Expertise
and insight
for the future

Hans Kanerva

Data Management Plan for the SGTS Nordic & Baltic PC Factory

Metropolia University of Applied Sciences

Master's Degree

Degree Programme in Business Informatics

Master's Thesis

30.11.2021

Metropolia
University of Applied Sciences

Author	Hans Kanerva
Title	Data Management Plan for the SGTS Nordic & Baltic PC Factory
Number of Pages	63 pages + 8 appendices
Date	30 November 2021
Degree	Master of Business Administration
Degree Programme	Business Informatics
Instructor	Antti Hovi, Senior Lecturer

The purpose of this study was to create a data management plan, which would address issues related to data management at the PC Factory activity within the Saint-Gobain Technology Services Nordic & Baltic (SGTS NB) organization. SGTS NB is an internal IT-service organization within the global enterprise Saint-Gobain, which deliver infrastructure services for the local Saint-Gobain companies in the Nordic & Baltic region and the PC Factory is a dedicated SGTS NB team, focused on delivering premium PC life-cycle services. The need for data management was identified before the study, but the necessary means and tools were missing. It was decided that the study was needed to find a solution for the issues faced and hence was the study commissioned by SGTS NB.

The current state of the PC Factory data management was investigated by using both qualitative and quantitative methods. The qualitative research consisted of six in-depth interviews with relevant stakeholders while the quantitative research consisted of a customer survey, which was addressed to a total of forty-six recipients. Issues were found in four areas of the current data management process, and the issues were related to missing skills, policies, reports, and current heavy manual processes. Best practices were found from available knowledge which identified the means and tools for solving the issues identified during the current state analysis phase.

The outcome of this thesis is a data management plan, based on the best practice, which address the data management issues within the PC Factory. The data management plan is targeted for a specific business problem, but it is written in a way which makes it possible to replicate, with modifications, to other parts of the organization. The data management plan addresses the six activities discovered during the analysis of the best practices, in addition to the data maturity assessment, which as a whole cover all aspects of the activities needed for working data management. A detailed action plan is prepared, as an appendix to the data management plan, to ensure a successful implementation of the data management plan.

It is agreed within SGTS NB that the data management plan will be facilitated within the PC Factory, as the benefits for the organization are clear and beneficial. The work will start in Q1-2022 and is estimated to be completed by the end of year 2022. By implementing the data management plan the PC Factory will be in control of data, will be able to utilize data better, and will be able to continue the development of data analytics to support the organization digital transformation in the future by continuing to build upon the data maturity presented in this thesis.

Keywords	Data management, Data Management plan, Data analytics, Data

Contents

List of Figures

1 Introduction

Data is digitalized information which is presented to any user through a data source. A data source can be any form of source providing the data as files, databases, web or streaming data (Talend, 2020). The various types of data and data sources are complicated to manage, and this is where the data management plan will assist the company to capture and help to utilize the data.

This thesis studies data management and develops a data management plan to help the case IT department to get in control of the data.

1.1 Business Context

The case company of this thesis is Saint-Gobain Technology Services Nordic & Baltic (SGTS NB). Saint-Gobain is a multi-national corporate which has activity in 70 countries with 171.000 employees. Saint-Gobain operates in multiple industries with a focus on wellbeing of people. As presented in its website,

> "Saint-Gobain designs, manufactures and distributes materials and solutions which are key ingredients in the wellbeing of each of us and the future of all. They can be found everywhere in our living places and our daily life: in buildings, transportation, infrastructure and in many industrial applications. They provide comfort, performance and safety while addressing the challenges of sustainable construction, resource efficiency and climate change" (Saint-Gobain, 2020).

The regional scope of Saint-Gobain Technology Services Nordic & Baltic (SGTS NB) covers seven countries including Denmark, Estonia, Finland, Latvia, Lithuania, Norway, and Sweden. The corporate structure consists of 40 legal companies spread all over the seven countries whereas the holding company Saint-Gobain Nordic A/S is in Denmark. Industries covered in the region are construction products, flat & automotive glass, and abrasives. Saint-Gobain has a total of 650 sites in the region and these sites consists of offices, plants, shops, and warehouses. The Group has a total of approx. 13.000 employees in the region using over 10.000 PCs.

The case organization of this thesis is an internal IT organization delivering IT infrastructure services for Saint-Gobain Technology Services Nordic & Baltic (SGTS NB). The PC Factory is a SGTS NB internal service delivery team, and it is belonging to the User & Site Support department providing IT support services to the customers. The PC Factory's role is to deliver premium PC related life-cycle services for the SG entities. PC life-cycle services include PC purchasing & invoicing, management of a local PC stock, PC imaging & installation, PC logistics, IT equipment recycling and reporting.

1.2 Business Challenge, Objective and Outcome

Currently, the services by the case company are delivered by using three suppliers covering equipment manufacturing, logistics and recycling services. There are also two internal departments supporting the activity which are Finance and the ERP Competency center. The case company is providing tools and systems for the service delivery. All these different stakeholders generate data within their systems and services which is provided to the SGTS NB´s PC Factory in various formats.

The challenge faced by the case company is unidentified and unused data. Lack of control of the data means that the business does not know what data is available and from where it is generated which has business consequences. The lack of data control leads to the lack of insight during decision making that, in turn, might lead to poor decisions or even decisions made on false information or individual assumptions. Detailed reporting – that is currently missing – would support the decision making and give transparency for decision makers, internal customers, and internal auditors when the activity is audited. Unidentified data means that there are data sources which are generating data, but the data isn't efficiently used. The consequence of this is that data is often extracted ad-hoc and treated manually which means that a lot of manual work in the process and leads easily to human errors.

To fix this problem, the case organization needs to better control the data, which is not possible without a proper data management plan. Such a plan needs to be developed in order to stay in control of each part and detail of the delivered services and comply with the case company´s rules and guidelines but also to be transparent toward the decision makers and internal customers, so that they can rely on the internal service provider.

Accordingly, the objective of this thesis is *to develop a Data Management Plan for reaching an acceptable level of descriptive data analytics*. The outcome of the thesis is a *Data Management Plan*.

1.3 Thesis Outline

The scope of this thesis is to develop a data management plan for reaching an acceptable level of descriptive data analytics for SGTS NB´s PC Factory. This thesis will cover the research of practices in use but respecting the case company´s internal rules and guidelines. For example, the case company restricts usage of certain solutions, systems, or processes due to security risks and this must be respected. The fundamentals of data management will be introduced in this thesis and the final data management plan will be built upon the findings taking the restrictions into account.

This thesis is written in seven sections. Section 1 is the introduction giving the background and context to the thesis. Methods and materials are described in Section 2. Section 3 is dedicated for the current state analysis of data management and data use in the case company. Existing knowledge and best practice are presented in Section 4. Section 5 presents an initial data management plan is co-created with the key stakeholders based on the findings from Sections 3 and 4. The data management plan is presented to the company stakeholders for validation in Section 6. Section 7 is dedicated to an executive summary and wrap up of the thesis, but here are also next steps and recommendations presented.

2 Method and Material

This section describes the research approach and research design, and data collection methods used in the study.

2.1 Research Approach

The research approach and methods tell the reader how the research is carried out. There might be different answers to the question what research is. In the end, research is about making diligent searches, inquiries, investigations or even experiments to discover new facts or findings (Adams et al., 2014: 1-2).

First, in terms of the research family, *applied research* differs from *basic* research as it used to answer specific question which might be a real question, problem, reasons for failure or success, or gaining a better understanding of the topic (Sreejesh et al., 2014). Applied research is defined by Adams et al. (2014) as follows: “Applied research is conducted when a decision must be made about a specific real-life problem” (Adams et al., 2014: 7).

Second, there are also different research methods to choose from, fitting different needs and targeting different research problems. *Quantitative* research methods are used for quantitative measurement which means usage of statistical analysis and metrics. Collection of data can be done through surveys and questionnaires. *Qualitative* research methods are based on qualitative data collection and non-numerical analysis of data aimed at the exploration and understanding of how the respondent experiences reality (Adams et al., 2014: 6). Even a mix of these research methods can be used to get a better and wider research result.

Third, as for the research strategy, case studies and action research are the most typical strategies for business research. Both, case studies and action research aim at exploring contexts and stakeholders, and participants points of views, or doing observations in the natural circumstances of the case which is being studied. For case studies, as well as for action researchers, it is important to interpret the data in the context related to the case (Simons, 2014: 5). One of the most typical research techniques for both the case study and action research is an interview. Interviews, especially semi-structured and

open-end interviews, give room for both questions and open discussion and pragmatic elaboration on topics instead of keeping to a structured interview (Leavy, 2014: 286).

In this thesis, a qualitative case study was selected as the research approach as this research approach fits the subject of the thesis. Qualitative data will give the necessary input for identifying the problem, while the case study approach will also ensure the use of theory on how to solve it. This case study will take other cases in consideration during the research phase. Interviews will be the primary data source for this thesis, especially semi-structured interviews. The interviews will be documented using field notes. Additionally, the thesis will organize a workshop with stakeholders and a customer survey in relation to investigating and improving data management.

2.2 Research Design

The research design of this thesis consists of five steps. Figure 1 also shows the data sources and outcomes in each step.

As shown in Figure 1, the thesis contains five consecutive steps, and the steps are shown in order of the research process in the thesis where the previous step builds the foundation for the following step. The research design steps are supported by three different data collection phases where each data collection phase is named Data 1, Data 2, or Data 3. These data collection phases are inputs for the steps which then generate the outputs, as shown in Figure 1 below.

Figure 1 below shows the research design of the study.

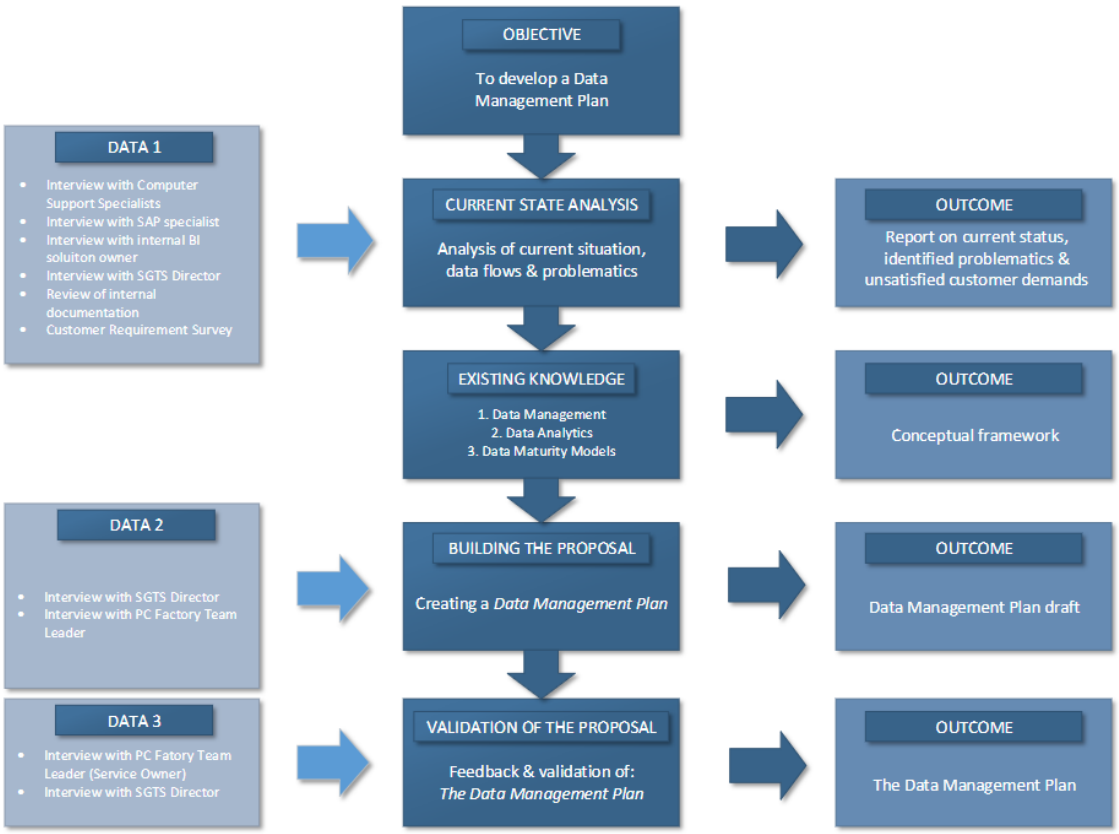


Figure 1. Research design of the thesis.

