturity model developed by Spruit et al. (2014) is to evaluate the maturity of the organization's current state of Master Data Management, including also the data point view and the focus on the operational process. Master Data Management (MDM) could be defined as *“the management of the consistent and uniform subset of business entities that describe the core activities of an enterprise”*. In the Table 22 the definition of MDM maturity levels and the relative brief description are provided.

Level	Description
1: Initial	A first awareness for issues regarding the topic of MDM has been raised on an operational level. Initial steps are initialized.
2: Repeatable	Measures from individuals are conducted to solve individual problems. No connection to other units or projects. Still operational.
3: Defined process	First collaborations take place on a tactical level. Awareness was created for the existence of other initiatives.
4: Managed and measurable	Best practices are in place for handling of MDM. There are defined processes on a tactical level.
5: Optimized	Optimized handling of MDM. The organization's efficiency has been improved. Tactical approach on the topic.

Table 22 Description of the derived MD3M Maturity levels (Spruit et al. 2014)

Moreover, the authors identified key topics and focus area, according to the main significant aspects of Master Data Management for organizations. These aspects were analysed at a right level of granularity and the model was applied following a loop approach. The model proposed is composed by 5 key topics and thirteen focus area.

Data Model	Definition of Master Data Master Data Model Data Landscape
Data Quality	Assessment of Data Quality Impact on Business Awareness of Quality Gaps improvement
Usage & Ownership	Data Usage Data Ownership Data Access
Data Protection	Data Protection
Maintenance	Storage Data Lifecycle

Table 23 Key topics and focus area in Master Data Management (Spruit et al. 2014)

## Big Data Maturity Model

Comuzzi et al., (2016) developed a MM centred on the organisations ability to leverage Big Data and the value created thanks to them. In fact, Big Data is an organisation-wide phenomenon and companies can exploit significant opportunities related to the availability of massive amount of both structured and unstructured data. Nevertheless, the Big Data potential is acknowledged, organisations are not yet able to make full advantage of it. According to Marr (2015) up to 50% of Big Data-related projects have never been finalised, since the market understanding of business implications related to the Big Data use is still scarce.

Data warehousing and Business Intelligence are two subjects strongly intertwined with Big Data. It is worth noting that both relatively to the first one (Eckerson, 2004; Sen et al., 2006) and the second one (Lahrman et al., 2011), different maturity models have been developed. Notwithstanding this fact, the scope of Big Data overcomes the data warehousing and business intelligence ones. More specifically, data warehouse is one of the many technologies used in Big Data, while business intelligence focus on the organisational decision-making process support (Comuzzi, & Patel, 2016). In the higher level of analysis, different domains were identified, together with the relative characterization by sub-domains (*Table 24*).

Domain	Sub-domain
Strategic Alignment	Strategy, Process
Organisation	People, Culture
Governance	Governance
Data	Management, Analytics
Information Technology	Infrastructure, Information Management

*Table 24 Research contribution Maturity model - Big Data*

## Industry 4.0 Maturity Model

Narrowing down to the domain under analysis, Schumacher (2016) developed a tool to assess the state-of-development in Industry 4.0 of manufacturing companies. The objective was to understand the overall impact of I4.0 on small-medium sized enterprises. The model

is composed by 62 maturity items, grouped in 9 company dimensions (strategy, leadership, customers, products, operations, culture, people, governance, technology), evaluated through 5 maturity levels, presented in form of standardized questionnaire. A relevant step of the model is the identification of item’s importance.

### Decision-making Maturity Model

Decision-making and knowledge management are two different concepts. The relationship between them was better explained by Holsapple and Whinston (1996) in their work: “Making a decision means that we are making a new piece of knowledge that did not exist before”. Therefore, it possible to state that the main aim of knowledge management is to provide support to decision-making.

Then also Kaner et al. (2004) developed a maturity model – Decision-making capability maturity model (DM-CMM) - dedicated to process-oriented knowledge-based decision-making, with the aim to identify and improve the ability to manage a given decision-making functionality. The DM-CMM presents two perspectives: the maturity levels refer to decision-making capability and the inter-levels maturation stages indicate the knowledge management capability. The model interrelates five maturity levels (ad-hoc, planned, defined, controlled, sustained) and four intermediary maturation stages (reception, revised, retained and reuse).

### Data productivity Maturity Model

Cianca (2018) in his work proposed a maturity model to measure data productivity in Industry 4.0. The model is composed by the following main aspects: the data itself and the means and tools applied to exploit the data potential. Each aspect is then described by four independent dimensions, addressing the critical topics of each area.

Data		Models and Tools	
Completeness	Communication	Control	Elaboration
Access	Frequency	Knowledge Exploitation	Integration

Figure 49 Model Structure (Cianca 2018)

The dimensions defined for the *Data* area are completeness, communication, access and frequency. While the dimensions identified for the *Models and Tools* area are control, elaboration, knowledge exploitation and integration. Each dimension of the model was assessed according to the maturity approach identifying five different level of maturity for each one.

Reference	Maturity domains	Assessment approach
Lahrmann, Marx, Winter & Wortmann (2011)	Business Intelligence	No details about items and maturity levels
Eckerson (2004)	Data Warehouse	Assessment in 8 dimensions: prenatal, infant, child, teenager, adult and sage
Spruit & Pietzka (2015)	Master Data Management	Assessment of 5 key topics and thirteen focus area according to 5 maturity levels: initial, repeatable, defined process, managed and measurable, optimized
Comuzzi & Patel (2016)	Big Data	Maturity evaluation according to 5 domains (Strategic Alignment, Governance, Data, Organization and Technology) and 9 related sub-dimensions
Schumacher, Erol & Sihh (2016)	Industry 4.0	62 maturity items grouped into 9 company dimensions
Kaner & Karni (2004)	Knowledge Management in decision-making	Two perspectives: defined 5 maturity levels (ad-hoc, planned, defined, controlled, sustained) interrelated with 4 maturity stages
Cianca (2018)	Data productivity	Assessment of 2 main aspects through 8 dimensions in 5 maturity levels

Table 25 Research contributors for DPMM

Finally, in the *Table 25* all the previous works analyzed and considered as source of inspiration for the DPMM were proposed and compared in a synthetic way.

## 6.2.2 Data Productivity Maturity levels

Starting from the literature review conducted, it turned up a gap in the wide range of maturity models, related to the proposal for quantitative measurement of data-driven

decision-making process performance. Therefore, a model to evaluate the maturity level of the dataset used by the organisation to make a specific decision is proposed: The Data Productivity maturity model (DPMM). The maturity model designed takes into consideration all the common features generally presented by this kind of methodology.

The process area clearly concerns the dataset assessment related to the precise decision-making process. Following the theoretical model previously proposed, the three factors - the data availability, the data quality and the DSS performance - were identified as the main dimensions of the measurement process. The use of maturity models refers exclusively to the measurement phase of the model proposed. It intends to evaluate the ability of the company to collect, maintain and use data in making decisions.

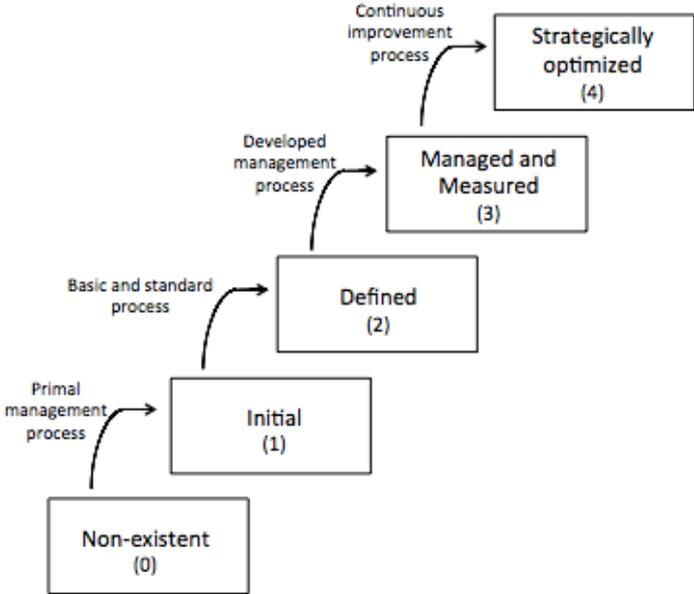
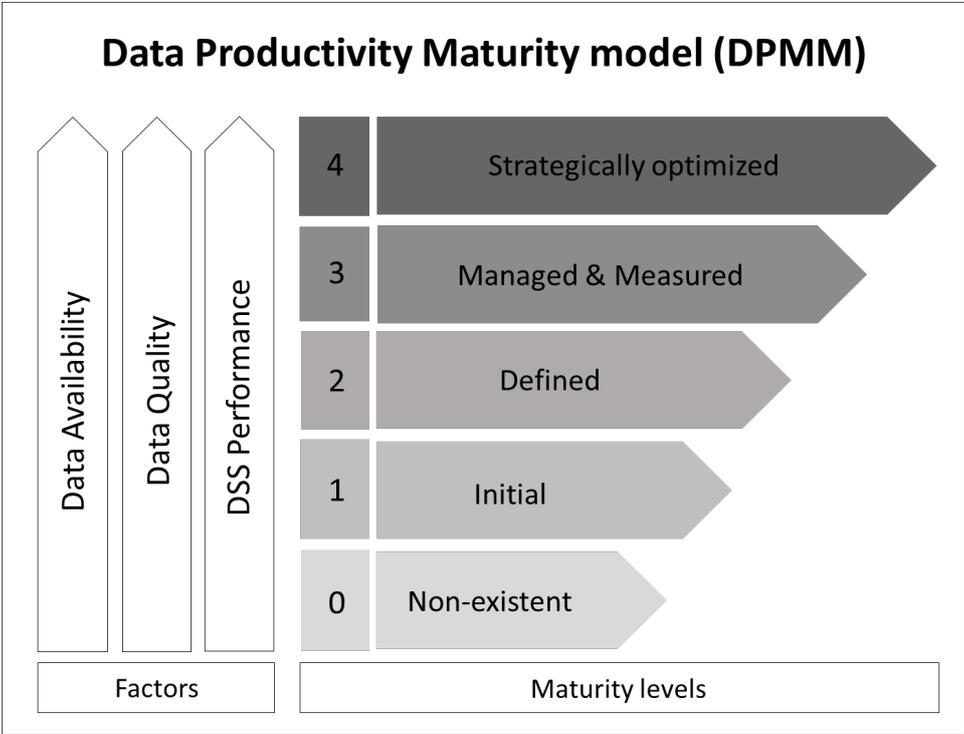


Figure 50 The five level of DPMM

The DPMM presents and interrelates five maturity levels - for each one was given a definition and a description – and four intermediary maturation stages. The selection of exactly five levels is not based on a precise theory. This number of levels was chosen first by Crosby (1979) for management intuition reasons and it is usually considered accepted for the maturity model formulation.

The five maturity levels identified are: non-existent, initial, defined, managed and measured and strategically optimized. These levels are related to the computation of each pillar’s sub-

factor identified. *Figure 51* provides a schematic view of the model and the state of the data-driven decision-making process at each maturity level. The arrows suggest the progressive direction from level to level together with the identification of maturity stages.



*Figure 51 Data Productivity Model*

Level 0 is the lowest one, characterized by ill-defined individualistic decision-making and total unconsciousness of the theme of data management in the specific process. Indeed, decisions are based upon judgement, following the individual knowledge resulted from unrecorded and tacit personal experience. On the other end of the spectrum, level 4 is the top-most level, indicating the optimized process in terms of data exploitation, aligned with organisation strategic objectives. Considered the novelty of the topic treated for the organisations, it was considered suitable to start the maturity assessment from level 0, instead of level 1. This was considered necessary since many organisations may be regarded at a familiarity level of the theme not yet sufficient.

**NON-EXISTENT** – No real sense of awareness on the topic. The organisations have no idea of the importance of collecting and managing data in the right way, with the aim to properly support the decision maker when taking decisions. Indeed, data are not collected, either

their quality or the supporting infrastructure are considered not capable to help the decider. The ignorance of this existing theme is considered no maturity at all; thus, the model cannot be applied.

**INITIAL** - A first and general awareness of the issue is present in the organisation, regarding the use of data in the decision-making process. Each feature starts to be managed, since it is present a primal management of the process and an informal and partially defined approach deployed in few areas, with different degree of effectiveness and sustainment. Processes of decision-making are based on a file drawer-type system of personal protocols, derived by personal experience. In this phase data management is in an irregular and undefined form, executed in an unconscious way with no uniformity and not following a specific aim. No plan has been made. The collection of data is partly carried out and the dataset might present low performances in terms of access and quality. Moreover, the Information System is considered low-end supportive.

**DEFINED** – Developed awareness of the acknowledged importance of structured data management in the decision-making process, thanks to the implementation of primary structures in the organisation, not yet institutionalised. Thus, the process could be described as formal and disciplined and it is based upon a model-based reasoning procedure. The data are managed through basic and standardized procedures and systematic approach/methodology developed across most areas, allowing a good sustainment. At this level the dataset could be considered of sufficient availability and of average quality. Finally, the information system is able to process and elaborate massive amount of structured data made available to the entire organisation, taking longer time than requested.

**MANAGED & MEASURED** – Best practices are in place for managing data in the decision-making process, using developed standardized and well-defined procedures measured and controlled across the organisation. Decision-making could be described as a model-based reasoning procedure, following a multi-step approach. Each step derives from rule results and outcome measures. Data management becomes a company core activity, pursuing the aims of developing, maintaining and operating the decision-making process. Indeed, the organisation considers data as a corporate asset to be treated according to a corporate

policy. Data availability and quality requirements are nearly completely met. The process is built on information exchange integration and interoperability across functional groups and it is fully planned and implemented. Both structured and unstructured data are processed through advanced data analysis techniques and algorithms.

***STRATEGICALLY OPTIMIZED*** – Exceptional and optimized handling of data in taking decisions. Data-driven decision-making process is well structured, controlled and collaborative and the relative efficiency has been improved. The system used for the decision-making has a reasoning-based procedure and it is established to learn about past experiences in similar contexts and improve the learning culture. In this phase the organisation is able to use its data to continuously improve and optimize the process, with the aim of enhancing at the highest level the data access, data quality and supporting infrastructure performance. The process results perfectly aligned with the company strategic business goals. The high potential growth of the company depends heavily on the excellent information exchange in collaboration between the company's functional groups and decision-making, mainly achieved through the great integration, interoperability, speed, soundness and security. Advanced analytics, such as machine learning and artificial intelligence solutions, play a major role for extracting valuable information from structured and unstructured data. These data are completely available to the organisation and of recognised quality.

The Data Productivity maturity model can be considered belonging to the maturity grids family, since the grid and the specific characteristics for each level were identified.

After having generally defined and described the maturity levels, at this stage it is necessary to decline them according to the dimensions of analysis taken under consideration, since different practices are associated to different sub-dimensions. In the following part, the defined levels per each dimension – Data Availability, Data Quality, DSS Performance - are going to be presented.

Level	Brief Description
Non-existent (Level 0)	No real sense of awareness of the topic. Ill-defined and individualistic decision-making process not supported by data
Initial (Level 1)	First and general awareness of the topic. Primal management of the process that is informal and partially defined
Defined (Level 2)	Decision-making process is defined and managed through basic and standardized procedures
Managed & Measured (Level 3)	Best practices are in place for data-driven decision-making. Standardized and well-defined procedures measured and controlled across the organisation
Strategically optimized (Level 4)	Optimized decision-making process aligned with organisations' strategic goals, characterized by well-defined structure, control and collaboration

*Table 26 Maturity Levels*

## 6.3 Data Productivity Index 2.0 – DPI 2.0

The aim of this work is to develop a metrics able to measure the data productivity, analysing how data effectively support and help user during a decision-making process. The starting point to develop the rate was the OEE metrics. It enables to build a model made by the three macro factors: Data availability, Data Quality and DSS performance. To assess each macro factor a deep literature study was carried out in order to understand and select the most appropriate sub factors to describe each macro factor. Finally, the model structure was developed selecting a set of features to describe macro factor identified.

Then each feature of the model was assessed by a qualitative approach: the maturity model. The main focus of this paragraph is to show how was made the transition from a qualitative approach to a quantitative one in order to obtain a comprehensive numeric index assessing the data productivity. The proposed index, named Data Productivity Index 2.0, emphasized how data are exploited by the user and by the information systems of a company, in order to become supporting means during a certain decision-making process.

### 6.3.1 From maturity to index

In the following part, the steps required to obtain the Data Productivity Index 2.0 are described.

In the last paragraph was deeply explained the identification of the generic maturity levels. Five maturity levels were identified starting from non-existent to strategically optimized. Each level generically describes the current state of a company in collecting, treating, processing and exploiting data for a particular decision-making process. The next step is the maturity model development for each feature of the model introduced.

The maturity approach is a qualitative assessment tool, deployed to obtain the current state of data, mitigating the assessment subjectivity through the presence of sentences. As mentioned, data evaluation is something extremely subjective, strictly related to the user and to the application context. Therefore, into this model it was taken the decision to firstly

assess data related to a specific decision-making process and secondly the feature evaluation, through maturity model, must be made by the data user. This last one should be someone that owns a high familiarity with both the selected decision-making process and with data collected by the company.

The key step to move from a qualitative approach to a quantitative one is to assign a numeric value that goes from 0 to 1 to each maturity level of each feature assessed. In this way the interviewee - the data user – selects, for each feature, the sentence describing the current state of data according to its perception; and beyond, each sentence is represented by a numeric value. Always keeping the score 0 for the level 0 and 1 for the last level of each maturity model, the other level values were chosen looking how the specific sentence is constructed for each dimension. Therefore, following a non-linear scale, different number have been adopted to each maturity level of each dimension, selecting arbitrarily the most suitable and representative value for the specific sentence.

0	<b>Non-existent</b>	No real sense of awareness on the topic	Relevance	<b>Data Availability</b>			<b>Data Quality</b>			<b>DSS Performance</b>			
1	<b>Initial</b>	A first awareness for issues regarding the use of data in the decision making process is achieved		Existence	Authorization & Accessibility	Timeliness	Accuracy & Credibility	Completeness	Presentation Quality & Consistency	Data Architecture & Integration	Data Processing	Supporting Infrastructure	Data Usage
2	<b>Defined</b>	Developed awareness of the necessary importance of a structured data management in the decision making process											
3	<b>Managed and Measured</b>	Best practices are in place for managing data in the decision making process											
4	<b>Strategically Optimized</b>	Optimized handling of data in taking decisions. The organization's efficiency in using data in decision-making process has been improved											

Figure 52 Model application

In detail the model is applied as follows: firstly, a list of data is provided as input of the model representing a set of *Interesting data* for the decision-making process under analysis. The aim of this input is to provide as much as possible comprehensive list of data that can be deployed and turn out to be useful for the user during the decision-making process selected. Secondly, it is asked to the interviewee to evaluate the relevance for each data of the list.

The user must evaluate how much interesting data are relevant and helpful, matching the theme at hand assigning a weight to each one. Indeed, the relevance dimension plays a different role within the model being deployed as weight for each data of the list provided as input.

Therefore, the relevance weight is applied to assess firstly the dimensions related to the first macro factor of the model, Data Availability, and then for the ones related to the second macro factor of the model, Data Quality. This is made through weighted average among the different evaluations given to each data. These computations are clearly explained in the following paragraph through the explicit formula representation. Finally, the dimensions related to the last macro factor of the model, DSS performance, are evaluated independently from the list given as input of the model.

Summarizing the model application is based on four main areas (*Figure 52*): Data relevance, Data Availability, Data quality and DSS performance. As just mentioned, each sub dimension is evaluated through five maturity levels and to each one a numeric value is assigned. The relevance is the only factor evaluated differently; indeed, it is deployed as a weight to aggregate the different evaluation given to each data of the list provided as input of the model.

In this following part are shown:

- the maturity models developed for each feature;
- the numeric value assigned to each maturity level of each feature;
- the formula applied to obtain the percentage representing each feature.

It is also presented how each future is aggregated in order to get a unique percentage to define firstly Data Availability, then Data Quality and lastly DSS Performance. Finally, the three macro factors are aggregated in order to gain the Data Productivity Index 2.0.

### 6.3.2 Data Relevance

The input of the model is a list of *Interesting data* for the specific decision-making process. To each data of the list, a weight that goes from 1 to 5, based on the data relevance for the specific decision-making process is assigned. The relevance to each data is given regardless

of which are those data that the company under analysis effectively collects and generates. The interviewee should assign a weight based on how much each data of the Interesting list could be helpful and useful in supporting the decision-making process selected. If on one hand, this evaluation seems to be totally subjective and based on the interviewee perceptions; on the other hand, this assessment is made by a person strictly related with the decision-making process selected that owns a high familiarity with the list of data proposed.

Level 0	Data not significant	1
Level 1	Data little significant	2
Level 2	Data averagely significant	3
Level 3	Data highly significant	4
Level 4	Data strictly necessary	5

Table 27 Relevance Assessment

Then, the relevance value is deployed as a weight to assess the other dimensions, more specifically to assess the data quality and availability features.

### 6.3.3 Data Availability

The assessment of each feature identified relatively the Data Availability dimension is done taking as object of analysis each single dimension specified in the list of *Interesting Data*. Considering the *Authorization and Accessibility* features, these two features were merged together for the maturity assessment after the application in the business environment through the use of case studies. Consequently, to the existence evaluation, all the other features will be evaluated only for those data estimated at a different existence maturity level from Level 0.

#### Existence

Firstly, the Existence of each data of the list is assessed in order to verify which is the degree of data collection of the company interviewed. Data existence is identified by the ratio between the existence maturity level of each data multiplied by its related weight and the sum of the weight of each interesting data. Therefore, the result of the formula is the percentage of data effectively collected and generated by the company.

To assess the percentage of existent data it was required to develop a maturity model since the list of data, given as an input of the model, is presented at an aggregate level. Therefore, it was not possible to deploy a simple Boolean variable (0/1) to assess this feature as happen in the previous work (Miragliotta et al. 2018). For example, assuming that the first data of the list is “Physical Products characteristics”, it could happen that a company generates just a part of the potential data collectable within this category, being therefore represented by the second or third maturity level.

Level 0	Data are not collected.	0
Level 1	The minor part of data is generated and collected.	0,3
Level 2	On average data are generated and collected.	0,6
Level 3	The major part of data is generated and collected.	0,8
Level 4	Data are fully generated and collected.	1

Table 28 Existence Assessment

$$\% \text{ Existence} = \frac{\sum_{i=1}^N R_i * E_i}{\sum_{i=1}^N R_i} \text{ where } \left\{ \begin{array}{l} i = 1 \dots n \text{ Interesting data for the decision – making process} \\ R = \text{Relevance's maturity level} \\ E = \text{Existence's maturity level} \end{array} \right\}$$

The other dimensions defined to evaluate Data Availability - Accessibility & Authorization and Timeliness - are assessed starting from the Existent Data, defined as a sub set of the Interesting data. This choice is driven by the fact that if a data is not collected and generated by the company, and therefore it is not physically present into the company’s databases, it does make sense to evaluate the other data aspects and features.