As seen from Figure 1, *the first step* sets the study objective after which the current state is analyzed in *step two*. The current state analysis is based on Data 1 collection which mainly consists of stakeholder interviews but also includes a customer survey as well as the internal documentation review. The intention with this step is to identify the current problems, identify where the department is lacking data management, and unsatisfied customer demand.

The third step of the research design is allocated for exploring existing knowledge. It looked into the knowledge and best practice currently available on developing data management plans and selected those elements that are most relevant for the study and can possibly be included in building the data management plan. The outcome of the third step is a conceptual framework for building a data management plan.

The fourth step is focused on building of the data management plan proposal. Data 2 collection, together with the outcome of step 3, will generate input for the creation of the proposal and the outcome of step four will be the initial draft of the data management plan. Data 2 collection will include interviews documented in field notes. The proposal is

based on both, the results of the current sate analysis as well as selected best practice and input collected from the organization stakeholders and employees.

The fifth step is the final step of the research design. Step 5 is validation of the proposal, meaning validation of the data management plan. Data 3 collection will include interviews with key stakeholders. These validations are conducted by interviews and tracked by field notes, in the same manner as Data 1 and Data 2 collections. The data management plan is validated by the key stakeholders who approve the data management plan, otherwise final modifications and updates are conducted based on the feedback during Data 3 interviews.

2.3 Data Collection and Analysis

Table 1 shows details of Data collections 1-3 used in this study.

Table 1. Details of data collection phases Data 1-3 used in this study.

#	Data Type	Data Source	Topic; Description	Date	Length	Documented As	Recording URL
Data 1: Current State Analysis							
1	Workshop	PC Factory team	Data source identification	22/03/2021	30 Min	Field Notes/Recording	https://web.microsoftstream.com/video/b058626a-7d7c-47b0-a315-d553effc8e7c
2	Interview	SAP FICO Team Leader	SAP data & SAP data extraction	24/03/2021	40 Min	Field Notes/Recording	https://web.microsoftstream.com/video/08bafa10-38ed-4963-84a0-53ce7a1af02c
3	Interview	Computer Support Specialist	PC Factory data management needs & current processes	22/03/2021	30 Min	Field Notes	
4	Interview	Junior Computer Support Specialist	PC Factory data management needs & current processes	22/03/2021	30 Min	Field Notes	
5	Interview	PC Factory Team Leader	PC Factory data management needs & current processes	29/03/2021	30 Min	Field Notes	
6	Interview	SGTS NB PowerBI solution owner	Current solution data import & reporting possibilities	25/03/2021	45 Min	Field Notes/Recording	https://web.microsoftstream.com/video/8904ab77-9742-4140-92e8-3a0a28209217
7	Interview	SGTS Director	Region management requirements towards PC Factory	16/04/2021	30 Min	Field Notes	
8	Questionnaire	PC Factory Customers	Customer current state feedback, new requirements & expectations	15/04/2021	-	Questionnaire Analysis Notes	
9	Internal Documentation	Review	Current state reporting review	16/04/2021	60 min	Field Notes	
Data 2: Building the Proposal							
10	Interview	SGTS Director	Region management requirements towards the DMP	01/11/2021	30 Min	Field Notes	
11	Interview	PC Factory Team Leader	Internal requirements for data management & data analytics	09/11/2021	30 Min	Field Notes	
Data 3: Validation of the Proposal							
12	Interview	SGTS Director	Region management validation & feedback of the proposal	26/11/2021	30 Min	Field Notes	
13	Interview	PC Factory Team Leader	Internal validation & feedback of the proposal	29/11/2021	30 Min	Field Notes	

As Table 1 shows, there were three main data collections stages in the thesis: *Data 1*, *Data 2*, and *Data 3*. *Data 1* collected the data for the current state analysis phase, *Data 2* for building the proposal, and *Data 3* for the validation and building the final proposal.

In *Data 1*, various stakeholders in different positions and organizations were interviewed: Team Leader from the SAP Competence Center, the current PowerBI solution owner

within the own organization, as well as all internal staff within the PC Factory team, including the PC Factory Team Leader. Each person interviewed had are participants or stakeholders in the service delivery of the PC Factory. In addition, a customer questionnaire was launched to collect customer feedback. Due to the internal constrains, the same questionnaire was used to collect new requirements which were used in *Data 2* phase. Internal documents were also reviewed to better understand the current state and current data management processes. The customer questionnaire was sent to the customers to collect customer feedback.

In the following data collection round, *Data 2* was collected data about new or missing requirements in addition to collecting input and feedback about the initial proposal. The primary method of data collection was interviews. The stakeholders interviewed in this round included: the SGTS NB Director to collect the region management requirements, and the PC Factory Team Leader to collect the internal PC Factory requirements. Data 2 was a crucial round to get the final requirements identified and documented as well as getting input and feedback on the initial proposal.

Data 3 was collected in the third and final round of the study. This round was used for collecting the final validation and approval of the proposed data management plan. The validation also contained discussions about the implementation and possible consequences on the organization. This round also identified final improvement suggestion by the key stakeholders interviewed for validation. In this round, the stakeholders interviewed included: the SGTS NB Director to collect the region management, and the PC Factory Team Leader to collect the internal PC Factory validations and approvals.

Thus, interviews made the primary method of data collection in this thesis. Preparation for the interviews included preparing questions and topics to be discussed still leaving room for additional questions and discussion. The interviews were conducted as semi-structured by telephone or Teams conferencing sessions. The current COVID-19 situation forced the interviews to be conducted by other means than face-2-face interviews. All interviews were documented in field notes. The questions and field notes for Data 1 interviews can be found in Appendix 1, questions and field notes for Data 2 in Appendix 4, and questions and field notes for Data 3 in Appendix 5.

The aim of the interviews was to collect information about the data generated by the different sources but also to identify possible new data that could be populated and be of benefit for the PC Factory. Suppliers were persuaded to talk about other customers' needs and generated data to find possible inspirations. The interviewer steered the interviews and asked additional questions if there was a need for it where the additional questions and answers were noted during the interviews.

In addition, the internal PC Factory workshop and the customer questionnaire supported the main methods of data collection. The workshop was conducted on-site in the PC Factory and the time was used efficiently to identify the different data sources used by the PC Factory. The field notes from the workshop can be found in Appendix 3. The questionnaire was chosen as a data collection method because of a large customer base which would have been difficult to interview in the timeframe of the thesis. The questionnaire answers were analyzed and used in Data 1 but was also utilized in Data 2 as inputs for the proposal. The questionnaire, answer data and analysis are found in Appendix 2.

Finally, this study included analysis of a number of internal documents. The documents included in the study are listed in Table 2.

Table 2. Internal documents used in the current state analysis, Data 1.

#	Name of document	Number of pages/Other Content	Description
A	SAINT-GOBAIN FINLAND OY 03-2021 Pulse-Diagostics-v1.xls	5 sheets	DHL delivery report
B	CustomReport20210412130002-10167.html	1 page	Lenovo delivery report
C	KO-99926566 – Saint-Gobain Finland Oy – Finland – Ojakkala - Report.xls	6 sheets	Tier1Asset recycling report

As shown in Table 2, these documents were analyzed for the Data 1 round, the current state analysis, to get an understanding of how the data is currently presented and utilized in the daily PC Factory activity.

The textual data was analyzed using Thematic/ content analysis. All primary collected data was documented in recordings and notes. Assumptions were left out of the analysis and all data can be found Appendixes 1 to 5 which are including the data collection.

2.4 Validity and Reliability

To ensure quality in research, various criteria can be selected and focused on by researchers. Most typically, in case studies, following Yin (2004, 2009), validity and reliability are discussed as research criteria. These criteria can be further divided into construct validity, internal validity and external validity (Yin, 2009: 40-44). *Construct validity* is identification of the correct operational measures for the concepts being studied (Yin, 2009: 40). To ensure the *construct validity* of the thesis, several methods of data collection (e.g. triangulation) were used based on several data sources. The data collection methods used were interviews, a questionnaire, and a workshop. The interviews were conducted with a broad selection of internal and external stakeholders; also, the internal team was reached via a common workshop, and the customers were addressed through a survey. The internal validity can only be validated after the study and final plan are complete. *External validity* is about knowing if the study's findings can be generalized (Yin, 2009: 43). In this study the final data management plan can be utilized in other domains of the organization and therefore the external validity can be considered as positive. All of these measures were aimed to ensure the validity of the study.

Reliability, according to Yin (2009: 40) requires that the research steps in a study can be repeated with the same results. In this study, the conceptual framework is based on the current best practice and data collected from current employees and stakeholders. The study is considered reliable at the time of conducting the study but it will be outdated in the near future due to the emerging development of IT in general.

Next, the results of the current state analysis, the biggest part of the data collection, are presented and discussed in Section 3 of this study.

3 Current State Analysis of Data Management in SGTS NB PC Factory

In this section of the thesis the current state of data management within the case company is presented and analyzed. The focus is on understanding the data knowledge, utilization, capabilities, and possibilities within the case company and to identify strengths and weaknesses in the current way of working.

3.1 Overview of the Current State Analysis

The goal of the current state analysis was to identify the current state of data management within the case company. Interviews were the primary data collection sources. Also, internal team workshop and customer questionnaire were used as data sources with the purpose of identifying customer input. This data collection ensured enough data to complete a thorough analysis for better understanding of the current state. The current state analysis was conducted in six steps.

First, the internal workshop was conducted virtually over Microsoft Teams due to the ongoing pandemic situation. The invitations were sent two weeks prior to the session with the agenda of the workshop and templates prepared before the session. This approach ensured an effective workshop focusing on completing the data source information in data source cards. The workshop was recorded and documented in field notes, to ensure that nothing was missed during the current state analysis.