### Accessibility & Authorization

A single maturity model is adopted to evaluate both the accessibility and the authorization dimensions. This choice was made after having carried out the pilot case studies demonstrating the high correlation between these two features. Therefore, a unique maturity model was considered sufficient to obtain a comprehensive evaluation. Accessibility & Authorization is identified by the ratio between the Accessibility- Authorization maturity level of each data multiplied by its related weight and the sum of

each Existent data’s weight. Finally, the result of the formula is the percentage of accessible and authorized data, defined as a subset of the Existent Data.

Level 0	Data are not accessible.	0
Level 1	Data are partly accessible, since the lacking necessary authorisation.	0,3
Level 2	Data access process is difficult and time consuming; it is enabled by the data owner who is an individual or a department.	0,7
Level 3	Data access is simply enabled by authorization given by the department or individual who owns and manages them.	0,9
Level 4	Data are quickly accessible, when needed, to everyone in the organization.	1

Table 29 Accessibility & Authorization assessment

$$\% \text{ Acc. \& Auth.} = \frac{\sum_{j=1}^M R_j * A_j}{\sum_{j=1}^M R_j} \text{ where } \left\{ \begin{array}{l} j = 1 \dots m \text{ Existent data for the decision making process} \\ R = \text{Relevance's maturity level} \\ A = \text{Accessibility - Authorization maturity level} \end{array} \right\}$$

Table 29 shows the scale adopted to this maturity model. As shown, it is assigned to the second level of maturity a more than sufficient score (70%) since the sentence describing this maturity level emphasises that even if the process of data accessibility and authorization could be time consuming, it enables the decision maker to directly access data. Moreover, it was taken the decision to assign 0,9 (90%) to the third maturity level since the sentences built to describe the levels, stress a slight difference between the third and the fourth one. This last choice recurs in almost all the dimensions introduced in the model.

### Timeliness

The last dimension required to evaluate data availability is data timeliness. As before, it is identified by the ratio between the Timeliness maturity level of each data multiplied by its related weight and the sum of each Existent data’s weight. The result of the formula is the percentage of timely data. Even in this case it was taken the decision to assign 90% to the third maturity level due to the sentence construction, that describes an almost complete maturity in data updating, to support the decision-making process under analysis.

Level 0	Data are not updated.	0
Level 1	Data are not regularly updated and suitable to support the specific decision-making process	0,3
Level 2	Data are occasionally current and updated to support the specific decision-making process, the time interval from data generation to processing does not fully meet the requirements.	0,6
Level 3	Data are enough current and updated to support the specific decision-making process in a proper way.	0,9
Level 4	Data are as current as possible being regularly updated; minimising the time delay from data generation and acquisition to utilization.	1

Table 30 Timeliness assessment

$$\% \text{ Timeliness} = \frac{\sum_{j=1}^M R_j * T_j}{\sum_{j=1}^M R_j} \text{ where } \left\{ \begin{array}{l} j = 1 \dots m \text{ Existent data for the decision making process} \\ R = \text{Relevance's maturity level} \\ T = \text{Timeliness maturity level} \end{array} \right\}$$

Finally, Data availability is assessed by combined these three dimensions:

$$\text{Data Availability (\%)} = \text{Existence} * \frac{\text{Accessibility \& Authorization} + \text{Timeliness}}{2}$$

Therefore, the percentage of available data is given starting from the percentage of Existent data; then from this value it is identified the subset of accessible, authorized and timely data. This subset is detected making an average between the percentage value obtained by Accessibility & Authorization and Timeliness dimensions.

### 6.3.4 Data quality

The assessment of each feature identified relatively the Data Quality dimension is done taking as object of analysis each single dimension specified in the list of *Interesting Data*. The features *Accuracy & Credibility* were merged together for the same reasons explained in the previous case, how even the *Presentation Quality and Consistency*. The Data Quality will be evaluated only for the Existent Data, the ones for which an evaluation different from Level 0

was given. Therefore, also in this case, the other dimensions defined to evaluate Data Quality – Accuracy & Credibility, Completeness, Presentation Quality & Consistency - are assessed starting from the Existent Data, defined as a sub set of the Interesting data. This because if data are not physically present in the company systems and database, their features evaluation could not be performed.

### Accuracy & Credibility

Level 0	Data are completely inaccurate and of doubtful credibility.	0
Level 1	Data contain several errors, mainly due to manual collection and manipulation. They come from uncertified sources.	0,2
Level 2	Data are partly correct and credible. They come from data sources that could be certified and trustworthy.	0,6
Level 3	Most data are correct. They all come from certified and trustworthy sources.	0,9
Level 4	Data are highly reliable: they are automatically collected from specialized sources, and experts regularly audit and check the correctness of the data.	1

Table 31 Accuracy & Credibility assessment

$$Accuracy\&Credib. = \frac{\sum_{j=1}^M R_j * AC_j}{\sum_{j=1}^M R_j} \text{ where } \left\{ \begin{array}{l} j = 1 \dots m \text{ Existent data for the decision making process} \\ R = \text{Relevance's maturity level} \\ AC = \text{Accuracy - credibility maturity level} \end{array} \right.$$

As happen for Accessibility & Authorization a single maturity model is adopted to evaluate both the Accuracy and the Credibility dimensions. This was the result of the case study implementation required to validate and test the model, that underlines a strong correlation between these two features. Indeed, on one side data coming from certificate and specialized sources, have a high probability to be accurate and free of error; on the other side if data are coming from sources not frequently monitored and authenticated the probability to notice inaccuracy among the company datasets is significant. Therefore, the Accuracy & Credibility dimension is identified by the ratio between the Accuracy & Credibility maturity level of each data multiplied by its related weight and the sum of each Existent data’s weight. The result of the formula is the percentage of accurate and credible data.

## Completeness

Level 0	Data are incomplete to take a specific decision.	0
Level 1	The minority part of data is considered complete to satisfy the needs of the specific decision-making process.	0,2
Level 2	Data are almost sufficiently complete to satisfy the needs of the specific decision-making process, nevertheless the presence of several missing values.	0,6
Level 3	Data are considered appropriately complete to supporting the specific decision-making process.	0,9
Level 4	Data are complete, any value is missing, having all required part of an entity's description.	1

Table 32 Completeness assessment

$$Completeness = \frac{\sum_{j=1}^M R_j * C_j}{\sum_{j=1}^M R_j} \text{ where } \left\{ \begin{array}{l} j = 1 \dots m \text{ Existent data for the decision making process} \\ R = \text{Relevance's maturity level} \\ C = \text{Completeness maturity level} \end{array} \right\}$$

Completeness is the second dimension assessed to evaluate the Data quality. It is identified by the ratio between the Completeness maturity level of each data multiplied by its related weight and the sum of each Existent data's weight. The result of the formula is the percentage of complete data. The third level of this maturity model describes an almost full capability of the company in obtaining complete data to support the decision-making process under analysis. Indeed, as happen in almost all the dimensions, the third maturity level is converted in a numeric value equal to 0,9. It means that the 90% of the data under assessment are considered complete by the user, properly supporting him within the decision-making process.

## Presentation Quality & Consistency

Similar to Accuracy & Credibility dimensions, a common maturity model is deployed to assess both the Presentation Quality and the Consistency features. Even in this case, it is identified by the ratio between the Presentation Quality & Consistency maturity level of each data multiplied by its related weight and the sum of each Existent data's weight. The result of the formula is the percentage of well- presented and consistent data.

Level 0	Data are not interpretable, and they don't have a consistent and concise representation.	0
Level 1	Data are difficult to understand causing some ambiguity in their interpretation and the format presents different meaning across time and space.	0,2
Level 2	On average data are comprehensible, interpretable, and partially consistent, presenting some discrepancies in the format across time and space.	0,5
Level 3	Data are concise, compact and clear represented not leaving much room for interpretation. They are consistent in almost their entirety, presenting discrepancies across time and space.	0,8
Level 4	Data are highly clear, understandable, compact and consistent over time and space (and after being processed).	1

Table 33 Presentation Quality & Consistency dimensions

$$P. \text{ Quality \& Cons.} = \frac{\sum_{j=1}^M R_j * PC_j}{\sum_{j=1}^M R_j} \text{ where } \begin{cases} j = 1 \dots m \text{ Existent data for the decision making process} \\ R = \text{Relevance's maturity level} \\ PC = \text{Presentation quality - consistency maturity level} \end{cases}$$

Therefore, the global percentage indicating the Data quality factor is given by the average among the Data Accuracy & Credibility, Completeness and Presentation Quality & Consistency.

$$\text{Data Quality (\%)} = \frac{\text{Accuracy\&Credibility} + \text{Completeness} + \text{P. Quality \& Cons.}}{3}$$

An average among the percentages obtained by the evaluations of each factor is made in order to achieve a comprehensive index for the Data quality macro factor. It was excluded the decision to aggregate these three dimensions by multiplying each other because in this study these features are considered independently by each other. It means for example that a data could be considered complete but not accurate and credible, or it could be consistent and well-presented but not complete. Hence, it is not possible to identify complete data as a subset of accurate and credible data; or it is not possible to define consistent and well-presented data as a subset of complete data. Therefore, the multiplication was considered a not appropriate approach.

### 6.3.5 DSS Performance

For this dimension, the maturity assessment is done taking as object of analysis the whole dataset available to the decision maker to take the specific decision. Except for the feature *Data Usage*, for which the evaluation is done taking as reference each single data considered existent.

#### Data Usage

Level 0	Data are not used to support the decision-making process.	0
Level 1	The minor part of data is used to support the decision-making process.	0,3
Level 2	On average data are used to support the decision-making process.	0,6
Level 3	The major part of data is used to support the decision-making process.	0,9
Level 4	Data are fully used to support the decision-making process.	1

Table 34 Data Usage assessment

$$Data\ Usage = \frac{\sum_{j=1}^M R_j * U_j}{\sum_{j=1}^M R_j} \text{ where } \left\{ \begin{array}{l} j = 1 \dots m \text{ Existent data for the decision making process} \\ R = \text{Relevance's maturity level} \\ U = \text{Completeness maturity level} \end{array} \right.$$

This last dimension aims to verify which is the subset of Existent data effectively deployed by the decision makers for the specific task. It is defined as the ratio between the Data Usage maturity level of each data multiplied by its related weight and the sum of each Existent data's weight. The result of the formula is the percentage of used data to support the specific decision-making process. To assess the percentage of used data it was required to develop a maturity model since, the list of data given as an input is presented at an aggregate level. Therefore, it is not possible for the user to affirm with a Boolean variable (0/1), if a particular data of the list is used or not to support the decision-making process under analysis. For example, assuming that the first data of the list is "Physical Products characteristics", it could happen that the user taking a particular decision, deploy only a part of the proposed data, being therefore represented by the first or second maturity level.

Data integration, processing and supporting infrastructure are the last three dimensions assessed. Their evaluation is made taking into consideration the whole dataset generated by the company, that could be deployed and exploited in order to give a support to the user

within a decision-making process. Therefore, from this point on the List of Interesting data taken as input and applied is not taken into consideration anymore. Therefore, any formula is required to assess each dimension and the percentage of each one is directly given by the maturity level identification.

## Data Integration

Level 0	Information System (IS) modules are not able at all to interface, allowing data integration.	0
Level 1	High probability to have interface and communication problems between different IS modules, in terms of technologies, processes and interactions between individuals.	0,3
Level 2	On average the IS modules are able to interface, addressing problems of data updating between different functional groups.	0,5
Level 3	The IS assures functional groups a good communication and interoperability.	0,9
Level 4	The IS of the overall organisation is scalable and flexible, allowing to achieve improved interoperability between functional groups, aligning communications	1

Table 35 Data Integration assessment

## Data Processing

Level 0	Data are not analysed.	0
Level 1	Data are partly analysed and presented in the form of indicators (e.g. KPIs).	0,3
Level 2	The majority of data is processed and presented in the form of significant indicators (e.g. KPIs), thanks to the aggregation and disaggregation.	0,6
Level 3	Structured and unstructured data are processed by advanced data analysis techniques and algorithms able to identify hidden patterns, supporting significantly the decision makers.	0,9
Level 4	Structured and unstructured data are processed by highly innovative and advanced algorithms, applying for example machine learning and artificial intelligence solutions, creating value added and supporting in an efficient way the decision makers.	1

Table 36 Data Processing assessment

## Supporting Infrastructure

Level 0	The underling infrastructure (hardware and software) is not able to analyse, support and process data.	0
Level 1	The underling infrastructure (hardware and software) is able to process big amount of data leading mainly to descriptive analysis on structured data.	0,3
Level 2	The underling infrastructure (hardware and software) is able to process big amount of data through averagely complex algorithm able to perform descriptive and predictive analysis on structured data. The infrastructure requires a significant timeframe in order to run the data analysis techniques.	0,6
Level 3	The underling infrastructure (hardware and software) is able to process big amount of data through descriptive and predictive analysis on structured and unstructured data, historical real time and advanced data visualization. For data analysis, the infrastructure requires a time frame considered appropriate and reasonable by the decision-maker.	0,9
Level 4	The underling infrastructure (hardware and software) is able to process big amount of structured and unstructured data supporting the most advanced and innovative data analysis techniques, performing perspective analysis. Data are processed on highly performing infrastructure, able to process the data on time and therefore providing answer to the decision makers in a short time frame.	1

Table 37 Supporting Infrastructure assessment

Therefore, the DSS performance value is computed as the average among Data Usage, Data Integration, Data Processing and Supporting Infrastructure.

$$DSS\ Performance\ (\%) = \frac{Integration + Processing + Supp.\ Infrastructure + Usage}{4}$$

As happens for the Data quality dimensions, it was decided to deploy an average instead than a multiplication since one factor does not exclude the other one. In this work, these four dimensions are viewed as four different aspects of the decision support system that could be evaluated and treated differently since from the conducted case studies there is no correlation evidence. However, it is not excluded the possible existence of correlation among these dimensions, to be studied better with future statistical analysis conducted on large-scale.

Finally, the Data Productivity Index (DPI2.0) is obtained by multiplying together the three macro factors. This choice, inspired by the OEE metrics, assumed that: the DSS performance is valuable only if working with high - quality data; and Data quality count only if data are affectively available to the user. This is a dropdown approach that conceptually assumes that DSS -Performance is a percentage of the Data Quality and this last one is a percentage of the Data availability.

This approach, as the OEE, has a geometric mean that punishes variability among components. For example, a company that owns a more stable index across the macro factors (e.g. 50% in each factor) will get a better global evaluation than a company performing extremely good in Data quality (e.g. 80%) but that has serious deficiencies at the DSS performance level (e.g. 20%).

$$\text{Data Productivity Index 2.0} = \text{Data Availability}(\%) * \text{Data Quality}(\%) * \text{DSS Performance}(\%)$$

# 7 Model Testing

The evaluation and consolidation of the model proposed was done using an observational methodology. The observational evaluation method was considered most suitable for our case, with the aim to develop and validate new theories by empirically observing the context analysed (Voss et al. 2002). Descriptive methodology was considered not comprehensive enough to evaluate innovative artifacts (Hevner, March, Park & Ram, 2004).

After having already identified the scenarios and the knowledge base to establish the model utility, mainly thanks to the previous studies, in this moment it was preferred to use the case study approach as research strategy. This decision was adopted since they could be defined as studies of precise events within real-life context, in particular the business one. Indeed, case studies are considered a preliminary research strategy suitable for research exploratory phase, supportive when the questions *why* and *how* are posed.

The case study application corresponds to the *deployment* phase of the model development, previously consisting in the model testing and finally in the application and standardization of the model. The Yin's methodology (2009) was used to ensure the validity of our research study. This section is going to present as the first item the design phase and as second one the real application of the model in the companies.

## 7.1 Context application

To design and formulate the model and consequently to compute the rate, the review of the existing literature on close topics was done, in order to develop an insightful proposal. Case study begins with the definition of the theme to be studied and development of the case studies design. In order to better define the topic of the model, it was possible to appeal to the components of research design drawn up by Yin (2009). The principal study question and its proposition concern how the productivity of data in the specific decision-making process could be measured in organisations. Therefore, the unit of analysis were chosen. The field of application considered are the *manufacturing and distribution companies* and the expert interviewed need to be people taking decision in the *Distribution planning Process*.

With the purpose to standardize and enhance the model applicability, it was considered necessary to restrict the field of application to one precise decision-making process. As first step, the decision to focus on a planning process was made, bearing in mind the significance of data for this kind of process. Data are essential requirements to reduce and respond in real-time to the occurrence of process variability faced in taking this kind of decisions. Indeed, the basis of a good planning process consists in a good data management, starting at data collection, organisation and storage, to come to the effective data exploitation. In particular, the model is applied in Distribution planning process, that is gaining more and more importance within organisations; in fact, distribution planners struggle in meeting the increased expectations. First of all, it is crucial to give a brief and generic definition of the distribution planning. It is a process referring to the systematic management of the flow of goods in the distribution network between sites. It is intended as the combination of transportation planning (outbound flows) and inventory planning and their subsequent optimization. Transportation planning concerns detailed knowledge of complete and outstanding orders from the company to warehouses and from the warehouses to the different possible customers. The inventory planning refers to the definition and management of stock levels at the stocking points.

In summary the nature of this Distribution planning process is highly data-intensive, deterministic and repetitive; for this reason, the distribution decisions leverage on information technology and optimization techniques to manage inventories and product flow thus meeting demand at minimum cost. This is possible resorting to Decision Support Systems.

Thanks to the use of case study, the authors are able to confirm the issue of relevance of the topic studied. Indeed, the experts interviewed supported our initial assumptions concerning the general necessity for organisations to monitor performances and therefore the necessity of really understanding and measuring their ability in data exploitation when taking decisions. Particularly, they were interested to obtain the evaluation achieved, showing their enthusiasm for the topic treated, both for its novelty and for the possible practical application of the instrument in the organisation. Indeed, companies fail in the assessment of their own capabilities in data exploitation, impeding in this way the identification of potential areas of improvement and implementation of coordinated measures.

## 7.2 Designing Case study

Due to the nature of our objective, a *multiple case study* has been conducted. Indeed, the evidence obtained from multiple cases is often considered more compelling and the overall study is therefore regarded as characterized by more robustness (Herriot & Firestone, 1983).

The first phase in designing case study consists in the definition and development of the theoretical model to validate. Therefore, the next important steps in design and data collection process are the case selection and the definition of specific measures. Each case of the sample should be carefully selected. While respecting the just presented application fields, other criteria were followed to select the sample. The starting point was considered the detection of the possible companies operating in manufacturing and distribution sectors. After the selection of the sector, other criteria to be met was considered the geographical proximity, with the aim to directly receive the information. Indeed, the best way identified to collect data from companies is to conduct a face-to-face interview, or if not possible performing a Skype call or a simple phone call. The companies selected are nearly all located in Northern Italy. The last step consisted in the verification of the organisations availability in participating to our study. This verification could be done performing a phone call or sending an email.

A further question to encounter has to do with the number of cases deemed necessary or sufficient for the study. Indeed, it is required to point out the necessity of reaching from 6 to 10 case studies arranged effectively to validate the model and to provide compelling support for the initial set of propositions on which the model is built. This considering the possibility that few cases (2 or 3) would be literal replications, predicting in this way similar results. This decision depends on the certainty you want to have about your multiple-case results.

Furthermore, it is important to take into consideration that if cases are in some way contradictory, the initial theoretical foundations must be revised and retested with a new set of cases. Redesign must take place before proceeding further with other case study. Exactly with this aim it were carried out the two pilot cases. Indeed, the case study design may sustain modifications in response to the possible discoveries during data collection. This

kind of revelations can be enormously important, leading to the alteration or adjustment of original theoretical framework presented. The loop procedure must be done without diminishing the rigor with which case studies are conducted. This kind of procedure and the redesign were considered an essential step to avoid accusation of distorting or ignoring the discovery, just not to change and reshape the original design proposed.

## 7.2.1 The Interview

Once the company expressed its willingness to participate, it was contacted by email. In case of face-to-face interview, with the aim to arrange it, a scheduled program of the meeting was sent. Whereas in case of Skype or Phone call, the email sent was different, since more details were provided in advance to the interviewee, to make the data collection simpler and to allow the interviewee to prepare about the topic treated. Moreover, also the Non-Disclosure Agreement (NDA) was sent as attachment in the email.

The duration of the interview could vary from one to two hours according to the company allotted time and it is structured in the following way:

1. Introduction to the thesis work and interview objectives
2. Overview of the company and focus on the distribution planning
3. Presentation and application of the model

Any interview was recorded under company consensus with the aim not to miss significant details and particulars provided by the experts.

To apply the model, an Excel spreadsheet was especially developed to gather information regarding the current performances of organisations in data exploitation. The document is a structured and guided instrument - presenting instructions and guidelines in all its parts - able to return a company evaluation, in the form of percentage. The Excel is mainly composed by six blocks. The *first part* presented an introduction and explanation of the tool proposed. Then in the *second part*, the general information about the company is requested to the interviewee, to give a contextualisation of the single case study. Concerning this part an unstructured approach has been adopted, allowing in this way the interviewee to freely talk about their company and the relative distribution planning process. In case all the

necessary information about the company was not collected, the specific questions were directed to the respondent. Particularly the general information needed concern the following arguments:

- The industry/sector of the respondent's company
- Background information about the context in which the company operates indicating if possible, the objectives, product lines, markets and customers
- The size of the company, in terms of turnover and employees
- The company positioning in the value chain and a general description of the relationship with suppliers and customers
- The role of the interviewee in the company
- A synthetic description of the Distribution planning process
- Indication of the familiarity with the Distribution planning process

In the *third block* of the Excel spread sheet, it was requested to the interviewee to evaluate the Relevance of a list of data identified as *Input of the model* for the decision-making process under analysis.

### Input of the model

Considering the distribution planning focus adopted by the model, it has been necessary to specifically develop a baseline list of data considered interesting for the precise process. This list was created with the aim to improve the applicability of the model; for this reason, the most suitable level of data aggregation was chosen in order to ensure the model application in every sector. The development of this list was done in accordance with expert opinions in the field, following a two-step approach. Firstly, the baseline list was built interviewing and requesting advices and opinions of experienced people and specialists of the sector. Therefore, through the pilot cases some modifications and clarifications of the list items were done thanks to the opinions of the respondents.

To make the evaluation more guided and intuitive, it was considered necessary to group the list items according to the possible area of belonging.

<p style="text-align: center;">Descriptive data (product-network)</p>	<p>Product descriptive characteristics (e.g. categories, line, size, colour, lifecycle, depreciation) and Product physical characteristics (e.g. packaging, weight)</p> <p>Product characteristics affecting lead time (e.g. product specific customs requirements)</p> <p>Distribution network characteristics and typology (e.g. nodes, capacity) and relative development (e.g. new openings, new markets)</p> <p>Transportation times per segment and possible frequency (e.g. flights planning)</p> <p>Total transportation times per segment (included administrative steps, customs, etc) and Total lead time door-to-door, with relative variability</p> <p>Transportation cost per segment (considering the different transportation modes)</p>
<p style="text-align: center;">Inventory and flow data</p>	<p>Warehouse inventory level (e.g. available, booked, saleable, reserved, damaged)</p> <p>Pipeline (transit) inventory level</p> <p>Incoming orders, backlog or outstanding</p>
<p style="text-align: center;">Market Data</p>	<p>Customer demand forecast (punctual, mean and dispersion)</p> <p>SLA (e.g. target service level, target availability rate, target stock level)</p> <p>Working calendar (holidays)</p> <p>Calendar promotional events</p> <p>Possible substitute products</p>

Table 38 Model input - Interesting Data for the Distribution Planning process

The *descriptive data* refer both to the product and to the network. Nowadays these data are considered crucial, since organisations' distribution strategies are determined by product characteristics. For this reason, the data identified in the list generally refer to the product characteristics in all its different aspects: properly descriptive, physical and the ones affecting the lead time. Furthermore, to have detailed information about the network improves the reliability and performances of distribution planning, pointing out the importance of these kinds of data. The list items to which it is getting reference are linked to the general characteristics of the network, typology, development, transportation times and costs.

The *inventory and flow data* refer principally to the inventory management. These are key data for the distribution planning, to be taken into account to ensure to the customers a satisfactory service level, avoiding distribution problems and promptly filling orders. Inventory levels usually include multiple types of stock. Warehouse inventory level refers to a set of different stock typologies present in the warehouse. While pipeline inventory refers to any inventory that are in the company shipping chain and it hasn't yet reached the ultimate destination. The understanding of inventory level allows to more accurately track how much cash is tied up in inventory and overheads. Moreover, data related to incoming order, outstanding orders and order backlog are considered significant in addition to these data.

Instead the *market data* identified are those kinds of data that you have to consider in order to align the distribution planning with the market and to increase the organisation's responsiveness. The data identified related to this area are the ones concerning the customers (demand and service level), the calendar (holidays), promotional events (sell-in and sell-out) and the competitors.

The identified list of data must be evaluated independently from the data really owned by the organisation but looking only for the potential importance of the specific data under analysis for the distribution planning process.

The *fourth part* aims at gaining specific information related respectively to the Data Availability and Data Quality, through the support of the maturity models developed for

each feature identified for the two macro dimensions. The evaluation of each feature must be done for each data of the list; marking with an X the sentence that better identifies its current state. While the *fifth part* was directed to the assessment of the DSS performances, considering as object of the evaluation the entire dataset. Also, in this case, the respondent should choose for each feature, the sentence that better describes the company situation.

In the *last part* all the results obtained by the company relatively to the three different macro dimensions are going to be showed, together with the overall results obtained through the overall DPI 2.0 computation.

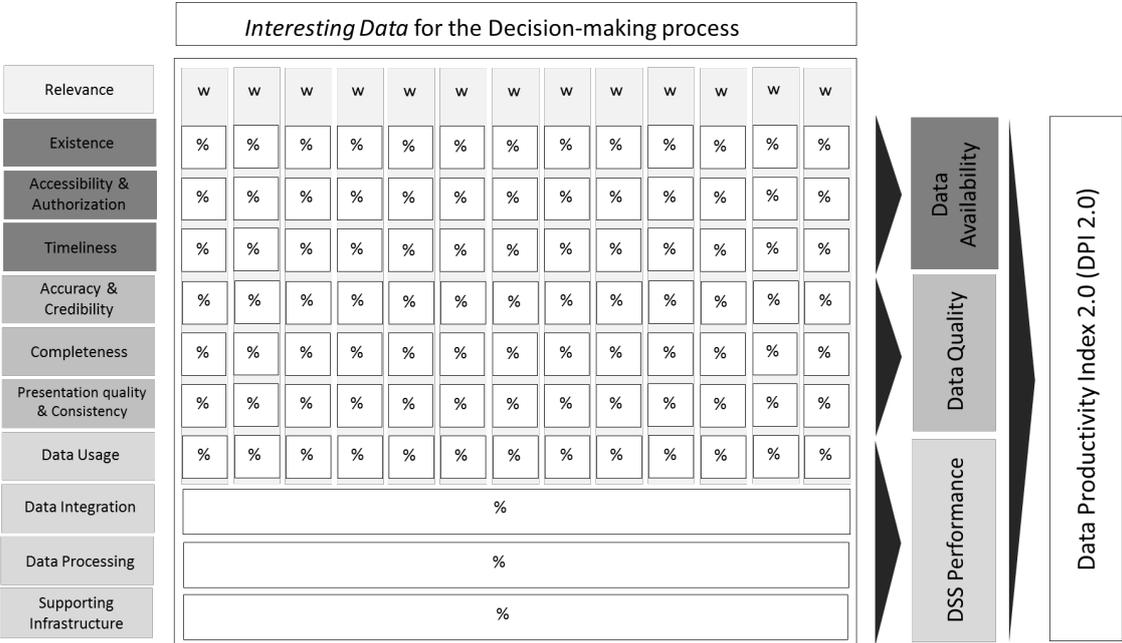


Figure 53 Excel Dashboard Computation

To sum up, *Figure 53* aim is to exhibit how the Excel Dashboard calculates the Data Productivity Index 2.0. Indeed, starting from the thirteen data provided as input of the model the percentage representing the features of Data Availability and Data Quality could be computed. Then aggregating the evaluation given to each data’s features the comprehensive evaluation of Data Existence, Accessibility & Authorization, Timeliness, Accuracy & Credibility, Completeness, Presentation Quality & Consistency are obtained. Finally aggregating the related features, the percentage of Data Availability and Data Quality is computed. Instead, the features related to the DSS-performance are calculated regardless

the Interesting list. Finally, aggregating the three macro factors the Data Productivity Index 2.0 is achieved.

## 7.2.2 Training and Preparation

Case study is a complex approach due to the lack of routine formula, therefore requiring training and preparation in order to be performed with high-quality. In order to be prepared for the interviews, it is fundamental to ask the proper questions, being adaptive and flexible. Indeed, interviews have been made across companies operating in different sectors, requiring the capability to adapt questions to each specific case. This capability was crucial especially in the first part of the interview, where information about company context and its distribution planning process are required. Moreover, even in the model implementation, where interviewees are required to complete the Excel Dashboard, the ability to make proper examples as much as possible related to the specific company sector has been decisive. This is of the utmost importance to exactly convey the questions and consequently to obtain more objective and precise evaluations. Interviewers must be prepared to deal with multiple cases, properly listening and interpreting all the information received.

## 7.2.3 Case Study Protocol

In order to perform multiple case study a Protocol is defined. This is a tool including all the procedure and generals' rules that must be followed while performing the interviews. Its definition increases the reliability of the case studies, becoming a guidance during the data collection.

The Protocol defined to carry out the case studies within this work, it is composed by five blocks:

**A. Definition of the project and case studies objectives:**

- The core project objective is the identification of a metrics able to measure how data support users during a specific decision-making process. Within the case studies, the model is tested to analyse if data support the decisions related to the Distribution Planning process.

- The cases studies aim is to verify and validate the model construction, testing the model applicability.
- B.** Case studies must be made by face -to-face, phone or skype call interviews. During the session data and information must be collected taking note and only in case an authorization is given - recording the whole procedure.
- C.** The interview must be structured as follows:
- 15 minutes to explain the general objective of the study and of the interview. This step is done with the support of a power point presentation.
  - 15 minutes to gain general information about the company through open questions, that could differ from one case to case.
  - 15 minutes to obtain information about the structure of the Distribution Planning process of the company. This is made by open questions that must be adapted to the specific company context.
  - 45 minutes to fill in the Excel Dashboard. The role of the interviewer in this part is to explain each block of the Excel Dashboard providing help and support during the compilation. Practical example should be made in order to help the interviewee to properly answer to the questions. Lastly the result achieved by the model proposed must be shown, giving the related explanation and motivation.

The time available to each part of the interview could changes according to the specific company availability.

- D.** Reporting case studies – the information recorded must be transcribed. The recorded information could be integrated with available sources (e.g. company official website or academic articles) but always quoting data sources and bibliography.
- E.** Evaluating case studies – the evaluation must be driven by the literature studied to develop the model.

## 7.2.4 Screening case study nominations

The companies interviewed, employed as cases studies, were selected following a defined process.

Firstly, a wide range of suitable companies have been identified among manufacturing and distribution companies. Concerning the manufacturing ones, a first screening was made excluding all that companies not managing directly the distribution processes; for instance, because the process is totally handled by third parties. Secondly a further screening was made. The companies which don't own structured and data monitored processes were excluded, being considered not suitable to the test the model proposed. The final sample of companies was selected among the suitable ones, selecting those companies whereby a direct contact with supply chain or logistic manager was available. Moreover, only the companies effectively demonstrating their interest in our study joined the research, giving their contribution.

The resulting sample of this selection process encompasses the 8 companies shown in *Table39*.

Company	Sector	Turnover (€)	Employee	Interviewee
<b>A</b>	Home care Retailer	300 million	1700	Supply Chain Manager
<b>B</b>	Pharmaceutical	800 million	2000	Logistics & Warehouse Managers
<b>C</b>	Home and Personal care	500 million	1000	Logistic & Customer service Manager
<b>D</b>	Mass Distribution	450 million	500	Integration Logistics process Manger
<b>E</b>	Mass Distribution	7700 million	23000	Head of Organisation department
<b>F</b>	Freight and logistics	16 million	15	Logistics Buyer & Administrative officer
<b>G</b>	Freight and logistics	220 million	350	Logistics & Engineering Manager
<b>H</b>	Fashion	280 million	1000	Distribution Planner

*Table 39 Company sample table*

## 7.2.5 Pilot Case studies

Among the eight companies selected, the first two interviews were deployed as Pilot case studies. Indeed, these two first cases have been performed in order to firstly improve the data collection process and particularly to refine the model, providing some conceptual clarification. Pilot cases are essential to understand if the model construction is appropriate and easily applicable. Therefore, these two cases have enabled some model variations leading to the final model construction.

The selection of the two companies deployed as pilot cases was driven by two main reasons. Firstly, the companies' locations were convenient and easily reachable in terms of geographical location; and secondly direct contacts with the heads of the Supply chain were available making the meetings set up easy.

The Pilot cases leads to the construction of a second and final version of the model: in the following paragraphs the term *Model Version 1.0* refers to the model before the pilot cases variations, while the term *Model Version 2.0* refers to the second and final model version.

## 7.3 Case study

### 7.3.1 CASE A – 1<sup>st</sup> Pilot Case study

Company A is an Italian retailer dealing mainly with home care products. Its turnover is around 300 million € per year achieved thanks to 1700 employees. The company develops its business through a network of both private and franchising shops, located across Italy. Globally, the company counts around 500 shops split equally by private and franchising. The products sold are purchased by two different supplier categories: the first one, accounting for the 60%-65% of the total, are offshore suppliers mostly coming from China, providing private label products. The second category, accounting for the remaining part 35%-40%, are Italian suppliers that supply goods of others multiple brands. The company being a retailer, does not have any production activities; their products are totally provided by suppliers delivering the required goods directly to the company warehouses.

The distribution process of the company is characterized by the presence of 5 “Smart warehouses” and 7 “Lung warehouses”. From the Smart warehouses the products are taken and delivered directly to the shops, while the Lung warehouses are simply used to keep the and stored goods when the Smart warehouses capacity is overtaken. The products stored in the lung warehouse are moved in the smart one when required. The reorder process made by the shops could be simply described as follows: an assortment of around 9000 products is available to each company shop; when one of these makes an order, the delivery is satisfied only if products are available in the Smart warehouses. Indeed, these are the real company warehouses owing an appropriate information system to efficiently handle the goods.

This first pilot case was performed by interviewing the supply chain manager of the company through a face-to-face interview of about two hours. The interviewee has a high familiarity with the whole company distribution planning process since he is the one in charge of the transportation and warehousing optimization. It is important to notice that the company is in a transitional phase, going towards a higher efficiency and digitalization. Therefore, the Data Productivity Index 2.0 (1,7%), obtained by the multiplication among the three macro

factors, indicates that the overall current state of the company is not satisfactory. Indeed, as also underlined by the interviewee, the company supply chain presents many inefficiencies, due also to the lack of data generation, collection and maintenance.

It is important to highlight that the results shown in *Table 40* are obtained by using the first version of the model.

Data Quality	Data Availability	DSS-Performance
13,5%	37,6%	33,1%
<i>Data Productivity Index 2.0 = 1,7%</i>		

*Table 40 CASE A Results - Model Version 1.0*

As a Pilot case, this first interview, was supposed to be useful to revise and adjust the model presented. But due to the level of company’s digitalization, its current ability in collect data and having automated processes, the company under analysis was not considered supportive, not providing a valuable feedback. The only worthy suggestion was a first indication that one data of the Interesting list, provided as input of the model, was not completely clear and understandable. The discussed data refer to the theoretical concept stating that distribution processes are different according to the way companies fulfil demand and thus to the types of goods handled. For instance, the distribution process changes if a company operates with a make to stock (MTS) approach or with a make to order (MTO).

Despite this it was taken the decision to not modify yet the Interesting list, but to perform a second pilot case presenting the same model, to verify if this specific data was not clear neither to the second company.

Moreover, a second pilot was considered absolutely essential in order to obtain more valuable and helpful opinions to revise and improve the model introduced.

### 7.3.2 CASE B – 2<sup>nd</sup> Pilot Case study

Company B is a multinational company operating worldwide in the pharma sector. Its business portfolio goes from pharmaceuticals, consumer health, animal health to Crop

Science, addressing doctors, veterinarians, farmers and consumers. In Italy it is composed by different corporations; the interview was made only to that part managing all the activities related to the human and animal health. The cooperation interviewed does not have any production activities, but it deals only with the goods distribution. Its turnover counts around 800 million € with 2000 employee. Its network is composed by wholesalers, pharmacies, nursing home, pet shop and mass market. This context is particularly complex due to two main reason: the high range of products available and the short lead time required by customers especially in case of lifesaver products.

The distribution planning process is structured as follows: the customer service manages all the orders received by the customers which are forwarded to the warehouses through SAP application. The company deploys two warehouses that, even if are not owned directly by it, are totally managed internally. Indeed, the task of the supply chain actors is to manage the warehouses' interface dealing with the trend of the stock, possible difference between the real stock in the suppliers' warehouses and the stock level traced within the company systems. Then also the transportation is performed by third parties but even in this case are totally managed internally by the company through information system interfaces. Therefore, the key activities performed within the distribution planning process are the management of the interfaces with warehouses and transporters.

The second Pilot Case was performed interviewing the Distribution & Logistic Manager and the Warehouse Manager; two individuals owing a high familiarity with the company distribution process and the related data. The interview was conducted through a face-to-face meeting with a duration of about two hours.

The results obtained by the implementation of the first version of the model are shown in *Table 41*.

Data Quality	Data Availability	DSS-Performance
56,0%	63,7%	65,2%
<i>Data Productivity Index 2.0 = 23,2%</i>		

Table 41 CASE B Results - Model Version 1.0

This Pilot case was very valuable to realize and understand the critical issues of the model. Consequently, after this second test, multiple variations have been made to the first model presented, leading to the construction of a second and final version of the model. Below are deeply discussed the lessons learnt from this second case study.

*Dimensions aggregations:* When developing and constructing the theoretical model a high number of dimensions was introduced. With a first study of each dimensions' definition and concept it was stressed a strictly correlation between some definitions. In detail, this was noticed for the concepts of Authorization & Accessibility, Accuracy & Credibility and finally Presentation quality & Consistency. Regardless this, into Model Version 1.0 each dimension was evaluated individually, through the development of a maturity model for each one. This was made to understand if these dimensions were effectively strictly correlated as suggested by the theory with a first model application. As expected, the test of the model suggests that single maturity models, merging together the concepts of Authorization & Accessibility, Accuracy & Credibility and lastly Presentation Quality & Consistency were enough comprehensive and sufficient to evaluate these features. Therefore, three common maturity model were developed to assess:

- Authorization & Accessibility: these two concepts were considered subsequent phases and therefore strictly interconnected. Indeed firstly, the data user has to verify if the process to access data is easy and quick, then after data are reached, the user must own the authorization to effectively deploy them.
- Accuracy & Credibility: it was noticed the high correlations between these two definitions. In fact, if data are coming from specialized and authenticated sources, they have a higher probability of being accurate and free of error. On the other side, if data are coming from uncertified and not monitored sources (e.g. manual collection), they have a high likelihood of having been manipulated, presenting several inaccuracies.
- Presentation Quality & Consistency: these two concepts were merged together since consistency could be viewed as a prerequisite to obtain well-presented data, according to the definition given of presentation quality. Better explaining, data are defined consistent if their format is kept constant across time and space, for example

automatically updating changes in code or data format. Instead, Presentation quality could be viewed as a broader concept, which aims to verify if data are easily understandable and interpretable. In the application of these two distinct theoretical concepts it was clear according to interviewees the assessment of these two features together is more practical and realistic in business context.

*Maturity model development:* this Pilot case leads to the reformulation of the sentences of almost all maturity models. These variations turn out to be crucial and extremely required in order to have a model applicable in reality. Indeed, the maturity models presented in the Version 1.0 of the framework, were found to be too academic and strict, being not meaningful to represent the current state of companies. Therefore, the maturity model of each dimension was revised, implementing the suggestions and the perceptions received during this pilot case, making the model more realistic.

*Maturity level score:* variations to the weight given to each maturity level were performed. Indeed, in the first version of the model it was adopted a unique numeric scale to each feature of the model, to go from a qualitative approach to a quantitative one. The scale adopted into the Model Version 1.0 is shown in Table 42.

Level 0	0
Level 1	0,3
Level 2	0,6
Level 3	0,8
Level 4	1

*Table 42 Model version 1.0 – Maturity level scores*

This scale turns out to be not explicative to evaluate each feature of the model. This happens especially after the maturity models reconstruction where the sentences representing each maturity level have been reformulated. Therefore, it was taken the decision to personalize the scale, adopting a different one to each model dimension. As deeply explained in the *Model Building* chapter, each dimension was identified by the most appropriate scale, assigning a numeric value that goes always from 0 to 1 to each maturity level. These numbers are chosen arbitrary selecting the most representative value to quantify each sentence. For instance, for the maturity models related to the data timeliness it was decided

to assign to the third level a numeric value equal to 0,9. It means that data represented by the third maturity level could be considered for the 90% timely, being regularly updated and therefore appropriate to support the Distribution planning processes.

*Model's Input:* as mentioned in the previous case, the first version of the Interesting list provided as input of the model encompassed an additional data related to the type of goods handled. Even this second Pilot case stressed that companies do not clear understand the reference dataset, having difficulty in linking this data of the list with generated or collectibles ones. Consequently, without a clear insight of the data under analysis, the evaluation becomes tough and complex for the interviewees. Therefore, it was taken the final decision to remove data related to the types of goods sold from the interesting list.

Except for this data, the Interesting list was considered by both Company A and B understandable and comprehensive, encompassing all the collectable data that may support the distribution planning process.

*Excel Dashboard presentation:* in the Model Version 1.0 the excel dashboard was made by 7 different blocks, the data quality and the data availability dimensions were evaluated separately. This choice was found to be not comfortable for the interviewee, requiring a considerable amount of time to fulfil the entire questionnaire. Therefore, the two blocks related to the data availability and data quality dimensions were merged into a unique block, making the Excel Dashboard more manageable and easier to be complied.

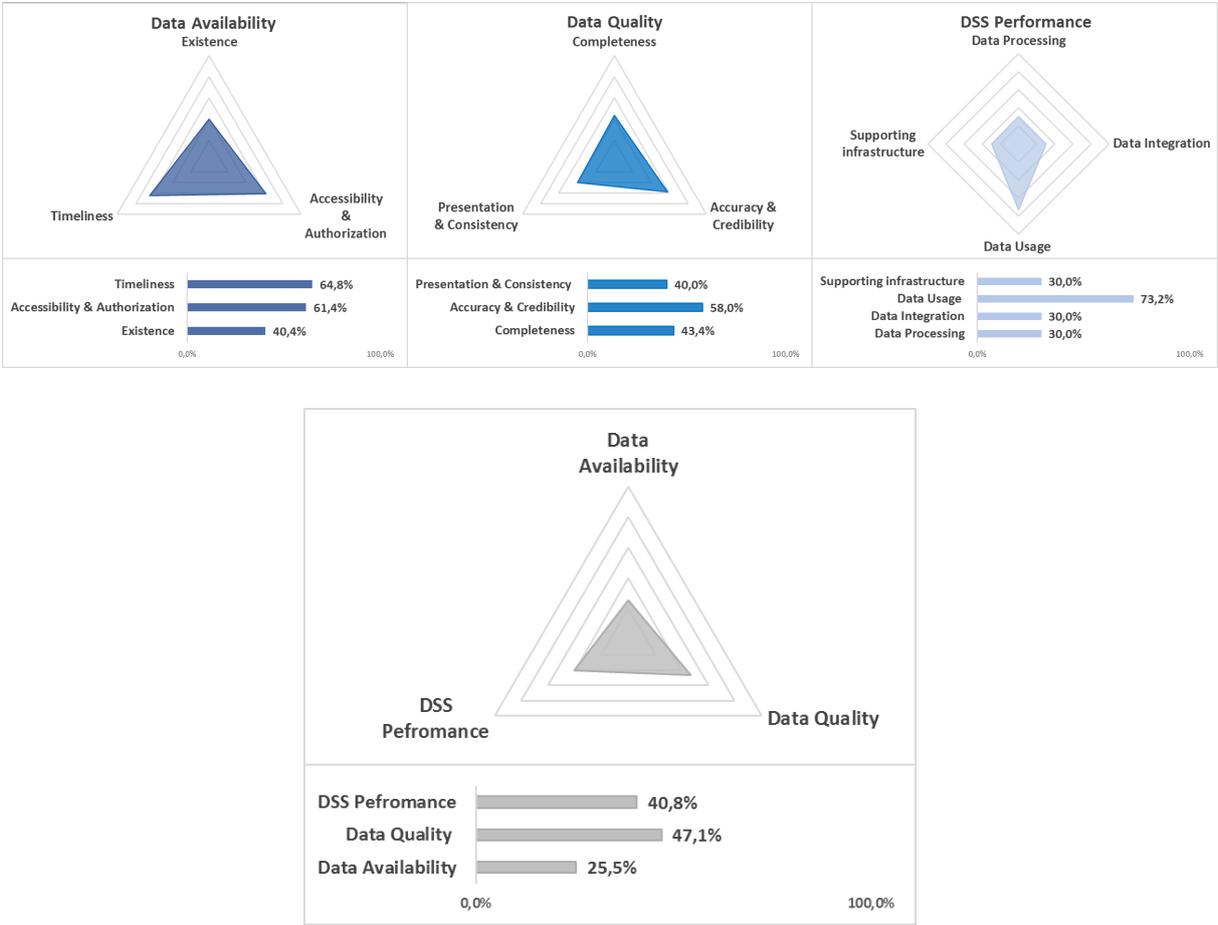
After the implementations of the changes driven by the interview, questions and suggestion received by the Company B, the final model version has been developed (Version 2.0).