Second, the interviews with the internal staff, including employees and team leader, were conducted to identify the current data management state. Weaknesses, strengths, opportunities, and threats were identified. The interviews of the ERP system team leader and PowerBI system owner ensured that details about data source alternatives and possibilities were identified. Both interviews were documented in field notes and recorded.

Third, a management interview was conducted with the SGTS Nordic & Baltic Director. The goal with this interview was to understand possible weaknesses in the data management from the management point of view.

Fourth, a customer questionnaire was conducted to identify the customer experience and expectations feedback. Weaknesses in the customer interaction were identified and

noted. This was essential as the data management is also linked to customer data reporting and visualization.

Fifth, current supplier reports were analyzed to identify possible lack of data and future development possibilities.

Finally, the strengths and weaknesses of the current data management practices were identified and summarized.

3.2 Description of Current Data Management Practices

The interviews of the internal staff gave a picture of the current data management process. This high-level process is visualized in Figure 2.

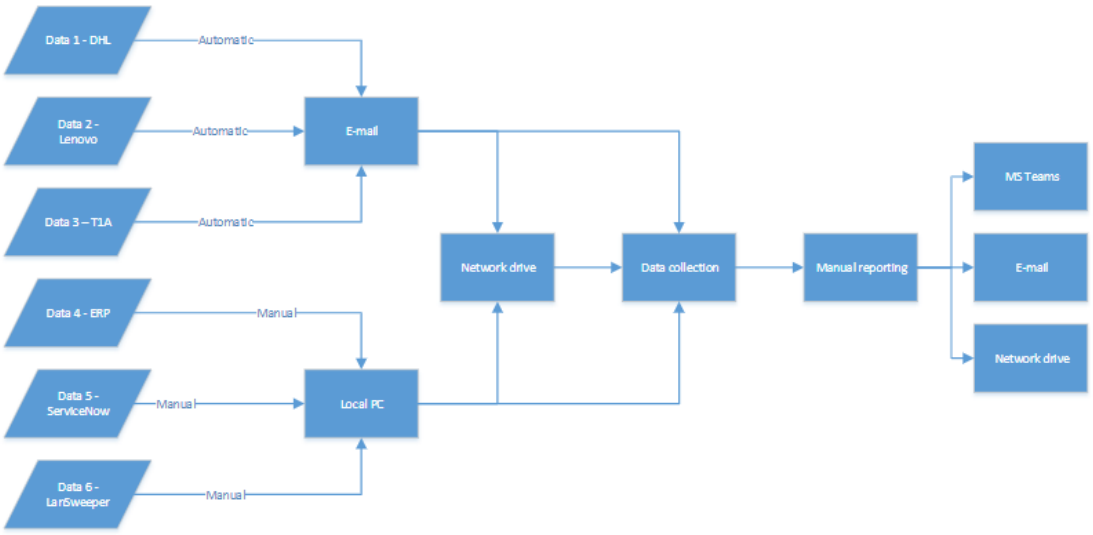


Figure 2. Current data management process in the SGTS Nordic & Baltic PC Factory

As seen from Figure 2, six data sources were identified during the workshop. The six data sources are:

1. Logistics data from *DHL*

2. PC supply & delivery data from *Lenovo*

3. PC recycling data from *Tier1Asset*
4. *ERP* data from SAP
5. Ticket data from *ServiceNow*
6. Asset data from *LanSweeper*

Three out of six data sources, which are *DHL, Lenovo,* and *Tier1Asset*, are generating data and reports automatically while the data is received by e-mail. The remaining three data sources, which are *ERP, ServiceNow,* and *LanSweeper*, content is either populated or generated manually with a need for manual storage operations.

The data storage location isn't agreed, as common internal practices aren't agreed, which has led to those three different storage locations have been identified as shown in Figure 2.

1. The team's common e-mail box
2. The team's common network drive
3. The employees' own PCs.

The needed data is manually retrieved by an employee from the required data storage location for further processing on a case when there is a need for data sharing, reporting or data visualization.

All reports or visualization is done manually with Microsoft Excel as the tool on ad hoc basis as the common tool PowerBI hasn't been facilitated in the complete organization. There is a lack of resources with the necessary skills and competence which can drive the internal implementation initiative. Reports and visualizations are shared with the necessary recipients, e.g., internal customers needing asset lists for budgeting by Microsoft Teams, e-mail, or by placement on a network drive. This is considered by everyone involved as an ineffective, time consuming, uncontrolled and an unreliable way of reporting and data visualization. This input can also be found in the customer questionnaire feedback by anonymous respondents when responding to the question about reason for dissatisfaction:

“Many things take too long time.” (Anonymous respondent, 2021)

The same message could also be interpreted through the customer feedback where the respondent was asked if the she or he knows where to find data:

“Yes - I ask SGTS USS Tina” (Anonymous respondent, 2021)

Thus, the current data management process can be divided in five stages: data sources, data storage, data processing, reporting, and data sharing. These stages are described in more detail below.

3.2.1 Data sources

The case company is working with six different data sources. The current data sources are well known by the case company staff. Each data source populate data which is important for the case company activity. The data format populated by each data source depends on how the data is populated. Data populated by supplier systems is in either Excel or HTML format, which is good as the date can easily be used in the following stages of the process. HTML can be replaced with Excel by changing the export settings in the supplier system. The three internal data source systems populate data in Excel format as well.

Supplier data is received automatically through e-mail delivery as scheduled reports. The internal data must be populated and exported manually or manually printed.

The storage of the data received isn’t automized which means that the data storage is sporadic, and location not agreed. One supplier report, DHL, is included in a routine where the data is manually stored from the e-mail to a specific network drive. Other suppliers´ data reports are stored in the e-mail box. Internal data is usually stored on the employee PC and occasionally moved to a common storage location like a network drive in case the data is to be shared among several persons.

Table 3. Data sources (summary from the workshop).

Data Source	Data Format	Alternative Data Format	Data Delivery	Data Storage	Format Approved	Delivery Approved	Storage Approved
DHL	Excel	N/A	E-mail	Network drive	Yes	Yes	Yes
Lenovo	HTML	Excel	E-mail	E-mail	Yes	Yes	No
Tier1Asset	Excel	N/A	E-mail	E-mail	Yes	Yes	No
ERP	Excel	N/A	Manual Export	PC	Yes	No	No
ServiceNow	Excel	N/A	Manual Export	PC	Yes	No	No
LanSweeper	Excel	N/A	Manual Export	PC	Yes	No	No

As seen in *Table 3*, all data formats of the data sources were approved by the case company during the workshop. Three data source data delivery formats as well as five data sources storage locations weren't approved by the case company.

The format of the data can be harmonized to the same format, which is Excel. The only data source, *Lenovo*, which do not have Excel as a primary is having Excel as an alternative data format. It is important that the documentation data storage location is known and documented, which isn't currently the case. Supplier reports were reviewed, validated and the current data structure was approved by the case company.

As a conclusion, the data sources are well known, and identified. Proper data source documentation is missing though. The data format is a strength as Excel is considered a good data source format which can easily be used in the case company business intelligence tool, as mentioned by Lars Rohlen during the interview on the 25th of March 2021 (Rohlen 2021).

3.2.2 Data storage

The data storage location hasn't been agreed internally within the case company. This is one of the main problems with the current data management processes. The various data storage locations have led to that data isn't available for all relevant stakeholders as mentioned by Jukka-Pekka Knuutinen and illustrated in the following quote:

> Not exactly, I know some but not all. I have some difficulties in finding the storage locations. (Knuutinen 2021)

Tuomas Paavola mentioned during the interview that the spread of data storge locations as one of the greatest challenges in the current data management process (Paavola 2021).

The data is currently stored on at least three different locations, but it remains unclear if there are even further unidentified used locations. As a conclusion, the data storage location is one of the greatest challenges in the current process.

3.2.3 Data processing

Data processing is currently a fully manual process. The person carrying out the data processing will have to collect data from different sources and process the collected data manually. The processing is done with Excel. The challenge with the current process is that it is time consuming and increase the risk of human errors.

The case company has built a business intelligence system around PowerBI, as Lars Rohlen mentioned during the interview (Rohlen 2021). The business intelligence system isn't used by the case company. The main reason for not using the business intelligence system is current knowledge about data processing which was raised by Jere Kallioniemi during the interview (Kallioniemi 2021).

3.2.4 Reporting

Reporting in the case company is about telling and visualizing what has happened within the activity, meaning presenting historical data and information. The case company doesn't currently have any structured nor automatic reporting. All other current reporting activities are done manually on an ad-hoc basis and based on the manual data processing stage outcome and specific requests. The current reporting is done primarily in Excel but there are some common reports uploaded to the PowerBI business intelligence system. It is considered that the current report content isn't sufficient as mentioned by Jörgen Liberg, financial data reporting is missing (Liberg 2021). The customer reporting is missing a structure and even some mandatory reports. This lack of current reports was mentioned by one recipient in the customer questionnaire (Appendix 2).

The case company is lacking reporting knowledge and skills which is also considered as a challenge as it is difficult to create the necessary reports without the knowledge and skills.

As a conclusion, the reporting is currently a manual process and require a lot of effort and time. The current process is therefore a very resource demanding process. The business intelligence system is existing, but the use is very limited.

3.2.5 Data sharing

The current reports are shared through Teams, by e-mail, or uploaded on a common network share. The reports located in the business intelligence system has their data manually updated in the data processing stage and then manually refreshed in the business intelligence tool. There isn't any existing agreed policy for report sharing within the case company, which has led to that the reports are shared by three different means. This has led to that the internal staff is unaware of the report location and use excessive time in trying to locate the reports. This finding was also supported by the customer questionnaire results. The common opinion collected in the interviews and customer questionnaire is that PowerBI is considered as good and sufficient tool for sharing reports and data.

3.2.6 Data Analytics Maturity

Data analytics is a word that is currently replaced with reporting within the case company and therefore is data analytics maturity an unknown term. The case company isn't currently measuring data analytics maturity at all, as this hasn't previously been raised as a topic and thus no plans for data analytics measurement is put in place.

Data analytics maturity is essential for measuring the current and target state of data analytics within the case company and the topic must therefore be added in the data management plan.

3.3 Key Findings from the Current State Analysis

As seen from the description above, the current state of data management within the case company is relying on manual and ad hoc actions. Only a few data sources populate data automatically which makes the data processing time consuming as data must be fetched, stored, extracted, manipulated, and visualized manually. More detailed conclusions are summarized below based on the SWOT analysis.

3.3.1 Strengths and Weaknesses of the Current Data Management Process

This section summarizes the strengths and weaknesses revealed based on the analysis of Data 1. Opportunities and threats were also identified to get the full overall picture of the current state. The opportunities and threats will be taken in consideration while building the proposal in Section 5. The results are summarized in a SWOT analysis matrix and presented in Figure 3.