The interviews to the Company A and B were recorded under consensus. This enabled to retrieve a comprehensive amount of information concerning the companies, the overall distribution process, the state of each data of the interesting list, and the information technology infrastructure. Therefore, it was possible to reconstruct the assessment in order to apply the revised and update model (Model Version 2.0). The re-assessment of the company's current state according to the modifications done was made starting from the evaluation obtained in the Model Version 1.0.

Next the results of Company A and Company B obtained deploying the Model Version 2.0 are discussed.

### Company A – Model Version 2.0

Testing again the model for the Company A the results obtained are shown in *Figure 54*. Nevertheless, the overall index increased going from 1,7% to 4,9 % this result is still considered not satisfactory since the low level of process automation and digitalisation deployed in the organisation. Indeed, only now the company is becoming aware about the importance of collecting data and the possible value retrievable from them.



**Data Productivity Index = 4,9%**

Figure 54 CASE A Results - Model Version 2.0

Data Availability is equal to 25,5%; this low result is mainly due the fact that even if the company owns a discrete data accessibility and timeliness, it generates only a small part of

the potential data that could support the distribution planning process (data existence equal to 40,4%). While *Data Quality* results equal to 47,1%; generally, the data used in the distribution planning decisions are not considered sufficiently complete and well-presented to support the decisions, whereas the dataset accuracy was on average evaluated sufficient. Lastly concerning the *DSS Performance* (40,8%), if on one hand the company ability to process, analyses, efficiently stores and exchanges data is quite limited; on the other hand, the data collected and generated are actually deployed to support the distribution planning process.

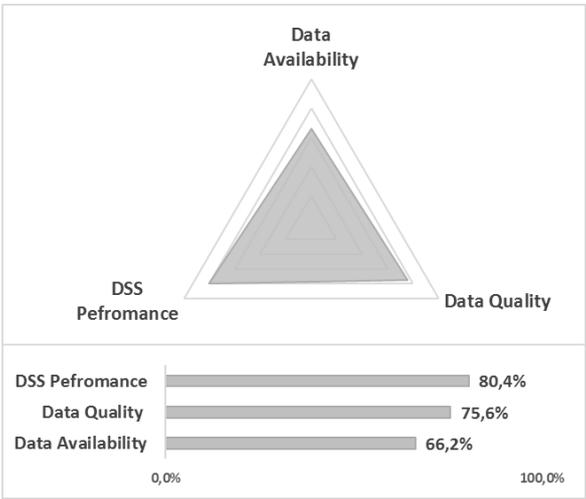
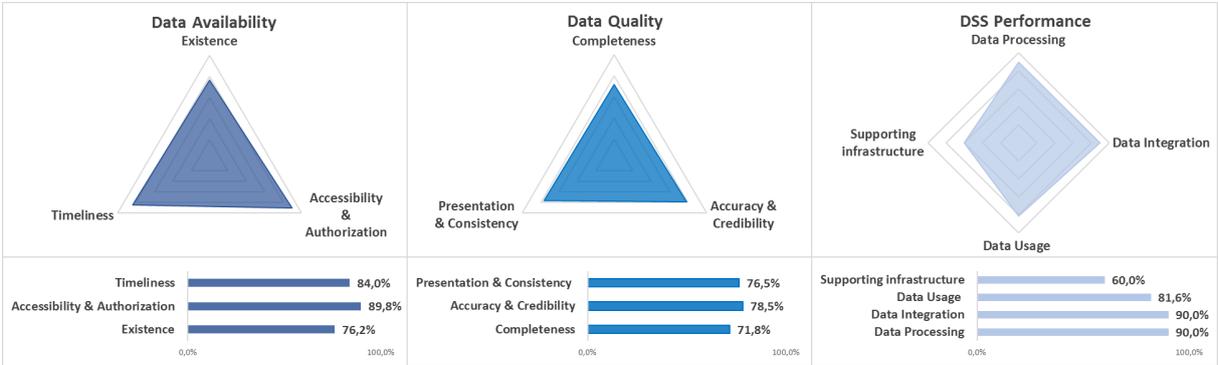
Therefore, globally the company has multiple inefficiency presenting significant lack in the data management field.

## Company B - Model Version 2.0

Some remarkable notes are driven by the evaluation discussion of each data. Firstly, concerning the relevance assessment it is important to notice that Company B operates with a high products variety requiring different management. For example, when judging the transportation time on one side, it becomes highly relevant for some type of products such as lifesaver requiring the score 5; and on the other side it is averagely relevant for product like pet care demanding the score 3. Therefore, it was decided by the interviewee to assign an average between the two scores.

*Data Availability* factor is equal to 66,2%. While the company well-performed in term of data timeliness and Authorization & Accessibility, the existence percentage, equal to 76,2%, highly punishes the overall value. The reason behind the data Existence value are found in some data that if one hand it is considered relevant to take the distribution planning decisions are not properly generated and collected by the company. Particularly for the data about the Warehouse inventory level and the Incoming order, backlog or outstanding (Relevance given 5), the respondents expressed the company difficulty in collecting this data, considered potentially helpful in the optimization of the distribution planning. Moreover, other two data – working calendar, calendar promotional events - were not collected at all, since not considered significant for the context in which the company operates, not affecting the planning. The company has good and balanced performance in all

the data quality components, underlying that almost all the data collected by the company are accurate, credible, complete, consistent and well-presented. The resulting *Data Quality* factor is equal to 75,6%.




---

***Data Productivity Index 2.0 = 40,3 %***

---

*Figure 55 CASE B Results - Model Version 2.0*

The *DSS Performance* dimension that mostly affects negatively this factor evaluation (80,4%) is the performance of supporting infrastructure. Indeed, the company affirms that its infrastructure is able to process only structured data with simple analyses. The good results obtained in this last factor are due to the high percentage of data used to undertake the distribution planning decisions and to the high level of data integration. Indeed, data are shared and exchange across the organization through SAP application that enables the communication among the different company databases.

Thanks to the changes adopted in the model, the company obtained better results in all the index components in comparison to ones achieved with Model Version 1.0. Therefore, the Data Productivity Index 2.0 results to be equal to 40,3%.

The results obtained by Company B deploying the second and final version of the model are shown in *Figure 55*.

### 7.3.3 CASE C

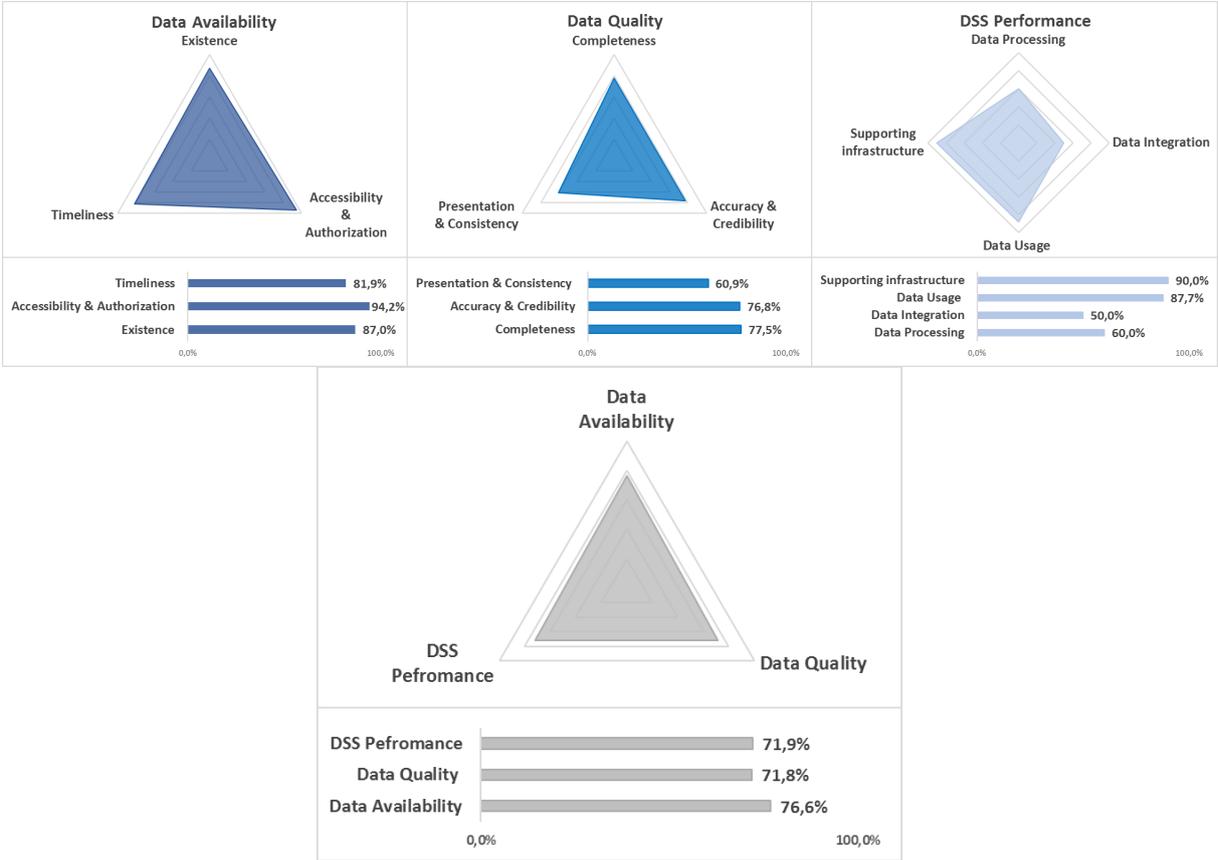
Company C is a chemical holding company principally offering home and personal care products. The group is leader in the personal, beauty and home care markets, characterized by a history of positive and dynamic growth and constant acquisitions, with a turnover of 500 million € and approximately 100 employees. The product lines offered are numerous: household-cleaning products, personal cleaning products, dermo cosmetics, perfumery, air fresheners, insecticides, hair care products, sunscreen lotions, children's product lines and tampons. The company operates at international level serving several markets, such as the mass-market, drugstores, perfumeries, professional channels, etc.

The organisation is well structured and deals with planning, production, new product development, storage and distribution. Particularly concerning the production, it is predominantly done in-house; therefore, the suppliers are mainly of raw materials. For one-half of the production, they handle also the manufacturing of dedicated packaging. Thus, their production starts from the product, to the cap and the bottle. Furthermore, they rely on external providers – about a dozen - that have to deal with the exhibitors packaging and/or special packaging. The customers' distribution channels are diversified:

- Drugstores, about 3000 customers
- Professional channel (hairdressing salons), about 2000 customers
- Organised large-scale distribution, about 1000 customers
- Perfumeries, about 1000 customers
- Export, about 40 countries with approximately 70 customers

The management of the distribution process is completely in-house. For the distribution of the 30% of the real production the company makes use of third-party logistic service

providers (3PLs). The strategic policies for distribution planning are decided by the logistics function basing on the inputs provided by the sales department. These inputs are the specific characteristics of the product, such as the total lead time, from the reception of the order to final delivery to customer. After having received the forecast by the demand planning and the production, the main task of the logistics function is to organise the distribution with national and regional partners (structured companies in transportation field). As concerns the product exportation, the logistics function defines the timing to the customers and organizes the distribution strategy with the foreign branches and/or the export direction.



**Data Productivity Index 2.0 = 39,6%**

Figure 56 CASE C Results

Subsequently, the results obtained from the coding of the phone interview with the logistic & customer service manager (i.e. the person in charge for the supervision of the logistic team and for the distribution planning decisions) are going to be provided and better

analysed. The interview was conducted by two analysts and had a duration of about one hour.

Presenting the relevance evaluation given to data, it is possible to immediately notice that the overall evaluation given on average is quite high. This means that the list items presented are deemed significant from their perspective in this decision-making process. The only data assessed as less important concerns the substitute products in the market, since these data do not affect the decisions in this field. Moving on to the data availability and data quality assessment part, firstly it is important to point out that the organisation owns all the data presented in the list, at different level of detail. The data are broadly accessible and timely, while the levels of accuracy, completeness and presentation quality are not considered satisfactory by the interviewee. Concerning the DSS performance evaluation, the company is well supported by the underlying infrastructure (hardware and software), while the level of data integration across the organisation and data processing are assessed as medium. Moreover, the data effectively used by the organisation when taking distribution planning decisions are few.

Concerning the general evaluations given, from the relevance to the data availability and data quality assessment it is important to specify that for their reality, it was considered necessary to take an average when evaluating each feature, considering the whole range of product lines. For instance, in case of professional products – representing about the 10% of the turnover – generally the evaluation would have been lower, just think to the manual handling of orders. In the company's view is not so crucial, in order to take decisions, to make effort in enhancing the quality of these kinds of data, considered the weight they have on the overall turnover. Indeed, if the respondent had to evaluate only core products, the evaluation given would have been different and of course higher.

The resulting evaluation in all the three dimensions are similar –Data availability (76,6%), Data Quality (71,8%), DSS Performance (71,9%) – highlighting a good level achieved for each one, and consequently a good evaluation obtained in the overall indicator of DPI 2.0 (39,6%).

### 7.3.4 CASE D

Company D is consortium of consumers' cooperatives in Italy. The system is composed by a national consortium having the central function, performing commercial activities (food and non-food), marketing and communication, on behalf of the cooperatives belonging to the system. The whole national consortium develops overall turnover of over 12 billion€, covering a market share in the Italian grocery sector of about the 17%. In 2006, the food and non-food businesses gave birth to two divisions. With the aim to promote the optimisation of logistics processes, for the food division 3 different departments were created.

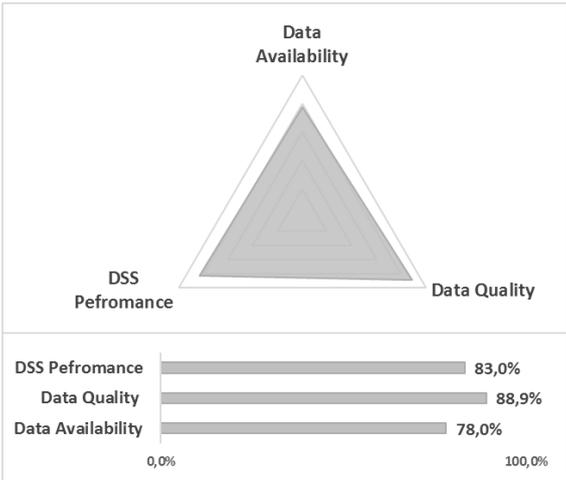
On the other side, concerning the non-food, the turnover is of 450 million € with 500 employees. The non-food logistics network is based on two central warehouses in Northern Italy. The first one is dedicated to seasonal and bulky products (e.g. Christmas tree, gardening supplies), while the second one to continuous products in assortment for the whole year. The delivery to the points of sales (PoS) is both direct and made through transit points, consisting of the food warehouses of the group. The two warehouses are aimed to serve and reach each week about 400-500 PoS – hypermarkets and supermarket - thus distributing throughout Italy.

The following results have been obtained from the interview with the integration logistics process manager, which deals with decision in the distribution field. The interview was conducted by phone by two analysts and had a duration of about one hour and a half. The analysis of the evidences collected is organised following the Excel dashboard structure.

Firstly, the part related to the relevance assessment was compiled by the respondent, giving very high scores to each data of the proposed list, from the descriptive data, to the flow and market data. This except for the substitute products, since in his view this kind of data was not considered significant in taking distribution decisions. It is important to highlight, according to the interviewee, the importance of product physical characteristics (e.g. size) in this context in order to take distribution planning decisions.

Moving over to data availability and quality assessment, as first point it is possible to analyse the evaluation given to the data existence. The organisation owns all the data listed, apart

from a partial collection of the inventory level at the PoS and the information related to substitute products in the market. Concerning the descriptive and flow data, the performance in the data availability and data quality are very high and satisfactory according to the respondent. While the data related to the market present some problems concerning the accuracy, completeness and the presentation quality. Furthermore, these types of data are not widely used when taking distribution planning decisions. As concerns the DSS performance dimension, the organisation uses software applications internally developed and managed, not purchased on the market. Consequently, the underlying infrastructure (hardware and software) is really supporting, ensuring in this way a good flexibility and ability to quickly process data, how even a good grade of interoperability across the organisation and a satisfactory level of data integration. The data are analysed and elaborated in different way, in order to obtain significant indicators.



**Data Productivity Index 2.0 = 57,5%**

Figure 57 CASE D Results

Therefore, considering the high level achieved in the Data Quality (88,9%) and DSS performance (83,0%) and the good result in the Data Availability (78,0%), the resulting overall DPI 2.0 is equal to 57,5%.

### 7.3.5 CASE E

Company E is one of the largest Italian chains in the sector of the organized large-scale distribution, operating in a network of more than 150 superstores and supermarkets mainly located in northern Italy. The company promotes product innovation, environmental preservation and consumer protection, with high quality and customer satisfaction objectives. The group is endowed with a wide selection of private label products and by fresh products, with an overall number of employees equal to 23000 and with a turnover of about 8 billion €. The suppliers are of different sizes, from the big pasta producer to the small wine grower) and the relationship with them are managed through dedicated buyers. The relationships with clients are direct, thanks to the large number of contact points.

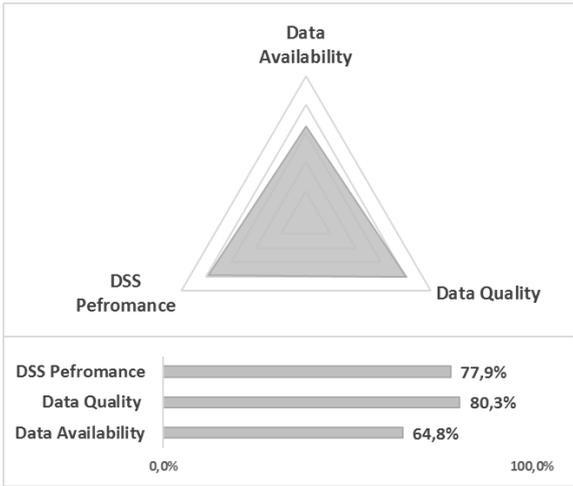
The organisation makes use of three main distribution centres in northern Italy, serving all the stores of the chain. The inventory management is handled by the procurement and the transportation by the logistic function. In this specific area, the decisions are taken, bearing in mind quantitative analysis of turnover rate and transportation costs.

In the following part, the analysis of the results achieved by the compilation of the Excel dashboard by the head of organisation department (i.e. the person in charge for the supervision of the entire team dedicated also to distribution planning decisions) is showed. The interview was conducted by two analysts and had approximately a duration of one hour.

Starting with the data relevance evaluation part, a really diversified evaluation of the list of data was given by the respondent. High significance is recognized to the descriptive characteristics of the product, transportation times and to the target service level, in alignment with their strategic objectives. Considered the sector in which the company operates, the market data are extremely important, such as the promotional events calendar. On the contrary few data received a very low evaluation by the interviewee, such as the in-transit stock level, considered their just-in-time (JIT) approach. This is mainly due to

their way of working, trying to produce and make available in real-time what the consumer market wants. Another data considered irrelevant concerns the possible substitute products.

In the Data Availability part, the first feature analysed is the data existence. Generally, it is possible to say that the whole list of data is generated, apart from the substitute products. Moreover, at the end the resulting data existence feature has an evaluation equal to 75,3%, since the whole dataset is generated in incomplete manner. On average the respondent has the access to the information when needed and the data are considered timely. In terms of data quality, the data are sufficiently well presented, correct and complete, giving room for improvement. Concerning the product characteristics affecting the lead time (e.g. customs requirements), they are not sufficiently updated to take the decisions, not presenting a good level both in accuracy and completeness.




---

**Data Productivity Index 2.0 = 40,5%**

---

Figure 58 CASE E Results

The DSS performance showed a good supporting capability of the underlying infrastructures, in terms of massive amounts of data quickly processed and wide range of analytics techniques used. The information system modules averagely are able to communicate, thus presenting updating issues across the functional groups in the organisation.

With the aim to calculate the final DPI 2.0, the three dimensions overall evaluation was computed. The overall indicator for the Data availability dimension obtained a not very high score (64,8%), particularly as a result of the partial data collection. While the Data quality result (80,3%) shows a reasonable level achieved in all the sub-features. In the DSS performance macro area, the evaluations given allow to reach an overall good level (77,9%). Finally, the value reached in the DPI 2.0 is 40,5%.

### 7.3.6 CASE F

Company F is a consortium for the wholesale trade of tools. This service company handles the orders and warehouse management on the behalf of the member organisations - in total 25 - that are their only clients. The company is firmly established in the sector of industrial distribution representing a real growth perspective in the sector for the Italian market, with a turnover of 16 million € and 15 employees. Thanks to the consortium formula, the member organisations are able to maintain their autonomy and participate actively, benefiting from the wealth of classified information and expertise offered by the group.

According to the specific policies, member organisations send their orders in digital format (ftp files) referring to a catalogue. The catalogue contains the full product range available that can be ordered and provided only to the member organisations. The consortium on a day-to-day basis fulfils orders through two launches – one in the early morning, the other in the late afternoon - serving approximately 30 stores, since a member organisation could own more than one subsidiary.

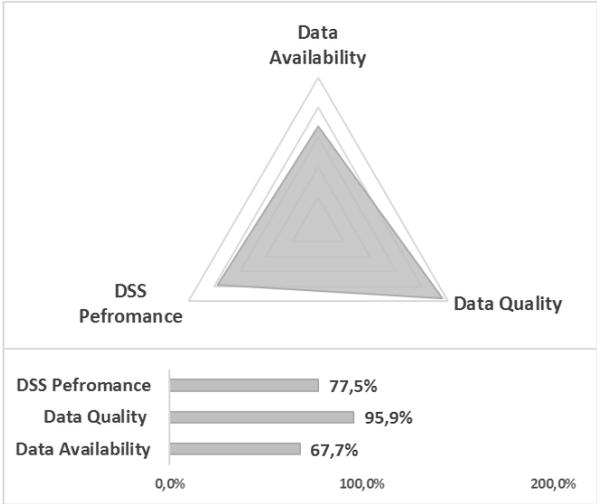
The interview was done in the presence of the general manager, the logistics buyer and the administrative officer and it was conducted by two analysts having a duration of about one hour. Based on the methodology explained in the previous section, after a presentation of the thesis objectives and a brief unstructured interview concerning the main companies'

activities, the model was submitted to the respondents. The Excel dashboard was mainly compiled by the logistics buyer and the administrative officer. Being in a distribution company, the persons interviewed are the ones really in touch with data and in charge for the distribution decisions.