STRENGTHS	WEAKNESSES
+ Good understanding of existing data sources + Existing company data management architecture & solution in place + Access to data sources + Data format + Data content	- Lack of internal practice of data storage - Lack of required reports - Lack of internal practice of report sharing - Lack of data processing knowledge & skills - Lack of reporting knowledge & skills
OPPORTUNITIES	**THREATS**
A Data visualization (reports / dashboards) A Support for decision-making A Satisfy customer demand by transparency A Data source connector possibilities	Δ Company security rules Δ Time & resources

Figure 3. SWOT analysis of the current data management practices.

As seen from Figure 3, the results include five strengths, five weaknesses, four opportunities, and two threats.

First, the most important strengths are the good understanding of existing data sources and a current or possible compliant data format. These strengths are the foundation for building the data management plan. Other identified strengths are current access to the data sources, modern existing data management architecture & solution, and compliant data content with the existing data management solution.

Second, as seen in Figure 3, there are several weaknesses identified. The most important weaknesses are related to the lack of knowledge and skills. The staff is lacking knowledge and skills in reporting and data processing. These weaknesses are directly impacting the way data is managed within the case company. Other important weaknesses are the lack of internal practices. Lacking practices in data storage and data sharing impact the accessibility of data. This weakness has a direct impact on the time

used by the staff for processing the data as the data must be found and accessed before the processing can start. Another weakness is the lack of required reports. Internal and external stakeholders are lacking reports which gives visibility to service and financial data. This means that the necessary service and activity transparency is lacking.

Third, four opportunities were identified in the SWOT analysis, and these opportunities are visible in Figure 3. The most important opportunities are the existing possibility for reporting using PowerBI visualization and dashboards. PowerBI can be used as support for decision-making and to satisfy the customer demand for transparency. The final opportunity, possibility of using data connectors between PowerBI and third-party systems, is an opportunity with potential but will require a lot of resources to get to work as the connector establishment will require planning, documentation, and even consultancy.

Fourth, two threats were identified and listed in Figure 3. Both identified threats are important. The case company security rules and restrictions are likely to cause issues with access to data which might have an impact on the final data management plan. Time and resources are scarce which means that the implementation of the data management plan might be delayed or even postponed unless the data management plan is considered prioritized before other projects in the pipeline.

3.3.2 Key Challenges and Business Impact

This section discusses the key challenges and their business impact which were identified during the current state analysis. The key challenges and identified business impact are listed in Table 4.

Table 4. Key Challenges and Business Impact.

Stage	Challenge	Impact
Data Storage	Data storage location unknown by the staff. No internal agreement, practice, or policy for common data storage practices.	1. Impossible to utilize the data in an efficient way as the staff can't find the data. 2. Loss of staff efficiency due to lost working time.
Data Processing	A manual data processing process.	1. Loss of staff efficiency due to lost working time. 2. Increased risk of human errors.
Data Processing	Lack of data processing knowledge and skills	1. Unutilized system and data. 2. Loss of staff efficiency due to lost working time.
Reporting	Lack of reporting knowledge and skills	1. Missing reports for both internal and external stakeholders. 2. Loss of staff efficiency due to lost working time as the staff use extra time on creating non-relevant reports.
Reporting	Lack of reporting/reports	1. Missing insight to financial performance which has an impact on the decision-making. 2. Customer transparency is missing and well as customer report for supporting the customer decision-making.
Data Sharing	No Internal agreement, practice, or policy for common data sharing practices.	1. Loss of data. 2. Loss of staff efficiency due to lost working time. 3. Impact on decision-making as necessary data is not found or inaccessible.

As can be seen from Table 4, there are five challenges listed in the *Challenge* column. Each challenge has two or three direct business impacts identified and these are listed in the *Impact* column. The stage, to which each challenge is connected, is marked in the *Stage* column.

First, the challenge related to the *Data Storage* stage has the following consequences on the business: unutilized data and requiring additional time utilization from the staff. The first impact consequence is reduction in transparency and decision-making and the second impact is unnecessary human resource utilization. An internal policy is needed for rectifying the challenge in the *Data Storage* stage. The internal policy must include the necessary framework and guidelines for data storage within the case company.

Second, the challenges related to the *Data Processing* stage is impacting the business as following: manual processing of data and inefficiency in the process due to lack of knowledge and skills. The consequence is that a lot of work hours are put on the data processing stage through manual labor, but the outcome of the stage isn't fulfilling the needs by the stakeholders. Data remain unutilized and impact the decision-making. Manual processing also increases the risk of human errors in the process. There is a great need for knowledge and skills around the data processing stage and available tools as this will ensure automation and digitalization of the process stage.

Third, the challenges related to the *Reporting* stage are like challenges related to *Data Processing* stage. The challenges, as seen in Table 4, are impacting the decision-making, transparency, and unnecessary human resource utilization. The lack of knowledge and skills has a direct impact on the other challenge, which is the lack of report. The reporting can be built to cover the stakeholder needs by building the necessary competence within the case company.

Fourth, the challenge related to the *Data Sharing* stage is similar to the challenge in the *Data Storage* stage. The case company is missing an internal agreement, practice, or policy for how the data is shared among the stakeholders. The current state, with a lack of this kind of practice, has led to that the populated and shared data isn't used and this has a direct impact on the decision-making. There is a need to define an internal policy for the *Data Sharing* stage where the common framework of data sharing is described.

The challenges listed in Table 4 must be taken in consideration when building the proposal for the *Data Management Plan*. The search for existing knowledge and good practices around the topic is described in the following section, *4. Existing knowledge and best practice for Data Management*.

4 Existing Knowledge and Good Practice for Data Management

This section of the thesis discusses existing knowledge in the areas of data, data management, data analytics, and organization data maturity. Existing knowledge about data management, data analytics, and data maturity was essential for the thesis as this knowledge is the basis for *Data Management Plan*.

4.1 Data and Data Management

It is essential to start with understanding what data is and what kind of data is existing.

Data is digitalized information which is presented to any user through a data source. A data source can be any form of source providing the data as files, databases, web-content or streaming data. (Talend, 2020)

According to Delen (2015: Chapter 1), data can be seen as a valuable business asset if the business utilize and treat it in an efficient way. Data is used by businesses to make faster and accurate business decisions. (Delen, 2015: Chapter 1)

In a business there is structured and unstructured data. The difference between structured and unstructured data is that the structured data is well defined, and it is typically repetitive. Repetitive data means that the data structure reoccur repeatedly, and a good example is repeated sales of a product where each sale generate identical data records in a structured way. Structured data is easy to handle in databases as the records are clearly structured and defined. (Inmon et. al, 2019: Chapter 1.1)

Unstructured data can be divided in two types. These types are repetitive and non-repetitive unstructured data. Repetitive unstructured data is where the action is repeated but the data content differ each time. An example for this is where a machine is reading bypassing trains and railroad cars. Non-repetitive unstructured data is data where the action and data content differ each time. E-mails are an example of non-repetitive unstructured data as each e-mail message is different and is received sporadically. Unstructured data isn't well suited with databases due to their irregular form. (Inmon et. al, 2019: Chapter 1.1)

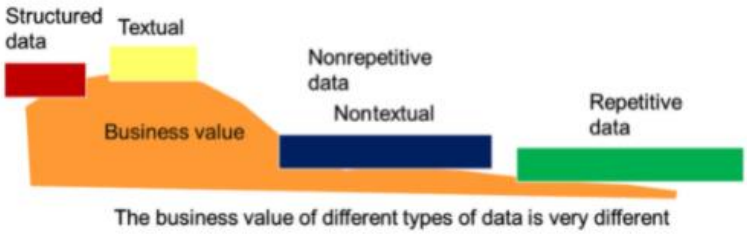


Figure 4. Business value varies dramatically across different types of data (Inmon et. al, 2019).

As seen from Figure 4, *Business value varies dramatically across different types of data* chart by Inmon et. al (2019), different data has different business value. This means that it is important that data is treated in the right way, within the business, so that the business get as much value out of the data as possible.

Data management is usually needed in medium size or large companies where finding data is time consuming. In small companies, the data and applications are close to the users and help for accessing the data is close as well. The data management need arise when the organization grow and the accessibility to data is delayed because of scattered knowledge, delay in the decision-making process, or lack of data ownership.

There are different definitions of data management but one of the simplest definitions is: “data management is a function which ensure that data, information in other words, is managed in a way so that it is suitable for communication, reading, or processing. The definition *data management* is often associated with enterprise-wide data services, but it hasn’t to be so in reality” (Gordon, 2013: Chapter 1.3).

Data management has two functions according to Gordon (2013: Chapter 1.3):

1. It supports *strategically* the usage of business data.

2. It supports *operationally* the development and maintenance of information systems.

According to Gordon (2013: Chapter 1.3), data management has also a set of responsibilities which are:

1. Ensure that data is recognized as a business-wide resource.

2. Improving and maintaining the data quality processes.

3. Ensure information sharing and data definitions provisioning among all stakeholders. This also includes enforcement of the usage of these data definitions.

4. Ensure that the right people, *owners*, are accountable for the data definitions development.

5. Ensure that there is a single data source within the business. This could also be called a *Master data* source.

Data Management has six activities and each one of these has a set of deliverables as shown in Figure 6 (Gordon, 2013).

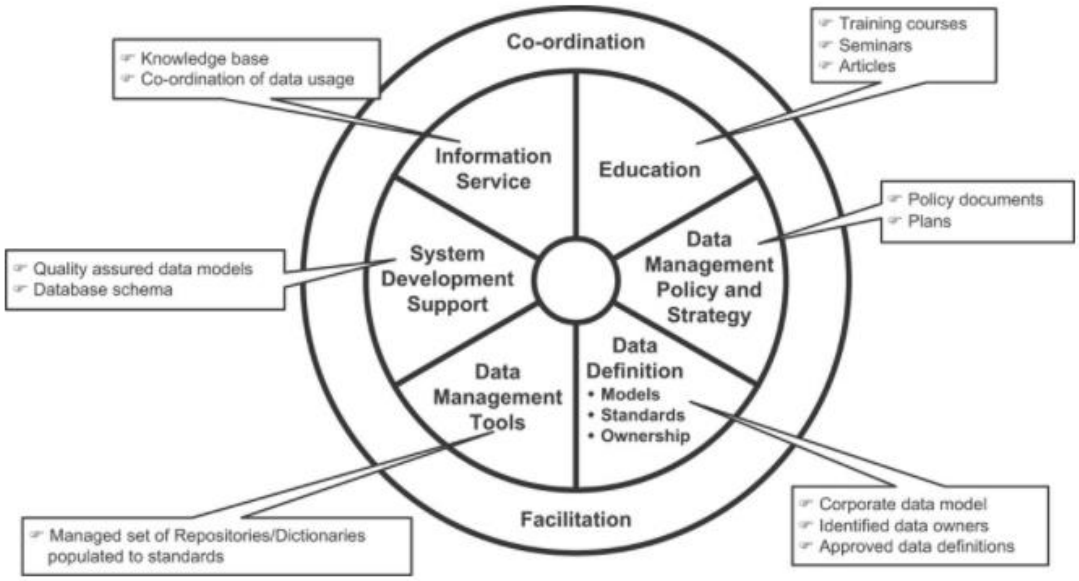


Figure 5. Data management activities and deliverables (Gordon, 2013).

The six activities, including deliverables are as following according to Gordon (2013):

1. *Education* of all concerned people as the message about the importance of data management must be spread within the organisation. Each individual need to understand the role they play in data management. Education can be done through *training courses*, *seminars,* and *reading & posting articles*.

2. *Data Management Policy and Strategy* set the framework for what the business expect of data management and from the staff. The data management policy should be facilitated with a top-down approach to get the needed result out of the policy content. The data management policy contains *policy documents* and *plans*.

3. The *data definition* activity is about defining what data is used within the organisation and appointing the suitable owner of each data definition. This activity is one of the main activities within data management. This activity is defining how data models are developed, what data formats to be used, and naming conventions to be implemented. Deliverables in this activity are a *corporate data model*, *identified data owners*, and *approved data definitions*.

4. *Data management tools* is about having the right tools for the right purpose. Data management is a complex function and require automation. Therefore, the needed tools must be identified and purchased. Processes around the consistent usage of the tools must be created as these processes are enforcing the usage of the tools. The deliverable for this activity is that a *managed set of repositories/ dictionaries are populated to standards* within the business.

5. The *system development support* activity ensure that future information system development is influenced by the existing data management. Data owners must interact with the users and system developers during the development project to ensure a successful transition project. The deliverables are *quality assured data models*, and a *database schema*.

6. The *information service* is the final activity. The information service is about sharing information and data within the business. Efficient information sharing ensure that data is used within the business and by the staff needing data. This activity enables data driven decision-making as it makes data available. The data managers are playing a key-role in this activity as they have most knowledge about available information and data. The deliverables in this activity are a *knowledge base*, and *co-ordination of data usage*.

According to Strengholt (2020; Chapter 1), the new trend is that companies are integrating their data with third parties and external platforms by using Application

Programming Interfaces (API). This means that the data is more decentralized and makes data management more difficult than before. This decentralization has negative consequences in the form of network issues and means that data integrations must be carefully planned to avoid unnecessary risks.

Gordon's circle of data management activities and deliverables shown in Figure 5 is a good example on how data management can easily be visualized. The visualized activities and deliverables will be the foundation of data management outcome of this thesis and are directly addressing the weaknesses identified during the *Current State Analysis*. Some of the mentioned deliverables must be adjusted and customized to fit the case company environment.

4.1.1 Data definitions

Data definitions can be considered an inventory of current data and data sources. The development of data definitions requires an internal framework of rules or guidelines which tells how data objects are created and maintained.

According to Gordon (2013: Chapter 5), the data definition guidelines are critical as they ensure that each data object is defined and documented in the same way, and they also define the minimum acceptable level of information per data object. The guidelines should include the following definitions per data object:

- A unique name or label (unique identifier).
- Any synonyms or aliases for the data object.
- A description of why the data is relevant.
- Used data formats.
- A valid value lists or validation criteria describing what kind of values are expected in the data object.
- Valid operations describing for which operations the data is used.

- Data ownership information.
- A description of for who the data is for.
- A data source description.
- Comments
- Configuration information if such exist.

A data definition can be simple as seen in Figure 7 by Gordon (2013).

Name or Label	Salary
Synonyms	Rumuneration
Significance Statement	has significance as the annual reward, expressed in whole pounds Sterling, paid to an employee of the company before the addition of any extra payments for overtime and/or performance bonuses and before the application of any statutory deductions
Format	Currency
Value List	Not applicable
Validation Criteria	> 0
Valid Operations	multiply by number; answer is currency divide by number; answer is currency add currency; answer is currency subtract currency; answer is currency divide by currency; answer is ratio
Ownership	HR Director
Users	HR Management System, Payroll System
Source	Interview Deputy HR Manager, 15 Sep 06
Comments	[None]
Date created	15 Sep 06
Author	K F Gordon
Date last updated	26 Feb 13

Figure 6. A data definition with validation criteria and valid operations (Gordon, 2013, Chapter 5).

A data definition template, like the example presented in *Figure 6*, is needed for successful completion of data source inventory and should be included in data management plan as an appendix.

4.1.2 Data Management Benefits for the Business

Working data management has a lot of benefits for a business but direct costs savings shouldn't be expected. There are two specific areas where business benefits can be expected. The first area is business related and the second area is IT and systems related. Data availability and data quality across the business is a great benefit as it support the decision-making process and ensure improvement in the overall business efficiency. Available data ensure that the business gains can be measured through financial figures, which make the benefit tangible and visible. IT and systems related benefits are collected through costs savings in application maintenance and increased development productivity. (Gordon, 2013: Chapter 1.3)

Data management can support the business to comply with regulatory, privacy, data protection, safety, and security requirements. Businesses must be able to provide authorities verifiable data about regulatory compliance. The business can be fined in case it fails to comply with these requirements. The business must also comply with privacy and data protection requirements set by authorities. The business and customer data must be protected from unauthorized access and misuse as this can lead to criminal activity as identity thefts, sabotage, breach of immaterial rights. (Berson et. al, 2011: Chapter 1)

Data definitions is essential for the data management plan as this is the activity which ensure that the data and its source is documented in a structured way. The initial data cards created for the case company will be updated and follow the structure presented in this section.

4.1.3 Data Management Best Practices

There are several best practices in *Data Management*, according to Talend (2021).

First, there is a need to create a *Data Management Plan* (DMP). This living document frames data management within a business and defines data usage, accessibility,

preferred file formats, naming conventions, archiving & backup, documentation location, suppliers, and data ownership. The document must be periodically revised to ensure that the DMP contain all changes which occur in the data environment. The DMP is also proof of data control for internal stakeholders like auditors.

Second, *the data storage* must be defined carefully. The data storage approach must be defined according to the business needs and can be a data warehouse or a data lake which is on-premises or in the cloud. The data storage structure must be defined as well. The data storage structure contains parameters which ensures that the stored data is consistent. These parameters are the naming of files, folders, directories, and users. Failing in enforcing these storage parameters will eventually lead to failure in data management.

As part of the above, security must also be enforced to secure the availability and integrity of data. A backup plan must be put in place to ensure that the data can be restored if lost. Also, detailed documentation of the data storage plan must be created to ensure continuity in case key personnel leaves the organisation.

Third, *the data sharing* with the businesspeople needs to be supported and defined. The right location and tools for accessing the data must be identified and deployed to the audience of the data. (Talend 2021.)

The best practice for data management, described by Talend (2021), is that the organization must develop and maintain processes which supports the data management, identifies inconsistencies & incomplete data, mitigate data quality risks, and defines how the company is collecting, storing, and processing data.

The best practice described in this section fits the need of the case company and the approach will be used in the *Data Management Plan*. The principles of the *Data Management Plan* are described in the next section of the thesis.

4.2 Data Management Plan

The *Data Management Plan (DMP)* needed by the case Company is an action plan that will support the company to reach a decent level of data analytics maturity. The DMP will

be built on the best practices described in *Chapter 4* of this thesis, but the DMP shall be customized according to the case Company needs.

Athuraliya (2021) defines an action plan as following: "*An action plan is a checklist for the steps or tasks you need to complete in order to achieve the goals you have set.*"

According to Athuraliya (2021), an action plan includes the following components:

- A clear and well-defined description of the goals of the action plan
- The necessary tasks and steps that have to be carried out to reach the goal of the action plan
- Definition of who oversees each task
- Time schedule, including deadlines and milestones
- Resources needed for each task
- KPIs for measuring the progress

The DMP shall be created in a text document which will act as the master document for the DMP. The DMP should be structured as any other formal document within the case Company, including formal *Table of Content*, but shall contain at least the following sections:

1. Introduction
2. Context and Objective of the DMP
3. Implementation of the DMP

   3.1 Education

   3.2 Data Management Policy and Strategy

3.3 Data Definitions

3.4 Data Management Tools

3.5 System Development Support

3.6 Information Service

4. Data Maturity Assessment

5. The Action Plan

6. Appendixes

The DMP is created with the *Five Pillars of the PC Factory Data Management* Plan in mind, which are described in detail in the coming sections of this chapter. The five pillars are:

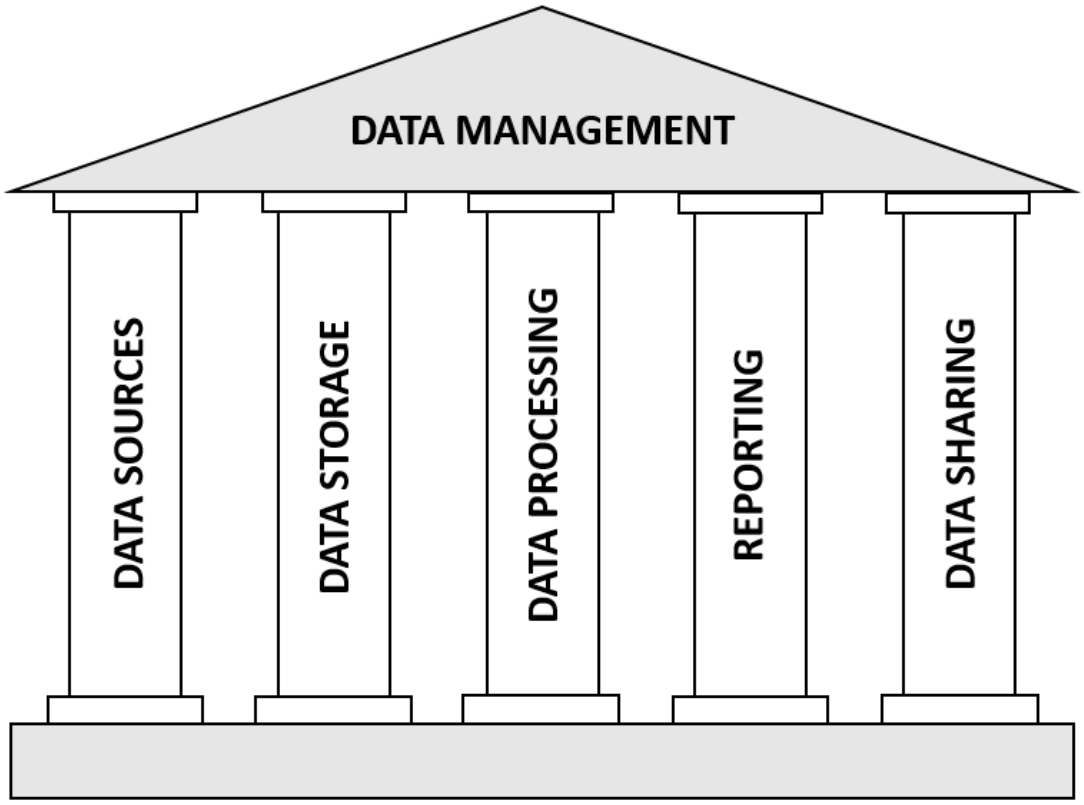


Figure 7. The Five Pillars of the PC Factory Data Management.