Concerning the part related to the relevance evaluation of the list of data, at first glance, it was possible to notice that few data – 4 exactly – received a very low evaluation, in contrast with all the others acknowledged as extremely important for their reality. For instance, the general characteristics of the product are essential for their business, while the network data are not likewise considered important, since the low evaluation received. Also the transportation times and cost are equally considered significant. As concerns the inventory and flow data, the customer inventory level was not deemed meaningful for their reality, whereas the possible customer demand forecast could be crucial to optimize the distribution planning. All the market data were assessed of high importance. Particularly in this case it is possible to point out the high evaluation given to substitutable products, in contrast with the evaluation normally given in all the other cases conducted. This is mainly due to their market position, since they should be able to constantly provide to member organisations updated product catalogue and solutions, by getting rid of those products considered obsolete.

The first thing possible to notice is that they do not own at all 5 data out of 13, achieving in the feature data existence 69,5%, penalising in this way the final result obtained in data availability. Concerning for example the data related to customers inventory, at the moment they do not own this kind of data, how even the customer demand forecast, since the lack of a demand planning function in the group. However, they work with data that could be generally defined of high quality, accurate and complete, but not always understandable and well presented. This is a basic necessity to allow the product management. Furthermore, the whole data collected in the organisation are then used by managers to take decisions. As concerns the DSS performance factor, they have an AS/400 to implement business management programmes and a specific software (ClickSoftware) offering the possibility to improve the efficiency and optimize the business. In general data are well integrated in the company, on average analysed and processed and the underlying infrastructure is

considered able to process huge amount of data, performing descriptive and predictive analysis




---

**Data Productivity Index 2.0 = 57,5%**

---

Figure 59 CASE F Results

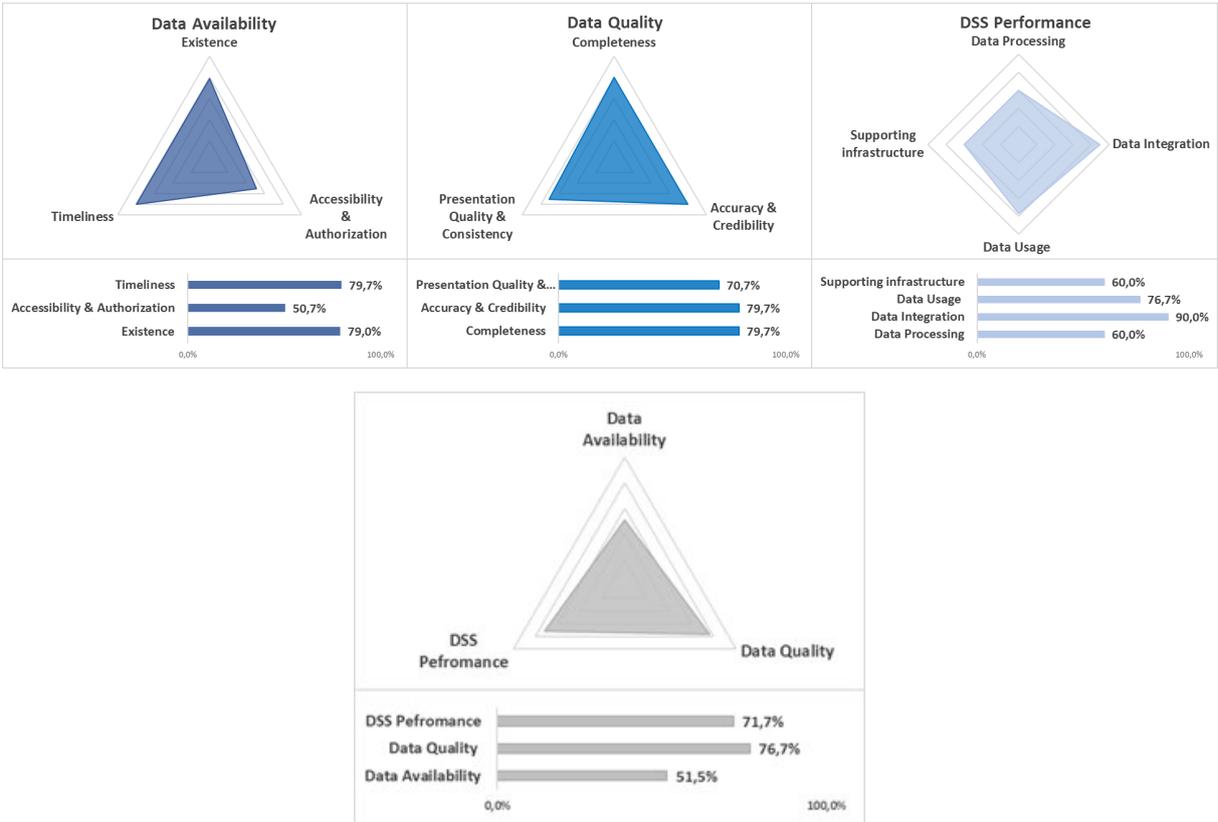
As can be seen in Figure 59, they achieved a very high level in the data quality factor (95,9%), since they have small datasets but well-maintained, and a good level in the DSS performance (77,5%). The few data collected weigh on the data availability score obtained (67,7%), even though the data are accessible across the organisation and timely. The DPI 2.0 result showed a quite high level of data exploitation in taking distribution-planning decisions (57,5%).

### 7.3.7 CASE G

Company G is a distribution company, specialized in the transportation of dry goods, at controlled temperature, operating in road-haulage and logistic field. The organisation, mainly at familiar management, provides logistics services, how even integrated logistics

services in Italy and Europe. It represents an established reality in the national and international market, with a turnover of about 220 million € and 350 employees.

The positioning in the chain is variable according to the suppliers and customers' requests. Indeed, they could manage the distribution cycle from the warehouse to the PoS, from the customer company to the final destination, ensuring a full service to the final customer. Considering that the inventory and transportation planning solutions are variable and strictly linked to the customers' needs, the company analyses and collaborates very closely with its customers. This in order to provide tailor made solutions and to enhance the synergies across the whole distribution network. Particularly, to ensure a sufficient presence nationwide, the organisation developed a network of localised platforms enabling storage, distribution and integrated logistics activities.



**Data Productivity Index 2.0 = 28,3%**

Figure 60 CASE G Results

In the following part the results of the Excel dashboard are going to be presented, obtained thanks to a phone call with the company business controller. The respondent was

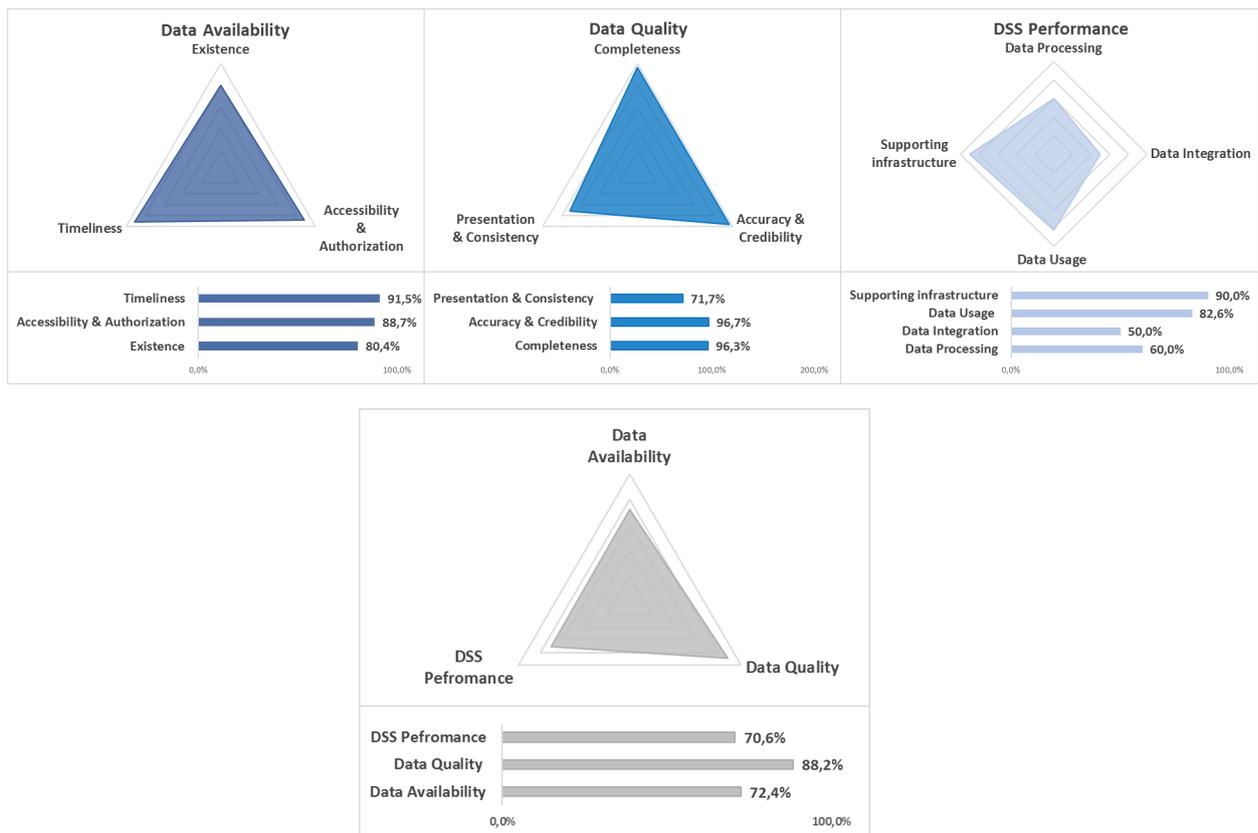
considered suitable, since he deals everyday with discussions and decisions belonging to the precise distribution field. The interview was conducted by two analysts with a duration of about one hour.

Starting from the analysis of the scores assigned by the respondent to evaluate the data relevance, it is possible to point out the great significance levels associated to the data specifically related to the description of the network, the relative flow and inventory level. This according to the reality in which the company operates. Whereas the product descriptive characteristics are considered of medium importance. Taken into account the close relationship the company aim to build with its customers, the customer demand forecast, and the promotional calendar are considered two essential data to be aligned with their needs. Since the strong relationships established with their customers, the interviewee defined the monitoring of substitute products not necessary. Moving to the data availability assessment, the organisation collects, even if partially, all the data presented in the list. For what concerns the data accessibility, the low evaluation given reflects the presence of some problems in the organisation, while the level achieved in the timeliness is medium. Regarding the data quality dimensions, the results achieved are quite good, even though the presence of some issues in the Data presentation dimension. Furthermore, it is important to highlight that in general all the data related to the market were evaluated with the lowest scores compared with the other data of the list. In DSS performance dimension, the integration of data across the organisation is considered by the interviewee quite satisfactory, while the majority of data are analysed and aggregated, in order to build significant metrics. The data effectively used in the distribution process by the respondent are mainly related to the descriptive, inventory and flow data area.

The result obtained in the Data Availability factor (51,5%) is strongly impacted by the low data accessibility level obtained. While the percentage obtained in the Data Quality dimension (76,7%) is the consequent result of the overall good level achieved in each sub-dimension. The resulting DSS performance (71,7%) is strongly penalised by the data processing and supporting infrastructure level achieved. Therefore, the consequent result obtained in the computation of the Data Productivity Index 2.0 is equal to 28,3%.

### 7.3.8 CASE H

Company H is an Italian company leader in the design, production and distribution of fashion, luxury, sports and performance eyewear; thus, showing off a vertically integrated business model covering the entire value chain build over the decades. The group has an overall turnover of about 9 billion € and more than 60 000 employees. The company is developed worldwide, covering more than 150 countries thanks to a network of approximately 9000 stores. Particularly, the business unit under analysis is the only one not submitting the general characteristics of the group just presented. Since the main business consists in apparel, footwear and accessories, the manufacturing is outsourced, not allowing the vertical integration. Specifically, for this business unit, the turnover is of about 280 million € with indicatively 1000 employees.



**Data Productivity Index = 45,1%**

Figure 61 CASE H Results

As concerns the distribution process and particularly the outbound flows, starting from the orders allocation according to the sales channels, when the product is available is then delivered to the point of sales. The network relies on 5 warehouses and it is supported by contract logistics. The business refers on specific society acting as interface with the different logistics providers.

The interview was conducted via Skype call with the distribution planner of the business unit just briefly presented (i.e. the person accountable for the optimization of the distribution planning and related decisions). It had total duration of approximately one hour and it was carried out by two analysts.

Next the company results, obtained through the compilation of the Excel dashboard, are presented and analysed. As first thing, it is important to analyse the relevance evaluation given by the interviewee relatively to the list of data presented. In particular, the highest score was assigned to the total delivery lead times and to the customer demand forecast. Indeed, the company stressed the importance to achieve high standards, both in terms of customer satisfaction and service level. The respondent underlined the necessity for this kind of business to define the distribution planning basing on the forecasts made by the demand planning function. Considering that the apparel products are completely produced in outsourcing, it is essential to rely on sound and valid forecasts, taking into consideration several factors, from the promotional events calendar to the different holidays in each county included in the interested value chain. For what concerns the data availability sub-dimensions evaluation, all the data presented in the initial list drafted are collected by the company, even if according to the respondent only few data could be defined completely generated. The results obtained in the other data availability dimension are considered generally satisfactory. Considering the data quality dimension, the data are evaluated as accurate (96,7%) and complete (96,3%), presenting some problems in the way they are presented. The evaluation achieved in the DSS performance dimension is the result of medium data integration across all the organisation functional groups and of a general analysis of the data collected. According to the respondent, previously working in another reality characterised by a strong digitalisation, the application software used is not enough performing and intuitive to support decision-making processes. While the support given by

the underlying infrastructure in the data analysis is considered really significant returning information in a timeframe considered useful to take the specific decision, getting over the mean of the interviewed companies. Finally, the data usage of the inventory and flow data is high, while the data related to the market are less used.

The overall value achieved in the final computation of the Data Productivity index is equal to 45,1%, particularly supported by the high evaluation obtained in the Data Quality dimension (88,2%). The resulting evaluation obtained in the Data Availability (72,4%) is mainly impacted by the data existence factor, since the partial data collection performed, while the DSS performance result (70,7%) is affected by the average values of data processing and data integration sub-dimensions.

### 7.3.9 Overall results

In this paragraph the results achieved by each case study are globally analysed and compared. In *Table 43* the Data productivity Index 2.0 reached by each company interviewed with its related macro factors are shown.

Company	Sector	Availability	Quality	DSS Performance	DPI 2.0
A	Home care retailers	25,5%	47,1%	40,8%	4,9%
B	Pharmaceutical	66,2%	75,6%	80,4%	40,3%
C	Home and Personal care	76,6%	71,8%	71,9%	39,6%
D	Mass Distribution	78,0%	88,9%	83,0%	57,5%
E	Mass Distribution	64,8%	80,3%	77,9%	40,5%
F	Freight and logistics	67,7%	95,9%	77,5%	50,3%
G	Freight and logistics	51,5%	76,7%	71,7%	28,3%
H	Fashion	72,4%	88,2%	70,6%	45,1%

*Table 43 Case studies results*

### Theme relevance

The test of the model through multiple cases studies enables firstly to verify the company reaction to the presentation of a total new metrics. Indeed, as deeply discussed within this work, literature does not provide any indication about a possible quantitative metrics capable to measure the data productivity. Currently any companies, in measuring its

performances, own an index able to trace how effectively data support a certain decision-making process. This lack within the company' performances measurement systems has been confirmed relevant by these case studies. Indeed, the expert interviewed supported our first assumptions concerning the need to have an insightful index to measure data productivity in decision-making process:

- Companies realized the potential of implementing Industry 4.0 practices
- Companies acknowledge the importance of collecting and generating high amount of data in supporting decision making
- Companies acknowledge the high value retrievable form the massive amount of data collected from different sources
- Companies lack a clear way to quantitatively assess their data set in terms of data quality and availability
- Companies acknowledge the importance of having an intuitive instrument able to measure and monitor the data productivity for a certain decision-making process.

Therefore, the model validation enables to understand that organizations have a strong interest on the subject covered in this work. This is proven also by the high response rate gained. Indeed 15 different companies have been contacted by email/phone by the authors and the 80% of these ones (12 of them) shown a high interest on the theme treated. Finally, the 53% of the companies (the 8 case studies discussed) effectively filled in the questionnaire presented, enabling to carry out the study. The companies interviewed showed their commitment not only through the compilation of the model, but also providing insightful suggestions, comments and feedbacks. Their interest is mainly focused on the results obtained, since the novelty of the monitoring tool proposed, how even on the possible improvement actions. Furthermore, it should be noticed that two of the enterprises that have been approached were not able to participate to the study because of internal privacy policy, even if they were delighted with the interest in this topic.

## Model Applicability

Moreover, case study was a useful approach to verify also the model applicability. Thus, especially performing the two pilot cases, it enables to understand if the Data Productivity model properly represent the business realities, in terms of data assessment features and related maturity models. This was particularly important for maturity models where with a first application they turn out to be too strict and academic, requiring multiple changes. Therefore, in this way it was developed an index that enables companies to achieve stimulating values. Secondly, the model application leads us to confirm that the interviewees do not have any difficulty when filling the model. On one side, the Excel Dashboard deployed was considered a manageable and intuitive tool; on the other side the questions asked through the maturity model adoption, were defined direct, clear and insightful, without leading to ambiguity.

## Factors analysis

Analysing the overall result of each company interviewed some observations are worthy of note. Two important aspect of the model must be underlined to better understand the company results:

- The relevance value to each data of the Interesting list is given by the user regardless of whether data are collected by the company.
- All the features introduced in the model are computed weighting the data evaluation with its related relevance.

Therefore, the results obtained in the Data productivity Index 2.0 and all its dimensions reflect these two aspects.

Observing the results obtained in each macro factor of the model, it is evident that that Data availability is the dimension that mostly penalizes the company results. In average the Data availability factor is equal to 62,83%, where the most performing one is the Case D (78,8%). Indeed, the model, as it is designed, highly punishes if a company does not collect data potentially useful in supporting a distribution planning process. Therefore, company must be aware that the Existent Data value has a predominate role in the Data Availability factor and

broadly in measuring the data productivity. This because, the computation is made by calculating firstly the percentage of existence data and secondly selecting which is the subset of the existent data that could be defined as timely and accessible. Moreover, if the company under analysis does not collect data that are considered by the interviewee highly relevant for the specific sector this weighs more in the Data Existence.

For instance, in the majority of cases, data related to the customer demand forecast are the most problematic ones. Indeed, if on one hand most of the interviewees have assigned to them a high relevance (5 = Data strictly necessary) recognising the high potential support that could be given by these data; on the other hand, the company actually does not own the right visibility. This result highlights the difficulty, typically present in business context, in achieving high level of integration and information exchange among the different actors playing in a supply chain. Another example refers to those data related to the possible substitute products that turn out to be not strictly relevant in most of the contexts, slightly affecting the Existent value in the event of not proper collection.

Generally, the data quality factor exhibits that most of the companies are capable to store data owing medium - high quality (in average 78%). Analysing the results achieved by the data quality features it could be underlined that companies have problems in the data representation. In fact, the decision maker has problems in data understandability and interpretability. The causes could be found or in the data collection process, where different sources and individuals treat and affect the data presentation, or in the inability of the system to maintain constant the data format.

Lastly the value obtained by the performances of the decision support systems show that companies interviewed averagely own infrastructure able to process and treat the massive amount of data collected. According to the valuation given by the interviewees, many companies relies on simple algorithms performing mostly descriptive and predictive analysis. Once more underlining the novelty of the topic treated for organisations.

Furthermore, it is possible to notice that the importance of market data is not yet acknowledged by companies, since the low score generally achieved both in the data collection and usage.

Finally, the Data Productivity Index 2.0 values achieved by the cases studies present a wide range that goes from 4,9% to 57,54%. The following classification (Table 44) was taken as a reference to classify the company size in term of turnover and employees.

Company size	Employees		Turnover (€)
Micro	Less than 10	OR	Less than 2 million
Small	Less than 50	OR	Less than 10 million
Medium	Less than 250	OR	Less than 50 million
Large	Over 250	AND	Over 50 million

Table 44 Company size classification

Therefore, all the companies interviewed could be classified as Large companies, except for Case F that is defined as a Small one. This was done to pursue the objective to analyse companies owing more structured processes. The large companies present averagely good results with a DPI 2.0 around 40%. These results strongly highlight the innovative nature of the topic treated. Indeed, even if enterprises are aware of data importance and the consequent impact that their deployment could have in the competitive advantages, companies still have some deficiencies in data collection, maintenance and processing. The only small one interviewed has a DPI value above the average (50,3%), showing despite its size, its capability in using data as a supportive means within the Distribution Planning process.

### 7.3.10 Judge the quality of the study

Yin proposes in his work (2009) four tests widely used to establish the quality of an empirical social research. Since case study could be defined as a social research, the four tests could be used in this case as criteria for the verification of the overall quality of the study.

The first tactic suggested by Yin (2009) to increase ‘Construct validity’ when doing case studies, is the usage of multiple sources of evidence. This happened in the data collection phase. Indeed, the model proposed is based - from its theoretical foundations, to its implementation through maturity model and consequently to DPI 2.0 development – on several studies and research from academia and models from practice. The whole set of

different types of existing models was used as basis for the final model development. Moreover, another related case study tactic is to have key informants to review the draft case study report. The consultation with expert opinions in the field taking place during the interviews, makes possible the review of the model design via evaluation loops, such as for instance the two pilot case studies. From a methodological point of view, the corrections made possible thanks to this procedure, increase the accuracy of the study, thus its construct validity.

In order to ensure 'External validity', it is advised to recur to the replication logic in multiple case studies. This test deals with the generalizability issue and is a major barrier when doing case studies. Indeed, each case was carefully selected, in order to have both literal replication of results and other cases to obtain also contrasting results (theoretical replication).

The 'reliability' intends to demonstrate that the study could be repeated, achieving the same results. For this reason, it was considered fundamental to correctly document the research procedure followed by establishing a case study protocol, with the main goal to reduce to the minimum the errors and biases in the research. Indeed, this study is thoroughly documented, enabling a researcher in the future to follow the same methodology and achieve the same results in another case company. Moreover, the documentation of the results could serve as input for the development of a case study database.