The Five Pillars of the PC Factory Data Management in the thesis should be referred to in the DMP so that the reader of the DMP get a clear picture of the target data management setup. The strengths and weaknesses identified during the *Current State Analysis* must be included in the DMP. The weaknesses are converted as action points in the action plan to ensure that the weaknesses are mitigated.

The DMP shall also contain the necessary templates to be used when following the DMP. The following templates shall exist as appendixes to the DMP:

1. Data Source Card -template

2. Report Card -template

The DMP shall take in consideration the findings from the current state analysis, including possible case Company specific information, constraints or policies that might influence any part of the action plan. The initial action plan, including the necessary action points and descriptions, should be prepared and be added in the DMP.

With this structure the DMP will fulfill its purpose for the case Company. The existing knowledge of the core of the DMP, the Five Pillars, are described next in the thesis.

4.2.1 Data Sources

A data source is where the data is initially created, or information digitalized. A data source can be any form of source providing the data as files, databases, web data, or streaming data. (Talend, 2020)

Microsoft (2021) states that the data from the data source can be located anywhere, on the same computer as where the data is consolidated and used, on another computer, or on some other location on the internal or external network.

According to Johnson (2020), any document which is organized for extraction of data can be used as a data source. Text or database files can be used in the same way meaning that there isn't any need of separating these two as different kind f data sources. Text files are usually data extracts from applications while direct database connection

over Open Database Connectivity (ODBC) can also be used to retrieve data directly from a database.

Many applications can utilize external data sources as sources of data. Microsoft products like PowerBI, Excel, Word, and Access are probably the most common applications with this feature. It is worth mentioning, and it is backed up by Johnson (2021), that a data source and its data content must be relevant for the purpose of use.

Data sources is key for data management as they provide the necessary data from various systems about the activity that is wanted to be analysed or reported. The next step is to understand where to place the data and this will be explained in the next section.

4.2.2 Data Storage

Data storage in the case Company context is about data location and accessibility. Data location is the placement of the data source files from where the data is imported to the necessary data analytics or reporting tool for processing. The correct placement of the data, files from the data sources in this context, within the IT environment is crucial because of accessibility for data administrators and analytics tools.

According to Gordon (2013), data accessibility can be divided in three parts:

1. Data Security
2. Data Integrity
3. Recovery of Data

These three parts will be discussed more in-depth in the following sections.

4.2.2.1 Data Security

Data security's fundamental task is to protect data against any unauthorized access from users or systems. This is about maintaining privacy, which means ensuring that data is

accessed and seen only by those who are entitled to do so, but also ensuring that the data is not intentionally corrupted. (Gordon, 2013: Chapter 8)

According to Gordon (2013: Chapter 8), access controls are needed to limit unauthorized access to the data. Access rights should be granted to a person or a group of persons who need the data for their work and the access should be granted through their login to the company systems. The access rights should be limited to the minimum and grant create, read, update, or delete rights based on the person's need. An audit trail should be implemented for the data sources, if only possible, as it records any changes done in the data source.

Data security is mandatory for ensuring accurate and reliable data, but the data security levels must be planned for. Data security can also be used for letting the users know where the data is by onboarding when access rights are granted.

4.2.2.2 Data Integrity

Data integrity is the opposite to data security. Data integrity ensures that data isn't manipulated by authorized users. Authorized users can modify the data on purpose or by mistake and the risk should be mitigated either by controls or training. The optimal solution is a combination of both, where the user knows what he is doing while there are controls in place to ensure the integrity. This kind of file integrity controls are difficult to establish and require 3rd party software. This is out of scope of the thesis and the data integrity focus should be put on training the users who are using or accessing the data.

4.2.2.3 Recovery of Data

Data recovery is about restoring data once it is corrupted or has disappeared. The company must set policies in place which assess the risks associated to the data loss and then ensure that necessary data restore means are available to guarantee a rapid recovery of the data. (Gordon, 2013: Chapter 8)

There are three types of failures and can be classified as following, as described by Gordon (2013: Chapter 8):

- Transaction failures
- System crashes
- Media failures

A transaction is when some operations are done with the data. This could be copy of the data source, as an example. System crashes are due to faults in the hardware where the data is stored or due to bugs in the data management software or the operating system software. System crashes can even be caused by external factors like a power outage which shuts the system down. A typical media failure occurs when a disk loose its data because of some malfunction during the read or write operations from or to the disk.

There are different ways how you can manage your data recovery but using a backup is the most used technique.

> "The main technique used to recover after system or media failures is to restore a backup of the database" (Gordon 2013; Chapter 8)

The need for a good data recovery process is obvious. The types of reasons of failures are also easy understand and the data backup technique and recovery time should be verified when creating the *Data Management Plan*.

4.2.3 Data Processing

Data processing, in the context of the case Company, is about extracting data from repositories and converting it to knowledge. According to Delen (2015: Chapter 3), this is also called data mining. There are several standardized processes for data mining and here I'll focus on one of the best-known processes, called *Cross-Industry Standard Process for Data Mining* (CRISP-DM).

CRISP-DM is the most frequently used standardized data mining process. CRISP-DM was founded in the 1990s in Europe, and the main objective for this development of this process was for it to act as a standard methodology for data mining projects. (Delen, 2015: Chapter 3)

The CRISP-DM process has six steps, which are presented in *Figure 8*.

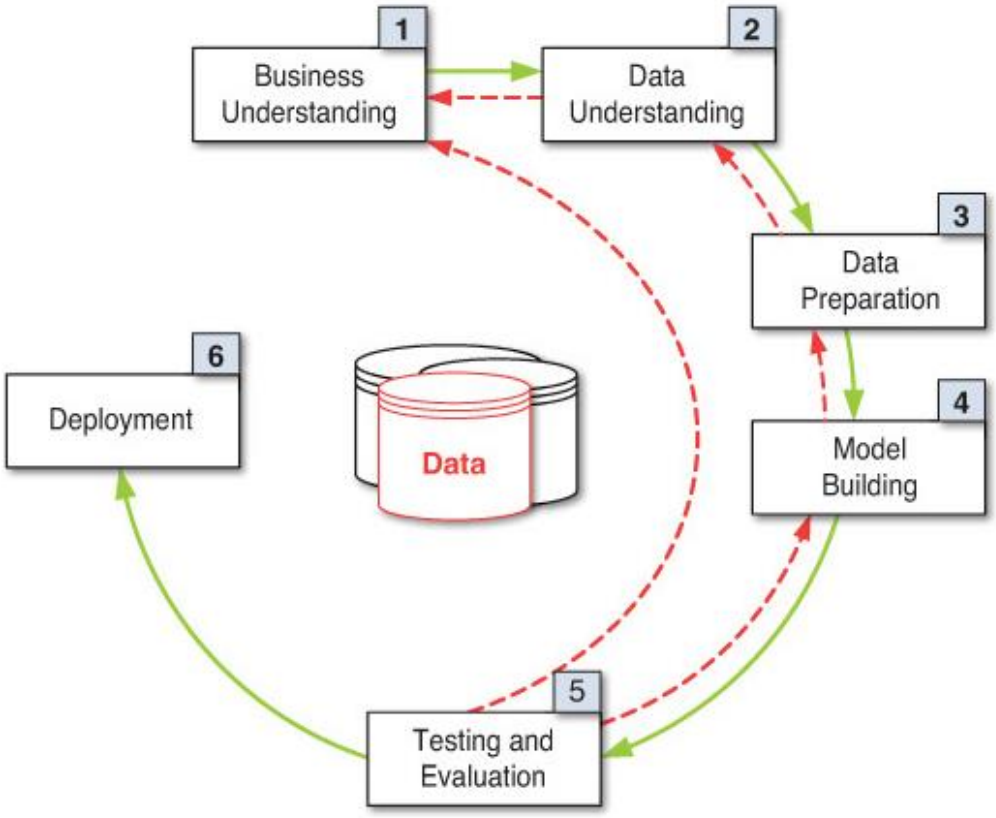


Figure 8. Cross-Industry Standard Process for Data Mining (CRISP-DM). (Delen, 2015)

Delen (2015: Chapter 3) describes the six steps of CRISP-DM as following:

Step 1: *Business Understanding* – Knowing the background of the task is essential for succeeding with the data mining project. The managerial need must be sorted out as well as the specification of the business objective. Specific goals should be defined, and a project plan should be created for reaching the goals. The project plan should ensure the collection of needed knowledge such as responsible person for data collection, data analysis, and reporting. Necessary budgets should be allocated to support the project in case the project of such dimension that the extra funding is needed.

Step 2: *Data Understanding* – Understanding the data is to match the data with the business problem the data is used for. Different outcomes require different data, and it is therefore essential that correct data is used for the purpose. Relevant data sources must be identified and utilized. The data mining task should be clearly defined as this is the foundation for identifying the correct data sources to be used. The used data sources should also be well known by the data analyst or processor. Data storage location, format, data automation state, update frequency, and person in charge of the data source is information that is required to understand the data.

Step 3: *Data Preparation* – Preparing data, or pre-processing in other words, is the most time-consuming step of the process. Data preparation is needed as data is usually incomplete (lacking attributes and/or values), noisy (containing errors), or inconsistent (discrepancies in codes or names). Rectifying the problems with data require a lot of effort and skills.

Step 4: *Model Building* – There are many available modelling techniques, and the technique should be chosen based on the specific business need. Using different techniques for a business problem can complement each other and verify the outcome of the model. It is important to understand that there isn't one specific technique that can be identified as the best alternative for a specific business problem. Therefore, it is important that the data analyst can use different techniques and find the right one which is appropriate for the business problem. Modelling is normally used in more advanced data analytics, like with predictive data analytics. Model building in the case company context can be compared with creation of reports.

Step 5: *Testing and Evaluation* – This step is ensuring that the developed model is assessed and evaluated. The model, or report in the case company context, is ensuring that the outcome meets the business objectives set in the first step. This is an important step in the process as it guarantees that the outcome is reliable and can be used by the stakeholders.

Step 6: *Deployment* – The deployment is about organizing and presenting the data is a way that the stakeholders can benefit from it. The deployment can be

in the format of a report. Often the deployment is done by someone else than the data analyst, e.g., the user of the data.

This data processing model is simple and can be modified to the case company need. The data management plan should contain inputs from this section.

4.2.4 Reporting

Data is all around, and data is converted to information. Information has been among the human population since the beginning of time. About 5.000 years ago someone decided to start to record important information, create the first report, in an easy way by writing on cave walls, according to Danciu and McKibben (2017: Chapter 1).

“Reporting can also be understood as the process of presenting the results of a series of research and analysis” (Finereport, 2021).

Logianalytics (2021) defines the meaning of reporting as collecting and presenting data for analysis purposes.

Whatever the report is, the report has always a goal or an objective. The report’s goal or objective is to deliver accurate and objective information is a structured way to meet the consumer’s expectation. Dashboard reports are the most common report format, but reports can also use tables, charts, or other widgets to present and display data. (Finereport, 2021)

There are five points of best practices when it comes to reporting, according to Logianalytics (2021):

1. Preparation of the reports with coming actions in mind, preparation for analysis in other words. The report and data should be formatted in a way that support quick and intuitive analysis. The column names should consistent and understandable by the users of the report.

2. Make the report short enough not to confuse the report users. Work together with the users to understand the need and find a format that is easy to interpret by the users.

3. Ensure data security and be careful with the handling of sensitive data. Grant only access that is needed for specific users or groups.

4. Create powerful web reports that are interactive and easy to navigate. Optimize the reports to limit the impact on the system.

5. The reporting layer should be created on top of the most frequently used data sources. A data source neutral reporting solution allows you to combine data from traditional and non-traditional data sources, such as databases, files, and web sites.

Preparing a report require definition of pre-requisite information about the report that should be created. According to Sisense (2021), a *Business Requirements Gathering* questionnaire could be used for collecting the necessary details prior to the creation of a dashboard, but the same can be applied for a report as well. The questions list is as following:

Define the dashboard's business need

● Who is going to use the dashboards?

● What are the dashboard business' needs? (Summarize/ analyze/ monitor)

● When will the dashboard be used? How timely must the data be?

Describe success and set KPIs

● Describe success: what is your desired business result?

● What KPIs represents this business result?

Identify cause and effect chain

● What behaviours or activities will drive desired business success?

● What are the expected intrinsic outcomes?

Build dashboards hierarchy and KPI architecture

● What measures represent intrinsic outcomes?

● How would you describe it? Build the dashboard hierarchy.

Understanding what a report is and what it requires to build one is essential for the outcome of this thesis. Information gathered in this section will assist the building of the *Report Card*, which will be added as an appendix in the data management plan. The *Report Card* will act as a template for report pre-requisites collection and documentation, which support the data management initiative.

4.2.5 Data Sharing

Data sharing, in the case company context, is about sharing the data with stakeholders for further use, or manipulation. Shared data can be used for decision making, e.g., IT budgeting purposes as the case company data contain customer company PC fleet data.

According to SCDS (S021), data sharing is referring to the framework and practices of data sharing instead of data itself. SCDS (2021) points also out that data sharing has existed for a long time, even before the era of computers and networks. The rapid development of technology and the digital environment has accelerated data sharing in a digital way during the past decade.

SCDS (2021) lists three elements that have impacted the way data is shared. First, the availability of data has increased significatly along with data quality. It is nowadays easy and cheap to store, process, and share data. Second, the culture around data has evolved. People understand better the value and possibilities of data compared to the past. Data is seen as an asset which contain a lot of value. Third, digital transformation is impacting everyones lives and bring implications of digital to everyone. Instances regulating this domain understand the risks and possibilities of data sharing, even if awareness must be raised on all levels of the society.

SCDS (2021) has also identifed opportunities related to data sharing. First, data can be easily shared between different stakeholders like companies and authorities. Data can be shared in a fast and secure way that respect laws and regulations. Second, data can

easily be combined from different sources, which increase the value for the user of the data. Third, data sharing support data-driven decision-making and lifts collaboration to a completely new level.

The value of data sharing is obvious and access to data for relevant stakeholders must be ensured. Data sharing can increase the company performance on many levels and even improve the individual productivity. Data sharing gives also the necessary transparency for the stakeholders that are interested in the activity, like auditors. These stakeholders can easily get a hold of the data for further analysis.

The Five Pillars of the PC Factory Data Management discussed in this section of the thesis is the foundation of the data management in the case company. Each pillar is described individually in this chapter as it is important to understand each pillar's relation to data management. Knowledge collected in this section is used to understand the relevance of each pillar, as well as to collect practical information for the initial proposal process, which is desribed in Chapter 5.

4.3 Data Analytics

Data is used by businesses to make faster and accurate business decisions. Business analytics is used to improve business performance by describing current performance and predicting future business scenarios. Business analytics is also used for solving complex business problems in different situations. (Delen, 2015: Chapter 1)

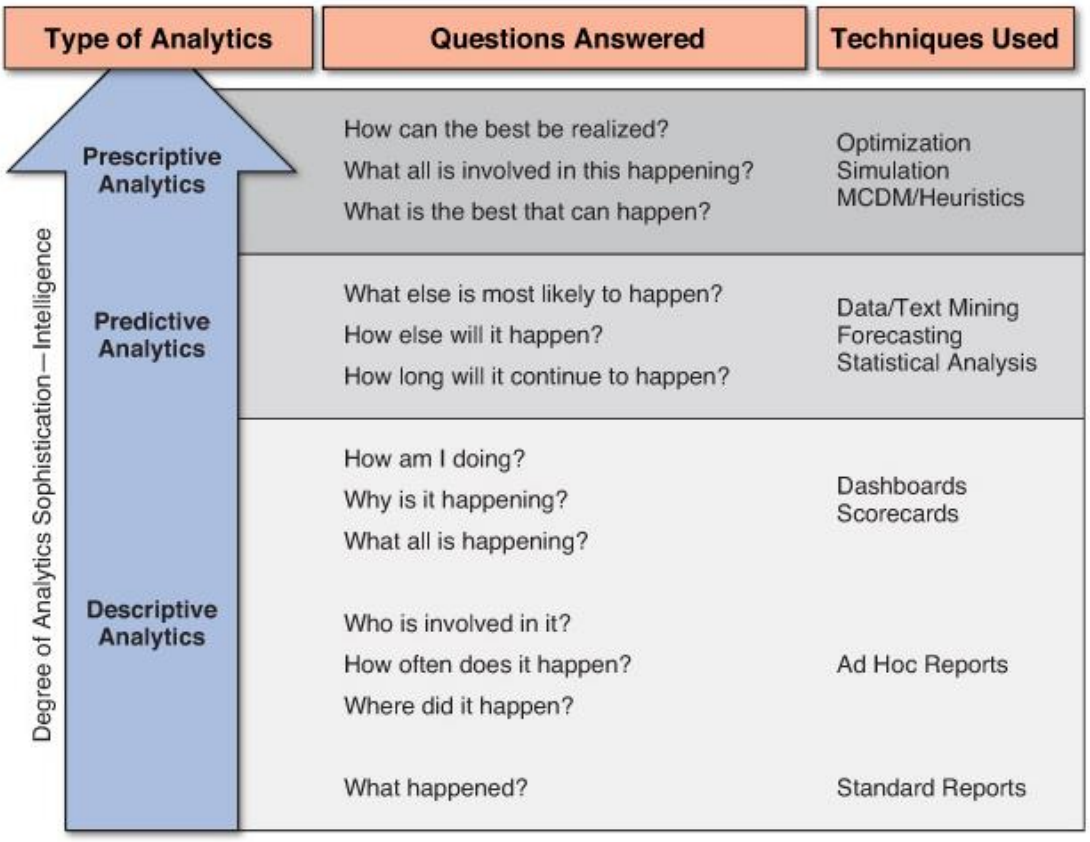


Figure 9. Three levels of analytics and their enabling techniques. (Delen, 2015)

There are three different levels of business analytics maturity, according to Delen (2015) as seen in *Figure 9*. The levels are read from the bottom upward, where the lowest level of analytics has least sophistication or intelligence.

Descriptive analytics is the first, lowest, level of analytics. Descriptive analytics is answering the questions about *what* has happened and *why* it happened. The techniques used at this level are basic reporting methods like usage of *dashboards*, *scorecards*, and *standard reports*. Also ad hoc reporting activities are a part of *descriptive analytics*.

Predictive analytics is the second level in the analytics maturity model. On this level the businesses are predicting *what* is going to happen by using more sophisticated means like *statistical analysis*, *forecasting*, *time-series,* and *data mining*. *Predictive analytics* can help the business to predict customer demand for individual products and required stock size for these products. Even weather forecasting, season sales peaks, and interest rate changes can be predicted with predictive analytics among other possibilities.

Prescriptive analytics is the third and most sophisticated level of business analytics. On this level the business is figuring out *what they should* do next to make the best out of the situation. Prescriptive analytics is using the most sophisticated techniques including *optimization*, *simulation*, and *neural networks*.

Business intelligence (BI) is often used as a synonym for *descriptive analytics* where *advanced analytics* is a synonym for *predictive* and *prescriptive analytics*.

The existing knowledge from this section describes what *descriptive analytics* is and the information will be utilized when building the data management plan.

4.3.1 Data Analytics Tools

There are a lot of different tools for data analytics on the market. “Business Intelligence (BI) software are tools which are designed to retrieve, analyse, transform and report data” (PAT, 2021).

The current leader on the market is Microsoft with its product *PowerBI*. PowerBI was in 2021 positioned for the third consecutive time as the leader in the *Gartner Magic Quadrant for Analytics and Business Intelligence Platforms*. (Ulagaratchagan, 2021)

The *2021 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms* is presented in *Figure 10*.

Figure 10. 2021 Gartner Magic Quadrant for analytics and business intelligence platforms. (Ulagaratchagan, 2021)

Microsoft *PowerBI* has the most complete vision and ability to execute, as presented in *Figure 10*. These features make the tool the most optimal tool on the market today. *Tableau* and *Qlick* is challenging *PowerBI* but isn't considered as mature as *PowerBI* is today.

The information in this section is clearly shows that Microsoft *PowerBI* as a tool is optimal for data analytics and reporting.

4.4 Data Maturity Models

Data maturity modelling helps the business to understand on which stage the company is when it comes to managing and utilizing data. Company performance is correlating to data maturity which makes it essential to know on which stage the company is at, as this is the foundation of making the plan for progressing to the next stage. It is a strict, quantifiable, and an objective view at the data use within a company, department line-of-business, or whatever scope is modeled. The current ratings are of interests and inform possible issues, bottlenecks, and urgency. What should be of any business' interest is what the next level in each category looks like and making plans to get there. (McKnight, 2020.)

Figure 11 represents an example of how McKnight (2020) sees what a data maturity model could look like.

Maturity Level	Level 1	Level 2	Level 3	Level 4	Level 5
Data Organization	*criteria*	*criteria*	*criteria*	*criteria*	*criteria*
Data Architecture	*criteria*	*criteria*	*criteria*	*criteria*	*criteria*
Data Technology	*criteria*	*criteria*	*criteria*	*criteria*	*criteria*
Data Processes	*criteria*	*criteria*	*criteria*	*criteria*	*criteria*

Figure 11. Example of a data maturity model (McKnight, 2020),

According to De Onis (2016), IT and the business must cooperate to create an operating model that solves specific business problems based on the extracted data. This require that the business invests wisely and utilize the right technology at the right time, deliver self-service analytics in a cost-efficient way. To get to the destination of the data journey, the business must know where it has been and where it currently is in data maturity. Maturity assessment is required as trust within the organization is built by having a clear plan and roadmap for building the data maturity for the business. Stakeholders are rarely

willing to accept predictions and more sophisticated analytics without trust in the assessment of the current situation through metrics and KPIs. (De Onis, 2016)

A maturity assessment should be done to establish a starting point of the development. The assessment should be periodically reassessed to ensure progress in the right direction. (Howson & Duncan, 2015)

There are many different data maturity models created by several different companies. This section reviews the most relevant maturity models which serves the purpose of the thesis.