Finally, 'Internal validity' is considered a concern only for explanatory case studies, the ones aiming at finding and establishing causal relationships. Our study was not conducted with this scope.

The adherence to the following criteria just presented attests the validity and soundness of the study proposed.

# 8 Conclusions

## 8.1 Findings

In this section, it will be briefly presented the findings of this thesis work in relation to their nature – academic and practical - and to the correspondent research questions.

### Academic Objectives

**RQ1.** The first research question aims at defining the most appropriate features for the data productivity assessment. The starting point was considered the factors proposed by Miragliotta et al. (2018) in their work, taking inspiration from the OEE metrics. Three main components were defined – Data Availability, Data Quality, DSS Performance - and to describe in an exhaustive manner each component, a set of sub-factors has been determined. These sub- factors were identified and defined in detail relying on the previous academic studies.

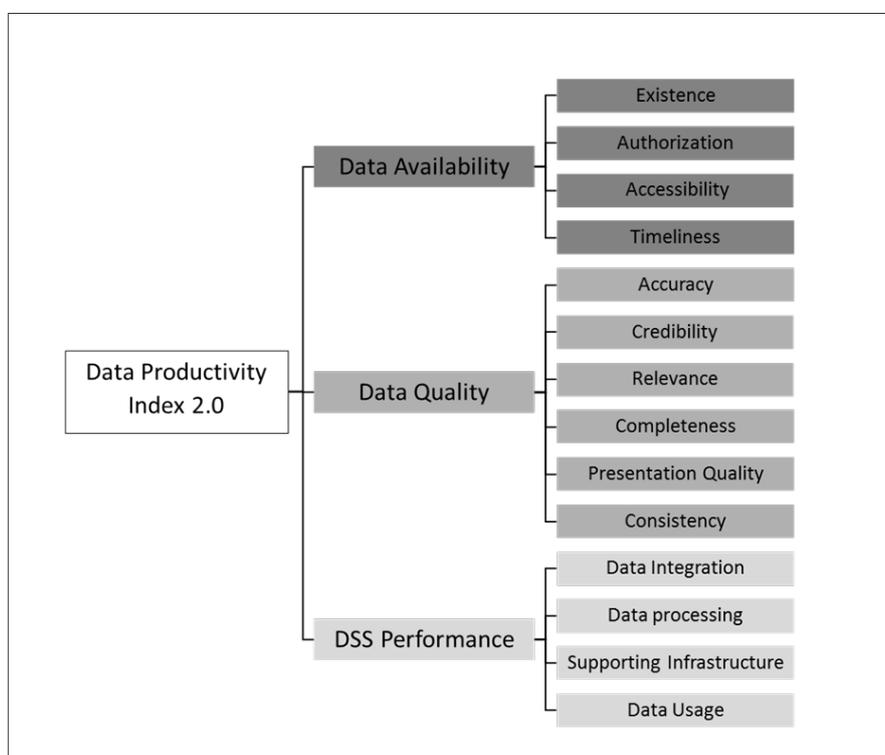


Figure 62 Data Productivity Index 2.0 - Dimensions

Concerning the Data Availability, the features deemed necessary to understand this dimension are the data existence, authorization, accessibility and timeliness. While the Data Quality is better defined through following features: the data accuracy, credibility, relevance, completeness, presentation quality and consistency. Finally, the sub-dimensions identified to describe the DSS Performance factor are the data integration, data processing, supporting infrastructure and data usage. In the graph all dimensions are clearly presented (*Figure 62*).

**RQ2.** In RQ2 the objective was to find a suitable approach to evaluate in a structured way the set of dimensions identified for the data assessment, in order to return an overall index. The calculation of this final numeric index, giving an indication of the company level of data exploitation in decision-making, was done relying on the support of maturity model. This is a qualitative approach, considered valid for the accomplishment of our scope. In the Data Productivity Maturity Model (DPMM), five maturity levels were defined to comprehensively describe and cover all the possible organisations' status. Thanks to use of the maturity approach, each dimension of the theoretical framework proposed was evaluated according to specifically developed sentences. The definition of sentences allows to reduce the user's subjectivity in the data evaluation.

To move from a qualitative assessment to a quantitative one, it was assigned a numeric value – from 0 to 1 - to each maturity level with the objective to obtain an overall numeric evaluation for the specific feature. Then these sub-dimensions were aggregated according to well-defined criteria in order to get for each macro-factor - Data Availability, Data Quality and DSS Performance - a unique percentage. Finally, the three macro-factors' values were merged together to achieve the Data Productivity Index 2.0 (DPI 2.0), in form of percentage.

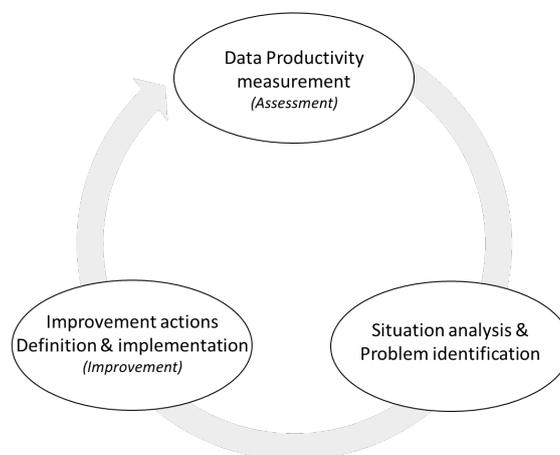
## Practical Objectives

**RQ3.** In this research question, the aim was to understand how is it possible to improve the applicability of the model in the business context, independently from the sector of belonging. As first step, it has been essential to detect the target of analysis in order to carry out this study. It was identified in organisations characterized by structured processes and monitored thanks to data. The RQ3 goal was mainly pursued following two paths. The first decision taken concerns the willingness to restrict the field of analysis to one specific decision-making process, thus focusing on the Distribution planning, intended as the optimization of the inventory management and transportation planning (only outbound flows). This decision enabled the definition of a list of data - Interesting Data - regarded as interesting for the decision-making process under analysis, identifying three main areas: product and network descriptive data, inventory and flow data and finally market data. According with the scope for which it was developed, the most suitable level of data aggregation was chosen. Thanks to this list definition, it was possible to define a standard input of the model, making in this way the index more comparable between organisations belonging also to different sectors. This has been possible also through the implementation of the maturity approach, mitigating the data assessment subjectivity with the introduction of sentences qualifying the specific maturity level. Furthermore, an Excel dashboard was especially developed to make structured and intuitive the evaluation for the user. The increased applicability of the model proposed was confirmed during the case studies performed and by the experts' positive opinions.

## 8.2 Improvement actions

The development and implementation of the Data Productivity maturity model (DPMM) provide an initial support for the organisations' improvement efforts in enhancing the level of data exploitation in decision-making. Indeed, according to Becker et al. (2009), maturity models aim at solving firstly the problem of defining an organisation current status concerning their capabilities and therefore deriving means of improvements. Also, Fraser et al. (2002) considers maturity model as integral part of an improvement process, and not only as absolute measure of performance.

Through an evolutionary improvement path - from obliviousness to a strategically optimized process – the model proposed suggests a staged improvement for each single feature previously presented. Each increment identifies a discrete change in the reference framework within which the individual must take a decision. Moreover, it could be considered as a guide leading toward the achievement of a desired level of competency or maturity. Indeed, the model provides the organisation with a clear basis for the gap analysis and allows the subsequent definition of possible directions for improvement actions and best practices per area. This facilitates the determination of key elements of current and potential process capabilities, identifying the most critical issue. Thus, enabling the organisation to consciously select a specific target level of maturity and then work towards that objective level.



*Figure 63 Phases for Data Productivity Assessment and Improvement*

In the development of the maturity model, three big improvement areas appeared, coinciding with the three macro factors identified for the data assessment: Data Availability, Data Quality, DSS Performance. These three-improvement areas can be approached either reactively or proactively. At first, it is possible to identify possible reactions to the situations; beyond that there are the proactive improvement actions.

Finally, the possible improvement practices can be classified into two main categories from the point of time, at which the snapshot is taken. Therefore, two general priority strategies were identified with a short-term and long-term perspective respectively, following the classification made by Batini et al. (2009).

For the short-term:

- *Data-driven strategies* aim at improving the data status, directly acting and modifying their value (e.g. the resolution of old mistakes and inefficiencies).

In the long-term:

- *Process-driven strategies* aim at improving the data status, identifying and deleting the root causes of errors and redesigning processes that create or modify the data (e.g. process redesigned to control the data format before storage or to avoid the manual input).

These two types of strategies apply a wide range of improvement techniques, including algorithms, heuristics and knowledge-based activities. Both the strategic approaches are considered important, since old mistakes and inefficiencies need to be resolved, as well as the relative root causes must be extinguished in order to prevent upcoming mistakes. However, generally the process-driven strategies involving process redesign are regarded as very expensive, whilst data-driven strategies are considered cost efficient in the short term, but costly in the long term.

Particularly, concerning the process-driven strategies, two main techniques are identifiable:

*Process control* refers to the implementation of check and control procedures monitoring the organisation current status, avoiding in this way the error propagation.

*Process redesign* refers to the process re-engineering with the aim to remove the possible root causes and to introduce new practices leading to the enhancement to a specific desired level. In case the process redesign is radical, reference is made to the business process reengineering that it is mainly composed of three activities. Starting from the process mapping and analysis of the as-is process – resorting to the BPMN 2.0 notation - it is designed the to-be process and finally implemented and continuously improved.

## 8.2.1 The Data Productivity Improvement matrix

In the light of the just presented subject, in this section it is proposed a simple classification of the organisations' data exploitation in decision-making, on which establish a basis of possible improvement practices to suggest and adopt in organisations.

At first, three improvement areas were identified in the model macro factors: Data Availability, Data Quality, DSS Performance. It is important to notice that these three areas are strictly connected. For instance, it is useless to collect data without being responsible for the quality aspect of data, how even to improve the data quality without the implementation of a suitable elaboration instrument. Therefore, it was possible, also according to the classification suggested by literature studies, to identify two main dimensions of classification - The Data Availability & Quality, and the DSS Performance – with the objective to design a matrix.

### First Dimension: Data Availability & Quality

To improve the *Data Availability & Quality* values companies should work on the feature defined in the proposed model. Particularly concerning data availability, the features presented are related to the Data Existence, Accessibility & Authorization and Timeliness. While, Data Quality in the Data Productivity model is characterized by three different data aspects: Accuracy & Credibility, Completeness and Presentation Quality & Consistency.

According to the model functioning, *Data Existence* is considered the most important feature, in comparison with the others, owing a greater weight into the model computation. Companies should have a greater sensibility aimed at a correct data collection and

generation. The strongly improvement of the company data acquisition processes have a long-term perspective since it implies the companies process redesign. To collect the massive amount of data required to properly support decision making, organizations should adopt the new technologies introduced in the industry 4.0 context such as sensors, CPS, IoT devices and communication network. Moreover, with the purpose to collect market data, companies should work on the realization of horizontal and vertical data integration within their supply chain in order to achieve an information alignment among the focal company, customers and suppliers. Multiple devices should be able to communicate among each other in order to achieve a perfect data alignment. Thanks to a more automated data generation process and data exchange among the different supply chain actors, an organization will strongly enhance its capability to generate all the potential supportive data for a certain decision-making process.

*Accessibility & Authorization* is the second factor for which it is possible to define improvement actions. Companies should provide the proper support to all the data users, making data easily accessible across the organization. The process of data accessibility and authorization should be optimized, avoiding wastage of time and resources. Companies must have a full awareness of the data needed by each actor, providing the appropriate means to fulfil the assigned tasks. Data users should be able to use and exploit data during decision-making process, owning the right to access the dataset needed to timely perform the task at hand.

In order to improve the *Data Timeliness*, the distinction among the long-term and short-term strategies could be made. An improvement technique coming from the data-driven strategies is the acquisition of new data in order to replace those ones generating problems. For instance, a database operating with non-current and obsolete data could be cleaned and refreshed with current ones, increasing instantaneously the data timeliness. This strategy is suitable for “one-time application” where a prompt data updating is needed making data ready-to-use. In order to obtain long-term improvement, companies should implement solutions enabling the information systems to reduce to the minimum the timeframe between data collection, processing and utilisation. For instance, companies should adopt the so-called real time infrastructures which are able to update data constantly and

automatically. Nowadays in big data context the real time concept is fundamental which required that data must be handled in the shortest possible time.

In real world data are often inaccurate and not credible containing multiple conflicts and errors. Data inaccuracy could be easily traceable and arranged through data-driven strategies, such as error localization and correction. This technique allows to detect and eliminate the quality error selecting those values not matching the quality rules. The data accuracy enhancement is made possible through the target establishment, defining comparable and reference values. Into our model data accuracy is associated with the concept of data credibility, by emphasising that if data are coming from specialized sources it is significantly higher the probability to gain free-of-error data. Finally, another data-driven technique applicable in order to obtain more credible and therefore accurate data is the source trustworthiness. Deploying this strategy, the data sources are at first selected analysing and evaluating the quality level of their data. Taking into consideration the long-term perspectives in order to improve the *Data Accuracy & Credibility*, the root causes of the identified problem must be recognised, redesigning the underlying processes. With a view of implementing a process-driven strategy, firstly the data sources selection process must be revised, effectively evaluating if the sources are specialised and certified and therefore providing high quality data. Then the process of data storage and maintenance must be studied and monitored, avoiding possible sources of inaccuracy such us data manipulation, falsification and manual processing.

To ensure *Data Completeness*, with short term objectives, multiple data driven techniques strategies could be implemented. Techniques as error localisation & correction might be deployed in order to firstly detect the localisations of the missing values. Indeed, these lacks could make a certain dataset not appropriate to support a decision-making process. Then the dataset under analysis must be fulfilled through data acquisition strategies, in order to introduce all the required parts of information entities. To be successful within the long term, the strategies implemented should be oriented to the whole process redesign, enabling companies to deal always with complete datasets. As happens for the other features, through a deeply understanding of the current state and the consequent identification of the problem sources, process redesign strategies might be implemented.

The last feature influencing the overall Data Quality value is the *Presentation Quality & Consistency*. Recalling that, presentation quality definition is related to the degree of data understandability and interpretability, while consistency is strictly connected to the preservation of the data format representation across time and space. Both the aspects could be treated implementing data driven strategies techniques by directly adjusting the data values. In short term inconsistency could be managed through standardization or normalization, replacing or supplementing values identified as not standard, with the ones matching the predefined rules. In order to obtain clearly understandable data, record linkage strategy could be a useful tool, detecting if different data representation referring to the same world object are applied in multiple tables, causing ambiguity. Therefore, the deployment of record linkage techniques is fundamental to keep to the minimum the matching errors. Moreover, data and schema integration strategies may result useful to improve both the presentation quality and consistency. These techniques aim is to eliminate conflicting data values coming from multiple sources or manual updating.

The achievement of consistent and well-presented data in a long-term perspective is possible through the underlined process redesign. For instance, both the features may be strongly improved by acting on three aspects:

- Clear definition of standardized rules for data representation
- Setting up of data updating monitoring able to avoid that two individuals could update at the same time the same data
- Introduction of an automatic mechanism for data updating

## Second dimension: DSS Performance

The second dimension of classification is the *DSS Performance*. The improvement related to this dimension are linked to the following main features: data integration, data processing, supporting infrastructure and data usage.

One main issue for organisations is the *Data Integration* across all the functional groups. This issue could arise considering all the different types of heterogeneity faced by data integration. These are classifiable in technological heterogeneities (i.e. mainly due to the

several representations of data in the different modules), schema heterogeneities (i.e. caused by the use of different data model and different representations for the same object) and instance-level heterogeneities (i.e. conflicting data values provided by distinguished sources). Data integration allows data access to users and offers a unified view from heterogeneous data sources. Indeed, to solve this kind of problem in the short term, it is possible to work on the elimination of redundancies and unnecessary use of memory, improving the data consistency, thus avoiding that the same data could appear twice with different values. In the long term the data integration could be achieved through the planned and coordinated redesign of the Information System (IS) architecture. This can be done integrating and enabling the interface of the different databases modules in a common global scheme, providing to each functional group a customized view. Through the IS redesign, it is possible to improve the performances of the overall cooperative IS of the company, in terms of exchanged data flows and central databases set up for coordination aims. In this way it is possible to achieve a better communication and interoperability between the different functional groups, avoiding the recurrence of updating issues.

The second DSS Performance feature for which it has been possible to define improvement actions is the *Data Processing*. In the short term, under the appropriate infrastructure support and after the definition of the specific questions and related measurement priorities, the organisation should develop efficient computer-assisted procedures. It is possible to treat data resorting to different techniques, such as:

- Probabilistic techniques, consisting in the broad set of analytical methods derived from the statistics and probability theory, varying from Bayesian networks to data mining.
- Empirical techniques, based on algorithmic techniques (e.g. sorting, tree analysis, neighbour comparison, and pruning).
- Knowledge-based techniques aimed at extracting knowledge from data files and applying reasoning strategies.

Since the huge amount of unstructured data collectable, it is crucial for organisations to be capable to process this kind of data and use them in decision-making. For these reasons, to analyse correlation and patterns for Big Data coming from many disparate sources, it is

necessary to use a robust ETL system to take the whole set of possible data into account and analyse it. In the long term, it is possible to develop and implement in the organisation highly innovative and advanced algorithms, applying machine learning and artificial intelligence solutions. In order to achieve such level of capability, it is required a redesign of the infrastructure, the definition of an ad-hoc monitoring system and the commitment of skilled people to implement it.

The *Supporting Infrastructure*, intended as the combination of hardware and software, is an essential aspect for the data processing and analysis. Organisations are unable to create value out of data without having suitable data infrastructures, capable to support more or less complex data analysis techniques. Indeed, to perform more sophisticated data analysis techniques on time, moving from simple descriptive analysis to more advanced prescriptive analysis; it is crucial to adopt a process-driven strategy, redesigning the infrastructure. The infrastructure functioning depends mainly on the skills and expertise of the software developer in charge for the writing code.

To improve the *Data Usage* feature, it is essential to work on the decision-maker full awareness of data potentiality in enhancing the decision-making process performances. Particularly, the existent data assessed as relevant by the decider, must be taken into consideration when making the specific decision. The decision-maker should improve its skills and the knowledge of the organisation dataset, with the aim to obtain a comprehensive view of the data considered relevant to take the precise decision in the specific sector.

From the intersection of the two dimensions – Data Availability & Quality and DSS Performance – a matrix made up of four quadrants is created. Each quadrant represents a class of companies, which requires a specific improvement strategy.

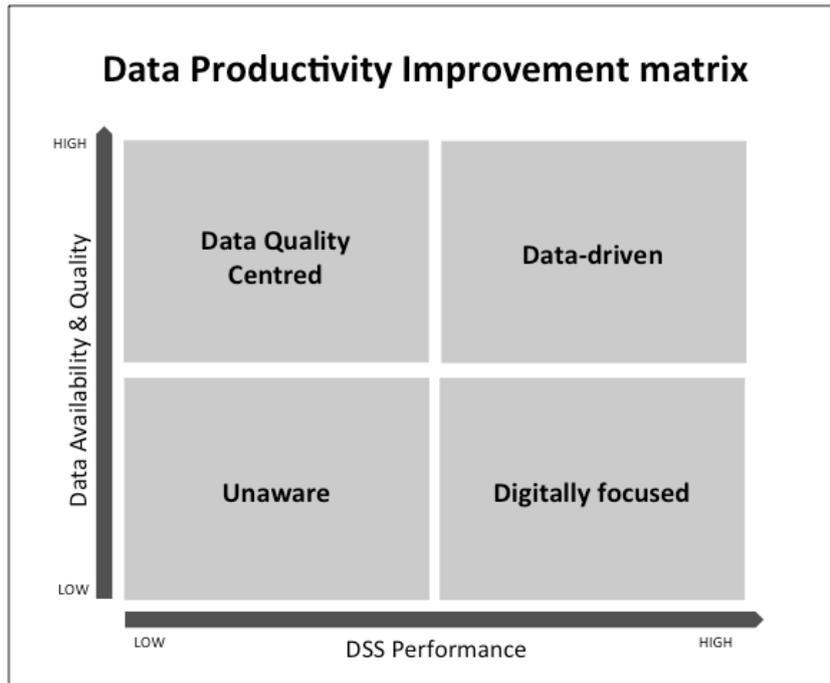


Figure 64 Data Productivity Improvement Matrix

Next each quadrant is going to be discussed:

## Unaware

Companies situated in this quadrant are characterized by low Data Availability & Quality and DSS Performance values. Comprehensively, the overall level of data exploitation is not considered sufficient to properly support decision-making processes. Companies present several deficiencies in the data collection process, being not capable to generate the right amount of data required to support a certain decision. Then, the collected data are not properly maintained, not treating and managing at all the data quality features. Moreover, also the company system does not support adequately the data user in the decision making. The underlying infrastructure is under-performing and enables only basic data analysis through simple and non-innovative algorithms. As a whole, unaware companies do not own data-driven and structured processes.

The multiple lacks characterizing these types of companies require radical changes. Process driven strategies must be implemented, firstly mapping the as-is situation to define the problem root causes and then redesigning the new processes. These strategies will make companies able to collect a higher amount of data identifying a proper way to store and

maintain data with high quality. The process redesign must cover also the IT infrastructure, allowing organizations to process and retrieve valuable information from the collected data.

### Data Quality Centered

Companies located in this quadrant own a high capability in Data Availability & Quality theme, while present inadequacies concerning the DSS performance. Indeed, the data collection activities could be considered progressive, generating most of the potentially useful data for a certain decision-making process. A particular attention to the data maintenance is taken in place, enabling decision makers to always deal with high quality data. The issues mostly concern the company decision support systems which is not nearly enough performing. The large amount of data collected could not be processed and analysed by the underlying infrastructure, not allowing to fully exploit the value retrievable from the generated data. Moreover, the system presents difficulty in the exchange and integration of data across the organization, leading to multiple duplications and errors.

Due to the medium-high companies' capability in collecting and maintaining data, the Availability & Quality dimension could be monitored and improved applying data-driven strategies. Indeed, the problems arising could be easily managed by directly modifying data. By contrast the decision support systems of the company located in this quadrant must be revised and updated. The supporting infrastructure must be redesigned in order to make it able to process massive amount of data and perform more complex and innovative algorithms. Procedure to monitor and check the processes must be implemented in order to continuously improve the performances.

### Digitally Focused

The companies localised in this quadrant are the ones characterized by high results achieved in the DSS Performance and low scores in the Data Availability & Quality. The underlying infrastructure used for the data analysis and processing is considered highly performing and able to support advanced and innovative analysis techniques. The functional groups are well aligned, and the IS of the overall organisation is scalable and flexible, achieving enhanced interoperability and coordination. The main issues are represented by the poor data collection and in general by the little care dedicated to the data quality features. Therefore,

nevertheless the supporting infrastructure is highly developed to process and analyse data on time, the technological and digital capabilities owned are not fully exploited in their potentiality to support the decision-making, since the current problematic status of data processed.

The first strategic indication suggested is to better define and understand the data needed according with the objectives and subsequently to improve the data collection. In general, to solve data availability and quality problems is essential to resort to the support of process-redesign techniques, in order to solve problems acting on the root causes. Moreover to maintain a good and stable level in the quality of data, it is necessary to implement also control procedure when new data are created, datasets are updated or new datasets are accessed by the process. Data-driven strategies are considered not enough in this case, since they are suitable for one-time application. The improvement approach suggested concerning the DSS performance relates to continuous improvement practices, since the good degree of digitalisation presented by the company.

## Data-Driven

With this term it is possible to identify the companies characterized by a good level both of Data Availability & Quality and of DSS Performance. These organisations are the ones better adapting to changes in the environment and really understanding the potentiality and the relevance of data for organisation, basing their decision-making process on evidence rather than intuition. According to the results obtained, the data are collected in a sufficient way and the company is able to exploit data since the good level of accuracy, presentation quality and consistency. Moreover, also the supporting infrastructure is considered highly performing, allowing to process and analyse huge amount of structured and unstructured data on time resorting to advanced and innovative techniques. The data flows are integrated across the functional groups, since the company IS ensures improved interoperability, communication and cooperation.

Generally, considered the good level achieved by the organisation, it possible to suggest the implementation of continuous improvement approach. Furthermore, organisations should apply a reactive strategy, inserting monitoring procedures both in the data production

process and in data analysis process, with the aim to avoid data degradation and error propagation.

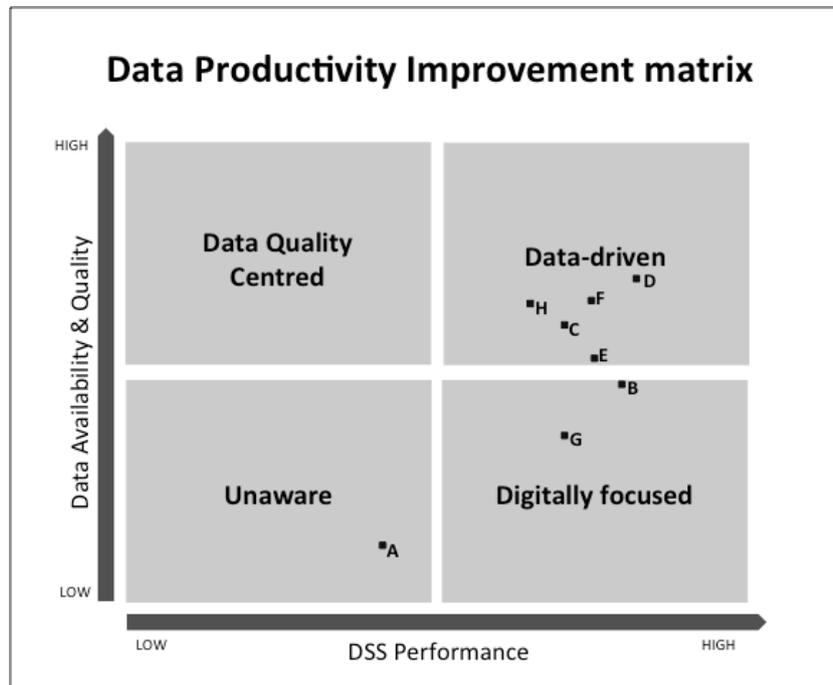


Figure 65 Data Productivity Improvement Matrix Application

In Figure 65 the resulting DPI 2.0 obtained by the companies' interviewed were reported in the Data Productivity Improvement matrix. The results were mapped according to the two dimensions previously identified, following also in this case a measurement scale from 0 (lowest evaluation) to 1 (highest evaluation). The horizontal dimension value is represented exactly by the percentage obtained by the company in the factor DSS percentage, while the value of vertical dimension is equal to the multiplication between the Data Availability and Data Quality percentages. At first glance it is possible to notice that only one company is positioned in the *Unaware* quadrant, showing in this way an overall low level of consciousness about the topic. Two companies are positioned in the *Digitally Focused* quadrant, considering the good performance of their Information System in processing data and the poor cure of the quality of their datasets. Whilst the majority of companies interviewed (5 out of 8) are localised in the *Data-driven* quadrant, close to the border of the digitally focused. Moving beyond the general good results achieved in both the dimensions, their position indicates the necessity to improve the Data Availability& Quality dimension.

## 8.3 Model Limitations & Future Research

This section is dedicated to the presentation of the main limits of the Data Productivity model developed in this thesis work, as well to the related future research directions. The limitations identified concern principally certain aspects of the model structure and the initial necessary assumption done. While the possible directions of research and applications identify the potential future actions to overcome the model limitation previously highlighted.

### Target Definition

The definition of the target achievable has been an important point in the development of the Data Productivity Index 2.0. The main inspiration at the basis of the model is the parallelism with the Overall Equipment Effectiveness (OEE). In the context of application, the reference standard set for the productivity of machinery is a value as close as possible to 100%. This even if the reference value obtained by the organisations' best-in-class performance is equal to 85%.

Therefore, our model was developed following this principal notion and giving to the companies as achievable standard for the DPI 2.0 the 100%. In particular, the model relies on the definition of specific maturity levels. The maturity levels were built and developed by making reference to a scale from 0 to 1. Since the model construction was entirely based on this principle, it is important to underline the possibility for organisations to achieve a result equal to 100%.

However, the results obtained by the model application in selected companies through case studies, are mainly around the 40%, with a maximum value obtained of about 57%. Considering the lack until now of a measurement tool for the data productivity in decision-making, this outcome could be considered as a good starting point for companies, giving room for a possible improvement. In this case, in contrast to the OEE, currently it is impossible to define a general best-in-class reference value of this metrics for organisations, since the narrow sample on which the model was tested. At the moment a rationale stating

for example that a result equal to 70-80% could be considered a reasonable standard, doesn't exist for this metrics. It is necessary to keep on moving towards a standardisation direction, working with standardisation bodies to ensure the model transposition in business context, with the purpose to achieve a universal benchmark

## Model Building

The Data Productivity model was developed through literature studies firstly with a view to define the most appropriate structure to evaluate companies' datasets. This could be seen as a first step in achieving a metrics able to give to company's manager a comprehensive evaluation of the current state of data. Future researches could be aimed at identifying a wider set of dimensions to introduce into the model structure, better describing the three macro-factors of the model.

Moreover, into the model proposed a basic assumption is that features are mutually independent (e.g. a data could be complete but not accurate, or it could be timely but not accessible). Statistical analysis conducted on large scale might detect some possible correlation among the dimensions, requiring variation in the model assumptions and on the practical computation made to obtain the DPI 2.0 value. Therefore, the possible dependencies among the different features could be discussed, testing the model on a bigger sample than the proposed one through a survey, and analysing the results with more accurate statistical analysis. On this basis, the introduction of new and more detailed features in the theoretical model structure does not necessarily correspond to a longer model application, since thanks to correlation studies it is possible to aggregate some dimensions.

For its nature the data assessment is affected by high subjectivity, thus strongly dependent on the user perceptions and on the application context, reducing the universal comparability of the index. Regardless in the development phase of the model different measures to mitigate the data assessment subjectivity have been taken, this is still an open issue that could be better improved and investigated. Indeed, in this work to mitigate the evaluation subjectivity, in the model application it was selected carefully the data user in order to interview a person potentially owing a high familiarity with both the decision process under analysis and the relative required data. Despite these mitigation strategies deployed, the

model does not take into consideration that an individual could don't have a full awareness of the current state of data. This limit particularly affects the Data Availability and Data Quality features while the dimensions characterizing the DSS Performance are ruled by a major objectivity.

Furthermore, in this work data features are assessed through the deployment of a qualitative approach – the maturity model - that even if it represents a structured approach, it is dominated by high subjectivity. Therefore, data subjectivity is still an open question, that requires detailed studies, that could be addressed discussing other possible quantitative or quantitative methods to evaluate the data dimensions.

Moving from the qualitative approach to a quantitative one with the aim to obtain as result a percentage, the model assigns a numeric value going from 0 to 1 to each maturity level of each feature assessed. This crucial point of the model is object of discussion, highly impacting on the Data Productivity 2.0 Index computation. Future research must strive for the maturity model calibration, understating if the value adopted by each maturity level properly represent the related sentences. Experimentation on large scale could be done, in order to validate different methodological approach and to establish the most appropriate one.

## Model Application

Concerning the application, the model required a constant maintenance principally on two aspects:

- Maturity models for definition need to be constantly updated, in order to properly and realistically represent the current state of business processes
- The input of the model - the Interesting Data – must be regularly updated and revised, in order to verify its completeness and appropriateness in describing data that could potential support a certain decision-making process.

In this work the model was tested for only a specific decision making – the distribution planning – for manufacturing and distribution companies. Future research could improve and enlarge the field of model applicability, implementing the model in different processes

and sectors, and therefore making the computation of the Data Productivity Index 2.0 possible for a higher set of decisions. To apply this metrics to all the enterprises' processes, firstly a classification of the decision-making process must be made, consequently analysing the potential list that could be considered a supportive and interesting instrument for the decision-making process under analysis.

In the model applicability another relevant issue is related to the object of analysis. In fact, the model input is a list of potential interesting data for the decision-making process under analysis, presented with a quite high aggregation level (e.g. Physical product characteristics). The main issue related to the model input is that data managed at different aggregation level, own different characteristics. For instance, considering this point, companies in giving data evaluation should averagely consider the current status of data for the different product lines owned. Indeed, according to the priority assigned by the business to the specific product lines, different assessment could be given. Therefore, future studies could be focused on the analysis and evaluation of data at different aggregation level, discussing which is the most appropriate one.

## Final Consideration

The model proposed is a first step in establishing a metrics able to measure how data support a decision-making process. The aim of this work is to introduce a new index into the companies' performance measurements systems, able to measure the degree of data exploitation, retrieving valuable information and effectively deploying data to optimize decisions-making processes. Despite the issues presented, the DPI 2.0 goes beyond the limits of the first version thanks to the improvement of different aspects. It provides a more comprehensive data assessment model structure, involving a high number of features. The model applicability is enhanced thanks to the focus on one specific decision-making process and on the consequent development of a general input enforceable in several sectors. Moreover, thanks to the maturity approach the model introduced in this work enables a more structured and guided way for the data evaluation. The model was presented and validated resorting to the application in eight different companies, adopting the case study methodology. This enabled the construction of a final model, aligned with the business needs.

Summarizing, the main goal of this thesis work was to develop a model applicable for all companies' sizes and all the different industries, relying on the maturity model approach and on the definition of a theoretical framework. This model has been proven to represent a good reference for organisations to assess and monitor the data exploitation in decision-making and benchmark it against other business realities.

# 9 Bibliography

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16(1), 3-9.

Ahituv, N. (1989). Assessing the value of information: Problems and approaches.

Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, 107-136.

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review*, 114, 416-436.

Asdecker, B., & Felch, V. (2018). Development of an Industry 4.0 maturity model for the delivery process in supply chains. *Journal of Modelling in Management*, 13(4), 840-883.

Australian Bureau of Statistics. "Transcript." 23 October 2009: 21.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, 41(3), 16.

Battista, C., & Schiraldi, M. M. (2013). The logistic maturity model: Application to a fashion company. *International Journal of Engineering Business Management*, 5(Godište 2013), 5-29.

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing maturity models for IT management. *Business & Information Systems Engineering*, 1(3), 213-222.

Bellinger, G., Castro, D., & Mills, A. (2004). Data, information, knowledge, and wisdom.

Benabdellah, A. C., Benghabrit, A., & Bouhaddou, I. (2016, November). Big data for supply chain management: Opportunities and challenges. In *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1-6). IEEE.

Bernstein, J. H. (2009). The data-information-knowledge-wisdom hierarchy and its antithesis.

Bovee, M., Srivastava, R. P., & Mak, B. (2003). A conceptual framework and belief-function approach to assessing overall information quality. *International journal of intelligent systems*, 18(1), 51-74.

Burgess, K., & Singh, P. J. (2006). A proposed integrated framework for analysing supply chains. *Supply Chain Management: An International Journal*, 11(4), 337-344.

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14.

Cappiello, C., Francalanci, C., & Pernici, B. (2004, June). Data quality assessment from the user's perspective. In *Proceedings of the 2004 international workshop on Information quality in information systems* (pp. 68-73). ACM.

Cianca, E (2018). Data productivity measurement in industry 4.0: a proposal for a maturity model.

Cleven, A., & Wortmann, F. (2010, January). Uncovering four strategies to approach master data management. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on* (pp. 1-10). IEEE.

Comuzzi, M., & Patel, A. (2016). How organisations leverage big data: A maturity model. *Industrial Management & Data Systems*, 116(8), 1468-1492.

Courtney, J. F. (2001). Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS. *Decision support systems*, 31(1), 17-38.

"Crosby, P. (1979). *Quality is Free: the Art of Making Quality Certain*. New York: McGraw-Hill."

Curtis, B., W.E. Hefley, and S.A. Miller, *The People Capability Maturity Model – Guidelines for Improving the Workforce*. 2 ed. SEI Series in Software Engineering. 2010, Boston, MA: Addison-Wesley.

Dal, B., Tugwell, P., & Greatbanks, R. (2000). Overall equipment effectiveness as a measure of operational improvement—a practical analysis. *International Journal of Operations & Production Management*, 20(12), 1488-1502.

Dalkir, K. (2013). *Knowledge management in theory and practice*. Routledge.

Davenport, T., & Prusak, L. (2000). *Working Knowledge - How Organizations Manage What They Know*. Boston, Massachusetts: Harvard Business School Press.

De Bruin, T., Freeze, R., Kaulkarni, U., & Rosemann, M. (2005). Understanding the main phases of developing a maturity assessment model.

De Carolis, A., Macchi, M., Kulvatunyou, B., Brundage, M. P., & Terzi, S. (2017, July). Maturity models and tools for enabling Smart Manufacturing Systems: comparison and reflections for future developments. In IFIP International Conference on Product Lifecycle Management (pp. 23-35). Springer, Cham.

Deshpande, S. P., & Thakare, V. M. (2010). Data mining system and applications: A review. *International Journal of Distributed and Parallel systems (IJDPS)*, 1(1), 32-44.

Diewert, E., & Lawrence, D. (1999). *Measuring New Zealand's Productivity* (No. 99/05). New Zealand Treasury.

Eckerson, W. (2004), "Gauge your data warehouse maturity", *DM Review*, Vol. 14 No. 11, pp. 34-35.

Erengüç, Ş. S., Simpson, N. C., & Vakharia, A. J. (1999). Integrated production/distribution planning in supply chains: An invited review. *European journal of operational research*, 115(2), 219-236.

Evans, J. R., & Lindner, C. H. (2012). Business analytics: the next frontier for decision sciences. *Decision Line*, 43(2), 4-6.

Faucher, J. B. P., Everett, A. M., & Lawson, R. (2008). Reconstituting knowledge management. *Journal of knowledge management*, 12(3), 3-16.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34.

Fisher, D. M. (2004). The business process maturity model: a practical approach for identifying opportunities for optimization. *Business Process Trends*, 9(4), 11-15.

Fraser, P., Moultrie, J., & Gregory, M. (2002). The use of maturity models/grids as a tool in assessing product development capability. In *Engineering Management Conference, 2002. IEMC'02. 2002 IEEE International* (Vol. 1, pp. 244-249). IEEE.

García-Mireles, G. A., Moraga, M. A., & García, F. (2012). Development of maturity models: a systematic literature review.

Geissbauer, R., Vedso, J., & Schrauf, S. (2016). *Industry 4.0: Building the digital enterprise*. Retrieved from PwC Website: <https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf>.

- Gericke, A., Rohner, P., & Winter, R. (2006). Networkability in the health care sector- Necessity, measurement and systematic development as the prerequisites for increasing the operational efficiency of administrative processes. *ACIS 2006 Proceedings*, 65.
- Gorla, N. (2003). Features to consider in a data warehousing system. *Communications of the ACM*, 46(11), 111-115.
- Gorry, G. A., & Scott Morton, M. S. (1971). A framework for management information systems.
- Grigorak, M., & Shkvar, O. (2011). A logistic approach for description of decision-making process. *Aviation*, 15(1), 21-24.
- "Harmon, P. (2004). Evaluating an Organization's Business Process Maturity, URL [http://www.bptrends.com/resources\\_publications.cfm](http://www.bptrends.com/resources_publications.cfm), Accessed 30 September 2005. "
- Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2016). The age of analytics: Competing in a data-driven world. McKinsey Global Institute, 4.
- Hermann, M., Pentek, T., & Otto, B. (2016, January). Design principles for industry 4.0 scenarios. In 2016 49th Hawaii international conference on system sciences (HICSS) (pp. 3928-3937). IEEE.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly*, 28(1), 6.
- Huan, S. H., Sheoran, S. K., & Wang, G. (2004). A review and analysis of supply chain operations reference (SCOR) model. *Supply Chain Management: An International Journal*, 9(1), 23-29.
- Huang, S. H., Dismukes, J. P., Shi, J., Su, Q. I., Razzak, M. A., Bodhale, R., & Robinson, D. E. (2003). Manufacturing productivity improvement using effectiveness metrics and simulation analysis. *International Journal of Production Research*, 41(3), 513-527.
- Joshi, K. D. (2001, January). A framework to study knowledge management behaviors during decision making. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (pp. 12-pp). IEEE.
- Kaner, M., & Karni, R. (2004). A capability maturity model for knowledge-based decisionmaking. *Information Knowledge Systems Management*, 4(4), 225-252.

- Kärkkäinen, H., & Silventoinen, A. (2015, October). Different approaches of the PLM maturity concept and their use domains—analysis of the state of the art. In IFIP International Conference on Product Lifecycle Management (pp. 89-102). Springer, Cham.
- Khan, M., Wu, X., Xu, X., & Dou, W. (2017, May). Big data challenges and opportunities in the hype of Industry 4.0. In 2017 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.
- Khatibian, N., Hasan gholoi pour, T., & Abedi Jafari, H. (2010). Measurement of knowledge management maturity level within organizations. *Business Strategy Series*, 11(1), 54-70.
- Klimko, G. (2001, November). Knowledge management and maturity models: Building common understanding. In *Proceedings of the 2nd European Conference on Knowledge Management* (pp. 269-278). Bled, Slovenia.
- Kościelniak, H., & Puto, A. (2015). BIG DATA in decision making processes of enterprises. *Procedia Computer Science*, 65, 1052-1058.
- Kyung-seok, R. (2005, February). A study on data quality management maturity model. In *Advanced Communication Technology, 2005, ICACT 2005. The 7th International Conference on* (Vol. 1, pp. 598-603). IEEE.
- Lahrmann, G., Marx, F., Winter, R. and Wortmann, F. (2011), "Business intelligence maturity: development and evaluation of a theoretical model", *Proceedings of the 44th Hawaii Conference on System Sciences (HICSS)*, pp. 1-10.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S. and Kruschwitz, N. (2011), "Big data, analytics and the path from insights to value", *MIT Sloan Management Review*, Vol. 52No. 2, pp. 21-32.
- Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & management*, 40(2), 133-146.
- Lichtenberg, Herb. "Production Losses: How to Find and Reduce Them." *Strategic Asset Management 1996-2011*, Samicorp, n.d.
- Liew, A. (2007). Understanding data, information, knowledge and their inter-relationships. *Journal of knowledge management practice*, 8(2), 1-16.
- Liu, L., & Chi, L. (2002). Evolutional Data Quality: A Theory-Specific View. In *IQ* (pp. 292-304).
- Liu, S., & Papageorgiou, L. G. (2013). Multiobjective optimisation of production, distribution and capacity planning of global supply chains in the process industry. *Omega*, 41(2), 369-382.

Ljungberg, Ö. (1998). Measurement of overall equipment effectiveness as a basis for TPM activities. *International Journal of Operations & Production Management*, 18(5), 495-507.

Lockamy III, A., & McCormack, K. (2004). Linking SCOR planning practices to supply chain performance: An exploratory study. *International journal of operations & production management*, 24(12), 1192-1218.

Lummus, R. R., Vokurka, R. J., & Alber, K. L. (1998). Strategic supply chain planning. *Production and inventory management Journal*, 39(3), 49.

"M. Kohlegger, R. Maier, and S. Thalmann, "Understanding Maturity Models Results of a structured Content Analysis," presented at the IKNOW '09 and I-SEMANTICS '09, Graz, Austria, 2009."

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

Marr, B. (2015). Where big data projects fail. *Forbes*. [Online]. [Cited: 20 August 2017.] <https://www.forbes.com/sites/bernardmarr/2015/03/17/where-big-data-projects-fail>.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.

"Mettler, T. (2009). A Design Science Research Perspective on Maturity Models in Information Systems,. *Design*, 41, 1-13."

Mettler, T., & Rohner, P. (2009, May). Situational maturity models as instrumental artifacts for organizational design. In *Proceedings of the 4th international conference on design science research in information systems and technology* (p. 22). ACM.

Mintzberg, H., & Waters, J. A. (1990). Does decision get in the way. *Organization Studies*, 11(1), 1-6.

Miragliotta, G., Sianesi, A., Convertini, E., & Distanto, R. (2018). Data driven management in Industry 4.0: a method to measure Data Productivity. *IFAC-PapersOnLine*, 51(11), 19-24.

Mohan, S. (2017) *Big Data: Transforming Logistics and Supply Chain*.

Naumann, F., & Freytag, J. C. (2005). Completeness of information sources. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, Institut für Informatik.

- Nemati, H. R., Steiger, D. M., Iyer, L. S., & Herschel, R. T. (2002). Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems*, 33(2), 143-161.
- Otto, B., & Hüner, K. M. (2009). Funktionsarchitektur für unternehmensweites Stammdatenmanagement Inhaltsverzeichnis. *Data Management*.
- Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability maturity model, version 1.1. *IEEE software*, 10(4), 18-27.
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218.
- Power, D. J. (2002). *Decision support systems: concepts and resources for managers*. Greenwood Publishing Group.
- Power, D. J. (2007). A brief history of decision support systems. *DSSResources. COM*, World Wide Web, <http://DSSResources.COM/history/dsshistory.html>, version, 4.
- Power, D. J. (2014). Using 'Big Data' for analytics and decision support. *Journal of Decision Systems*, 23(2), 222-228.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), 51-59.
- Romano, P., & Danese, P. (2010). *Supply chain management: "la" gestione dei processi di fornitura e distribuzione*. McGraw-Hill.
- Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*, 33(2), 163-180.
- Ryu, K. S., Park, J. S., & Park, J. H. (2006). A data quality management maturity model. *ETRI journal*, 28(2), 191-204.
- Saha, B., & Srivastava, D. (2014, March). Data quality: The other face of big data. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on* (pp. 1294-1297). IEEE.
- Schumacher, A., Erol, S., & Sihn, W. (2016). A maturity model for assessing industry 4.0 readiness and maturity of manufacturing enterprises. *Procedia CIRP*, 52, 161-166.
- Sen, A., Ramamurthy, K., & Sinha, A. P. (2012). A model of data warehousing process maturity. *IEEE Transactions on Software Engineering*, 38(2), 336-353.

Sen, A., Sinha, A. P., & Ramamurthy, K. (2006). Data warehousing process maturity: An exploratory study of factors influencing user perceptions. *IEEE Transactions on Engineering Management*, 53(3), 440-455.

Sianesi, A. (2014). *La gestione del sistema di produzione: pianificazione, programmazione, controllo, misura e miglioramento*. Etas.

Sidi, F., Panahy, P. H. S., Affendey, L. S., Jabar, M. A., Ibrahim, H., & Mustapha, A. (2012, March). Data quality: A survey of data quality dimensions. In *Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on* (pp. 300-304). IEEE.

Simon, H. A. (1960). *The new science of management decision*.

Smith, R., & Hawkins, B. (2004). *Lean maintenance: reduce costs, improve quality, and increase market share*. Elsevier.

Spina, G. (2012). *La gestione dell'impresa*. Etas.

Spruit, M., & Pietzka, K. (2014). The MD3M questionnaire: Assessing master data management maturity. Technical Report Series/Department of Information and Computing Sciences, Utrecht University, (UU-CS-2014-022).

Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, 51, 1068-1076.

Stadtler, H. (2005). Supply chain management and advanced planning—basics, overview and challenges. *European journal of operational research*, 163(3), 575-588.

Stank, T. P., & Goldsby, T. J. (2000). A framework for transportation decision making in an integrated supply chain. *Supply Chain Management: An International Journal*, 5(2), 71-78.

Strassmann, P. A. (2004). *Defining and measuring information productivity*. ComputerWorld.

Team, C. P. (2002). *Capability maturity model® integration (CMMI SM), version 1.1*. CMMI for Systems Engineering, Software Engineering, Integrated Product and Process Development, and Supplier Sourcing (CMMI-SE/SW/IPPD/SS, V1. 1).

Tuomi, I. (1999, January). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. In *Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on* (pp. 12-pp). IEEE.

Vercellis, C. (2009). *Business intelligence. Data mining and optimization for decision making*.

- Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86-95.
- Wang, R. Y., Reddy, M. P., & Kon, H. B. (1995). Toward quality data: An attribute-based approach. *Decision support systems*, 13(3-4), 349-372.
- Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer*, 40(9).
- Weber, C., Königsberger, J., Kassner, L., & Mitschang, B. (2017). M2DDM—a maturity model for data-driven manufacturing. *Procedia CIRP*, 63, 173-178.
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.
- Wee, D., Kelly, R., Cattell, J., & Breunig, M. (2015). *Industry 4.0-how to navigate digitization of the manufacturing sector*. McKinsey & Company, 58.
- Wettstein, T., & Kueng, P. (2002). A maturity model for performance measurement systems. *WIT Transactions on Information and Communication Technologies*, 26.
- Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1), 97-107.
- Yin, R. K. (2009). *Case study research: design and methods*.
- Zhou, K., Liu, T., & Zhou, L. (2015, August). Industry 4.0: Towards future industrial opportunities and challenges. In *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (pp. 2147-2152). IEEE.
- Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American society for information science and technology*, 58(4), 479-493.

# Annex

## A1. Excel Dashboard developed to apply Data Productivity 2.0 model

1<sup>st</sup> spreadsheet:

IL SEGUENTE FILE È COMPOSTO DA 5 PARTI:	
<b>1. Informazioni generali</b>	Nel primo foglio sono presenti una serie di domande mirate a delineare un'overview dell'azienda, con un particolare focus sul Distribution Planning process (gestione scorte, trasporti e ottimizzazione con l'interfaccia)
<b>2. Data Relevance</b>	Nel secondo foglio è presentato l'input del modello: una lista di aree di dato ritenute interessanti per il Distribution planning process. Per ciascun dato verrà richiesto di valutarne la rilevanza relativa al Distribution Planning process.
<b>3. Data Availability &amp; Quality</b>	Nel terzo foglio viene valutata la disponibilità e la qualità di ogni dato della lista in input per affrontare le decisioni relative al Distribution Planning process. La valutazione viene fatta attraverso i Maturity Models.
<b>4. DSS Performance</b>	Nel quinto foglio viene valutato come gli strumenti supportano il processo decisionale di Distribution Planning. La valutazione viene fatta attraverso i Maturity Models.
<b>5. Risultati</b>	Nel sesto foglio vengono presentati i risultati del modello, mostrando separatamente i valori ottenuti per Disponibilità dei dati, Qualità dei dati, e Performance dei sistemi di supporto al processo decisionale. Infine, è presentato l'indice complessivo per valutare quanto il patrimonio informativo è produttivo e supporta il processo decisionale di Distribution Planning.
<b>NOTA</b>	<b>Il seguente file è compilato in modo casuale per facilitarne la compilazione</b>

## 2<sup>ND</sup> spreadsheet:

ISTRUZIONI PER LA COMPILAZIONE	
Per ogni domanda fornire una risposta sintetica, utile per la contestualizzazione del modello proposto.	
<b>Domanda 1.</b>	<b>In quale settore opera l'azienda?</b>
<b>Domanda 2.</b>	<b>In quale contesto opera l'azienda in termini di: obiettivi, linee di prodotto, mercati, clienti?</b>
<b>Domanda 3.</b>	<b>Quale dimensione ha l'azienda in termini di fatturato e di numero di dipendenti?</b>
<b>Domanda 4.</b>	<b>Qual è il posizionamento dell'azienda nella filiera e come si relaziona con fornitori e clienti diretti?</b>
<b>Domanda 5.</b>	<b>Qual è il suo ruolo all'interno dell'azienda?</b>
<b>Domanda 6.</b>	<b>Descrivere sinteticamente il Distribution planning process dell'azienda inteso come l'insieme delle attività di inventory planning e transportation planning (struttura del processo, come vengono affrontate le decisioni, ruoli coinvolti)</b>
<b>Domanda 7.</b>	<b>Indicare la familiarità con il processo decisionale di Distribution Planning</b>

3<sup>rd</sup> spreadsheet:

ISTRUZIONI PER LA COMPILAZIONE			
<p>Di seguito è presentata una lista di aree di dato ritenute interessanti per il processo decisionale di <i>Distribution Planning</i> . In questa prima parte è richiesto di valutare la rilevanza assegnando un peso a</p> <p><b>Nota: La valutazione assegnata della rilevanza deve essere data indipendente da quali dati l'azienda effettivamente colleziona e utilizza.</b></p>			
<b>Rilevanza dei dati</b>			
<p><i>Relevance si riferisce alla misura in cui i dati sono compatibili con il tema analizzato e quanto sono utili per supportare lo specifico processo decisionale.</i></p>	<p><b>Quanto sono utili e rilevanti i dati per supportare lo specifico processo decisionale? Assegnare un punteggio da 1 a 5.</b></p>	1 = Dato non significativo	
		2 = Dato poco significativo	
		3 = Dato mediamente significativo	
		4 = Dato altamente significativo	
		5 = Dato strettamente necessario	
<b>Lista dati interessanti per il Distribution Planning Process</b>			<b>Peso</b>
<i>Dati descrittivi (prodotto-network)</i>	Caratteristiche merceologiche del prodotto (ex. categoria, linea, taglia, colore, ciclo di vita, deprezzamento ) e caratteristiche fisiche del prodotto (ex. imballo, peso)		
	Caratteristiche del prodotto che influenzano lead time (ex. requisiti doganali specifici del prodotto)		
	Caratteristiche e tipologia della rete distributiva (ex. nodi, capacità) e sua evoluzione (ex. nuove aperture, nuovi mercati)		
	Tempi di trasporto per tratta e frequenza possibile di trasporto (ex. piano voli)		
	Tempi totali per tratta (ex. inclusi step amministrativi, dogane) lead time totale door-to-door, con variabilità		
	Costo di trasporto per tratta (considerando le diverse modalità di trasporto)		
<b>Dati di flusso e Scorte</b>	Stato scorte a destinazione (disponibili, prenotate, vendibili, riservate, difettato)		
	Stato scorte in transito, ordini in lavorazione non ancora in transito (wip)		
	Stato ordini in arrivo, in backlog o inevasi		

<i>Dati di Mercato</i>	Previsione domanda del cliente (puntuale, media e dispersione)		
	SLA (ex. livello di servizio obiettivo, target availability rate, target stock level)		
	Calendario lavorativo a destinazione (festività), turni di ricezione merce a destinazione		
	Calendario eventi promozionali		
	Possibili Prodotti sostituti o sostituibili		

4<sup>th</sup> spreadsheet:

**ISTRUZIONI PER LA COMPILAZIONE**

In questa sezione viene valutata la Disponibilità (Availability) di ciascun dato attraverso diverse dimensioni. Ogni dimensione ha 5 livelli di maturità, per ognuna inserire una X nel livello di maturità che meglio rappresenta la situazione di ciascun dato. Esempio: per i dati *Caratteristiche merceologiche del prodotto* la frase che meglio rappresenta l'*Existence* è quella del livello 4 (4 = la maggior parte dei dati vengono generati e raccolti). Quindi la casella F13 dovrà essere contrassegnata con una X.

Dimensione	Definizione	Valutazione
<b>Existence</b>	<b>Existence</b> si riferisce alla misura in cui i dati sono effettivamente generati e collezionati.	1. I dati <b>non vengono raccolti</b> . 2. I dati vengono generati e raccolti in <b>piccola parte</b> . 3. I dati vengono <b>mediamente generati</b> e raccolti. 4. La <b>maggior parte</b> dei dati vengono generati e raccolti. 5. I dati sono <b>completamente generati</b> e collezionati.
<b>Accessibility &amp; Authorization</b>	<b>Authorization</b> si riferisce alla misura in cui un individuo o un'organizzazione ha il diritto di utilizzare un dato. <b>Accessibility</b> si riferisce alla misura in cui è difficile ottenere i dati.	1. I dati <b>non sono accessibili</b> . 2. I dati sono <b>accessibili solo in parte non disponendo dell'autorizzazione</b> necessaria. 3. I dati sono <b>accessibili</b> tramite un <b>processo di autorizzazione lungo e oneroso</b> , in quanto sono posseduti e controllati da un altro dipartimento. 4. I dati sono <b>accessibili semplicemente</b> , previa <b>autorizzazione</b> dal dipartimento che li possiede e gestisce. 5. I dati sono <b>velocemente accessibili a tutti</b> all'interno dell'organizzazione.
<b>Timeliness</b>	<b>Timeliness</b> si riferisce alla misura in cui i dati sono regolarmente aggiornati e se arrivano in tempo per rispondere a uno specifico task.	1. I dati <b>non sono aggiornati</b> . 2. I dati <b>non sono aggiornati in maniera regolare</b> , quindi non sono appropriati per supportare il determinato processo decisionale. 3. I dati sono <b>aggiornati occasionalmente</b> ; l'intervallo di tempo che intercorre tra la generazione e il trattamento dei dati non soddisfa pienamente i requisiti. 4. I dati sono <b>più che sufficientemente aggiornati e attuali</b> per supportare il determinato processo decisionale. 5. I dati sono il <b>più possibile attuali e regolarmente aggiornati</b> ; minimizzando al minimo il tempo che intercorre tra generazione e acquisizione dei dati e il loro utilizzo/trattamento.

<p><b>Accuracy &amp; Credibility</b></p>	<p><b>Accuracy</b> si riferisce alla misura in cui dati sono corretti (free-of-error). <b>Credibility</b> si riferisce alla misura in cui le fonti da cui provengono i dati sono controllate, certificate e affidabili.</p>	<ol style="list-style-type: none"> <li>1. I dati sono <b>completamente incorretti</b> e considerati <b>inaffidabili</b>.</li> <li>2. I dati presentano <b>numerosi errori</b> dovuti alla collezione manuale, possibile manipolazione, fonti non certificate.</li> <li>3. I dati sono <b>parzialmente corretti</b> e provengono da <b>fonti</b> che possono essere <b>considerate attendibili</b>.</li> <li>4. La <b>maggior parte</b> dei dati sono <b>corretti</b> e credibili. <b>Tutte le fonti</b> sono certificate e <b>attendibili</b>.</li> <li>5. I dati sono <b>altamente affidabili</b>: sono collezionati automaticamente da <b>fonti specializzate e autenticate</b>; esperti regolarmente revisionano e controllano l'effettiva correttezza dei dati.</li> </ol>
<p><b>Completeness</b></p>	<p><b>Completeness</b> si riferisce alla misura in cui nessun valore è mancante e i dati hanno una sufficiente profondità, larghezza ed estensione per uno specifico task.</p>	<ol style="list-style-type: none"> <li>1. I dati sono <b>incompleti</b> per prendere una determinata decisione.</li> <li>2. I dati <b>in piccola parte</b> sono considerati completi per prendere una determinata decisione.</li> <li>3. I dati sono considerati <b>sufficientemente completi</b> per prendere una determinata decisione, nonostante presentino <b>numerosi valori mancanti</b>.</li> <li>4. I dati <b>sono adeguatamente completi</b>, supportando in modo appropriato lo specifico processo decisionale.</li> <li>5. I dati sono <b>completi e nessun valore è mancante</b>, presentando tutte le parti richieste per la descrizione di un'entità.</li> </ol>
<p><b>Presentation Quality &amp; Consistency</b></p>	<p><b>Presentation quality</b> si riferisce alla misura in cui i dati sono chiari, comprensibili e rappresentati in maniera compatta. <b>Consistency</b> si riferisce alla misura in cui i dati sono presentati nello stesso formato nel tempo.</p>	<ol style="list-style-type: none"> <li>1. I dati <b>non sono assolutamente interpretabili</b>, in quanto non sono rappresentati in maniera chiara e concisa.</li> <li>2. I dati sono di <b>difficile interpretazione</b>, provocando <b>diverse ambiguità di significato</b> e il formato di rappresentazione non è costante nel tempo.</li> <li>3. I dati sono presentati in maniera <b>mediamente comprensibile ed interpretabile</b>, presentando comunque discrepanze di formato nel tempo.</li> <li>4. I dati sono <b>facilmente comprensibili e interpretabili</b>, presentando qualche <b>discrepanze del formato</b> nel tempo.</li> <li>5. I dati sono presentati in <b>maniera assolutamente chiara, concisa, comprensibile, compatta</b> e non presentano ambiguità di formato nel tempo.</li> </ol>
<p><b>Data usage</b></p>	<p><b>Data usage</b> si riferisce alla misura in cui i dati sono effettivamente utilizzati per rispondere allo specifico processo decisionale.</p>	<ol style="list-style-type: none"> <li>1. I dati <b>non vengono utilizzati</b> per supportare lo specifico processo decisionale.</li> <li>2. Una <b>piccola parte</b> dei dati è utilizzata per supportare lo specifico processo decisionale.</li> <li>3. I <b>dati vengono mediamente utilizzati</b> per supportare lo specifico processo decisionale.</li> <li>4. La <b>maggior parte</b> dei dati sono utilizzati per supportare lo specifico processo decisionale.</li> <li>5. I dati sono <b>completamente utilizzati</b> per supportare lo specifico processo decisionale.</li> </ol>

5<sup>th</sup> spreadsheet:

ISTRUZIONI PER LA COMPILAZIONE		
<p>In questa sezione vengono valutate le DSS-Performances attraverso tre dimensioni. Compilare la seguente parte considerando <u>L'INSIEME DEI DATI GENERATI E COLLEZIONATI</u> utili per il Distribution planning process. Ogni dimensione è valutata attraverso 5 livelli di maturità; contrassegnare con una X (nella casella colorata) la frase che meglio rappresenta la situazione attuale dei dati (generati e utili per il Distribution Planning process).</p>		
Dimensione	Definizione	Valutazione
<b>Data Integration</b>	Data Integration si riferisce al livello di integrazione dei dati che provengono da molteplici fonti (es. differenti databases, sensori).	<p>1. I moduli del Sistema Informativo (SI) <b>non sono assolutamente in grado di interfacciarsi</b> tra loro non permettendo quindi un integrazione dei dati</p> <p>2. Sono <b>altamente possibili problemi di comunicazione e di interfaccia</b> tra i diversi moduli del SI, a livello di tecnologie, processi e interazione degli individui</p> <p>3. I moduli del SI sono <b>mediamente in grado di interfacciarsi</b> tra di loro, presentando <b>problemi di aggiornamento</b> dei dati <b>tra i diversi gruppi funzionali</b>.</p> <p>4. Il SI assicura ai gruppi funzionali una <b>buona comunicazione e interoperabilità</b>.</p> <p>5. Il SI dell'intera organizzazione è <b>scalabile e flessibile</b>, consente <b>grande interoperabilità</b> tra i gruppi funzionali, armonizzando rapporti e comunicazioni</p>
<b>Data Processing</b>	Data-processing si riferisce a quei dati che (1) sono trattati e processati da algoritmi in modo da sfruttare il loro potenziale e (2) all'abilità di analizzare i dati su diverse dimensioni tramite database analitici, aggregando, disaggregando e presentando i dati attraverso indicatori per meglio supportare il decisore.	<p>1. I dati <b>non sono analizzati</b>.</p> <p>2. I dati sono <b>in parte analizzati</b> e presentati sottoforma di <b>indicatori (ex. KPI) più o meno significativi</b>.</p> <p>3. I dati sono per la <b>maggior parte analizzati</b> e presentati in modi differenti attraverso <b>indicatori significativi (eg KPI)</b>, aggregandoli e disaggregandoli.</p> <p>4. I dati strutturati e non strutturati vengono <b>processati attraverso tecniche di analisi avanzate e algoritmi</b> in grado di <b>identificare pattern nascosti</b>.</p> <p>5. Le decisioni sono supportate da <b>algoritmi altamente innovativi e avanzati, adottando ad esempio soluzioni di machine learning e intelligenza artificiale</b>, che processano dati strutturati e non strutturati.</p>

<b>Supporting Infrastructure</b>	Supporting infrastructure si riferisce al livello di adeguatezza della sottostante infrastruttura (hardware e software) che permette di supportare le tecniche avanzate di data-analysis come: data mining, computer based modelling e simulazione. I dati devono essere processati su infrastrutture (hardware e software) altamente performanti e da algoritmi efficienti con lo scopo di processare in tempo i dati per il processo decisionale sotto analisi.	1. L'infrastruttura sottostante (hardware e software) <b>non è in grado di fornire supporto e processare dati.</b>
		2. L'infrastruttura sottostante (hardware e software) permette <b>analisi descrittive</b> esclusivamente su <b>dati strutturati</b> .
		3. L'infrastruttura sottostante (hardware e software) è in grado di processare <b>grandi moli di dati strutturati</b> , performando <b>analisi descrittive and predictive</b> richiedendo un lasso di <b>tempo significativo</b> .
		4. L'infrastruttura è in grado di <b>processare dati strutturati e non strutturati</b> attraverso un ampio range di tecniche analitiche ( <b>analisi descrittive, predittive</b> ) in un <b>lasso di tempo</b> ritenuto <b>ragionevole</b> per supportare la decisione.
		5. L'infrastruttura è capace di <b>processare grandi moli di dati strutturati e non strutturati attraverso tecniche di analisi innovative e avanzate</b> , è altamente performante in termini di tempo (processa i dati in modo efficace ed efficiente), si adatta <b>velocemente</b> al cambiamento.

### 6<sup>th</sup> Spreadsheet:

DATA AVAILABILITY 100,00%		DATA QUALITY 100,00%		DSS PERFORMANCE 100,00%	
Data Existence	100,00%	Data Accuracy & Credibility	100,00%	Data Processing	100,00%
Data Accessibility & Authorization	100,00%	Data Completeness	100,00%	Data Integration	100,00%
Data Timeliness	100,00%	Data Presentation Quality & Consistency	100,00%	Supporting infrastructure	100,00%
				Data Usage	100,00%
<b>Data Productivity Index</b>					
<b>100,00%</b>					

## A2 Email Template to contact the selected companies

Gentilissimo,

Io e la mia collega per il nostro lavoro di tesi stiamo sviluppando un modello con lo scopo di misurare la produttività e il supporto dato ai processi decisionali dai dati raccolti in un'azienda. In questo momento stiamo ultimando il nostro strumento e ci piacerebbe poterlo testare nella vostra realtà. In particolare, vorremmo analizzare come i dati supportano le decisioni nel processo di Distribution Planning: inteso come ottimizzazione dei trasporti (outbound) e delle scorte.

Il questionario è diretto ad aziende manifatturiere caratterizzate da processi strutturati e monitorati attraverso l'uso dei dati. È preferibile che il questionario venga compilato da una persona che si occupa di decisioni in ambito Distribuito Planning.

L'applicazione del modello consta delle seguenti parti:

- informazioni generali sull'azienda e sul processo di Distribution Planning
- viene fornita preventivamente una lista di dati aggregati ritenuti interessanti per il processo decisionale (ex. Caratteristiche merceologiche del prodotto)
- ogni dato della lista sarà valutato attraverso un questionario strutturato secondo diversi aspetti (è richiesto di mettere una X nella frase che meglio rappresenta lo stato attuale del dato)

Esempio:

<b>Accuracy &amp; Credibility</b>	1. I dati sono <b>completamente incorretti</b> e considerati <b>inaffidabili</b>	
	2. I dati presentano <b>numerosi errori</b> dovuti alla collezione manuale, possibile manipolazione, fonti non certificate.	
	3. I dati sono <b>parzialmente corretti</b> e provengono da <b>fonti</b> che possono essere <b>considerate attendibili</b>	X
	4. La <b>maggior parte</b> dei dati sono <b>corretti</b> e credibili. <b>Tutte le fonti</b> sono certificate e <b>attendibili</b> .	
	5. I dati sono <b>altamente affidabili</b> : sono collezionati automaticamente da <b>fonti specializzate e autenticate</b> ; esperti regolarmente revisionano e controllano l'effettiva correttezza dei dati.	

Non saranno richiesti dati, ma valutazioni soggettive al fine di validare il modello.

Se è disponibile possiamo fissare un appuntamento, altrimenti possiamo inviarle tramite mail il file Excel da compilare e sentirci telefonicamente durante la compilazione o a posteriori.

Sono ben accetti consigli e critiche, considerato che siamo in fase di validazione del modello.