4.4.1 SAS Analytical Maturity

The SAS analytical maturity model is a tool for IT leaders who want to assess the organization's analytical capabilities and state. The involvement of IT, together with analytics and operations teams, are important but the ownership of the analytics, information, and content must be taken by the departments or teams which are utilizing these. (SAS, 2014)

Table 5. Analytic Maturity Scorecard (SAS, 2014).

Level of Maturity	Culture: Decision Makers Use of Data and Analysis	Internal Process Readiness	Analytical Capabilities	Data Environment: Infrastructure and Software
Level V	**Explorative:** Decision makers search for new ways to use advanced analytics to support business decisions.	**Enhanced:** Continually refining processes around data enhancement and analytic methods to optimize resources.	**Optimized:** Commitment to innovative analytic use for future growth and draws on advanced analytics and advances in new techniques.	**Championed:** Continuous improvement/learning to support the most difficult business challenges.
Level IV	**Empowered:** Decision makers leverage analytics across the organization to support business decisions.	**Optimized:** Widely deployed data processes support specific business insights.	**Assertive:** Management supports analytics to bring business units into alignment; analytics talent centralized; best practices shared.	**Quantitatively Managed:** Projects aligned to strategy; documented best practices; smoothly running advanced features and functions.
Level III	**Analytically Astute:** Decision makers adopt analytics for all decisions, including key investments and resource allocation.	**Implemented:** Common data management processes in place and use of data sets and analytics established for decision making.	**Open:** Good intentions, but slow to change; analytics development is constrained, yet departments have own experts/plans.	**Defined:** Clear project life cycles and processes; strong and flexible data environment; infrastructure and software indexed and retrievable; desire for new features.
Level II	**Analytically Aware:** Decision makers recognize benefits of analytics to support decision making but don't leverage analytics consistently.	**Siloed:** Development of department-level processes for data procurement, management and analytics.	**Aware:** Full benefits of analytics poorly understood, siloed and ad hoc activities, yet reasonable results.	**Managed:** Project-driven, often reactive; no best-in-class sharing; completeness unknown.
Level I	**Analytically Unaware:** Decision makers rely on perceptions, historical decisions and non-validated beliefs.	**Void:** No defined data management or analytic processes to support insight development or business decisions.	**Basic:** Lack of analytical skills or executive interest; poorly organized, reactive; considers historical reporting to be analytics.	**Basic:** Some projects have defined scope and objectives; inconsistency and duplication of software.

The *Analytical maturity model* by SAS (2014) consists of five levels and the assessment is done in four categories, as presented in *Table 5*. The categories represent four different parts of the organization readiness and capabilities to use data analytics. The categories presented by SAS (2014) are:

1. *Culture: Decision Makers Use of Data and Analysis*
2. *Internal Process Readiness*
3. *Analytical Capabilities*
4. *Data Environment: Infrastructure and Software*

Each of the five levels of maturity has specific objectives for each category, as presented in *Table 5*. The five levels of maturity by SAS (2014) are:

Level 1: The business decisions are made without facts. There aren't any defined processes in place for data management or analytics. There isn't any understanding of the data analytics requirements for the business and no consistency when driving projects related to data or data infrastructure.

Level 2: The business decisions are supported by analytics but not in a consistent way. The internal processes related to data management are department specific. There is some understanding of the data analytics requirements for the business and the infrastructure is existing and managed.

Level 3: Analytics is widely used by decision makers. The data management processes are defined and implemented. There is a good understanding of the data analytics requirements for the business, but the initiatives are still taken per department. The data infrastructure is existing, flexible, and available.

Level 4: All decisions are backed up by analytics. Data processes are developed further to support the business. The analytical capabilities are centralized and best practices shared within the company. There is advanced infrastructure and software implemented within the business.

Level 5: The business is exploring new ways to utilize analytics. The processes around data and data management are mature and continuously updated. The business is committed to continuously use and develop the capabilities

of data analytics. The infrastructure is developed to support the business needs.

There is a five-step journey to reach analytical maturity, according to SAS (2014). The steps are:

1. *Know what you're looking to achieve.*
2. *Commit resources and encourage ownership.*
3. *Communicate regularly and in a structured way.*
4. *Invest in analytics for the long term in a disciplined way.*
5. *Develop the capabilities to do predictive analytics.*

It is obvious that the current data analytics maturity level of the case Company, according to the SAS analytical maturity model, is *Level 1*. A second alternative for BI and analytics data maturity is needed to get a better view of what is expected to get to the next level. The alternative data maturity model is presented in the next section.

4.4.2 IT Score Overview for BI and Analytics

The Gartner IT Score maturity model is a management tool for IT leaders. The IT score maturity model is a tool for identification of the current maturity level of business intelligence and analytics within an organization. The tool will also point out which initiatives the organization must reach to achieve maturity levels, and as such the model also help the organization to create a roadmap for improvements needed for achieving the next maturity level. The maturity model can also be used when justifying the value of increasing the maturity of business intelligence and analytics to business managers. A business intelligence development program requires human resources, personal skills, defined processes, metrics, and the right technologies for the purpose. (Howson & Duncan, 2015)

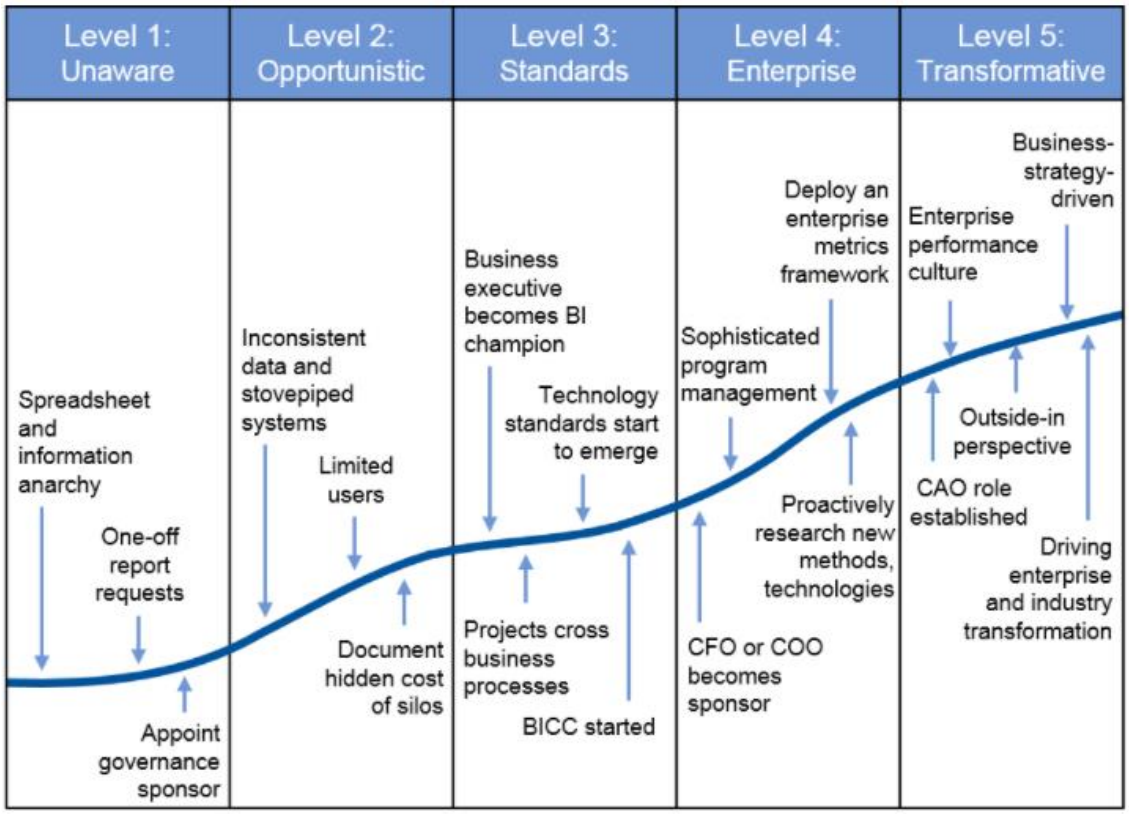


Figure 12. BI and Analytics Maturity Model (Howson & Duncan, 2015).

The *BI and Analytics Maturity Model* by Howson & Duncan (2015) consist of five levels as presented in Figure 12.

Level 1: *Unaware* – The demand for business intelligence and analytics is ad hoc. The decision-making is unstructured and lacking processes and practices. There is no information infrastructure, processes, or performance metrics.

Level 2: *Opportunistic* – Business units manage their own business intelligence and analytics projects. These business units have their own tools and information infrastructure. The business units have dedicated IT organizations. The process modeling is usually missing, and the usage of the existing tools isn't efficient. Analytics outcomes are mainly reporting, and dashboards which contain business or domain specific content.

Level 3: *Standards* – Company-wide coordination has started through the establishment of competence or analytics excellence centers. Technology

standards including architecture, business intelligence platforms and data warehouses are being put in place. No consistent sharing of data or analytics models between business units. Only some sharing of processes, and resources occur.

Level 4: *Enterprise* – The top management of the company is sponsoring business intelligence and analytics projects and initiatives. There is a company-wide framework of performance metrics which are directly linked to the company strategy and objectives. The company decision-making is supported by the established business intelligence applications. Common company rules, models, and practices are applied to mitigate issues with fragmented data.

Level 5: *Transformative* – Business intelligence and analytics has become a strategic initiative and is sponsored by the company chief executive officer (CEO) or chief analytics officer (CAO). All parts of the organization are working closely together with a focus on generating business value. Information and data are seen as a strategic asset, and these are used to generate income streams for the company. Processes and models are now standardized which support the decision-making.

The existing knowledge for *BI and Analytics Maturity Model* by Howson & Duncan (2015) confirm the data maturity level of the case Company as Level 1, as discovered in the previous section. These two data analytics maturity models will assist the case Company in understanding the current level of data analytics but also assist in the thesis when defining the DMP action plan.

4.5 Conceptual Framework of This Thesis

The conceptual framework for this thesis is presented in *Figure 13*. The conceptual framework consists of three main themes discussed above that are put together to build the foundation for a *Data Management Plan*. Each theme is addressing a set of questions which are relevant when building the proposal in Section 5 of the thesis.

Data Management	Data Analytics	Data Maturity Models
(Berson et. al., 2011) (Gordon, 2013) (Strengholt, 2020) (Talend, 2021)	(Delen, 2015) (Inmon et. al., 2019) (Talend, 2020)	(Howson & Duncan, 2021) (McKnight, 2020) (SAS, 2014)

• Answering the questions:
  • *What is data?*
  • *What kind of data exist?*
  • *What is Data Management?*
  • *What are the Data Management Best Practices?*
  • *What should a Data Management Plan (DMP) be and contain in this context?*
	• Answering the questions:
	  • *What is data analytics?*
	  • *Why is data analytics used?*
	  • *What does Descriptive Analytics mean?*
		• Answering the questions:
		  • *What are data maturity models?*
		  • *Data maturity models used today?*
		  • *How can current data maturity be measured?*
		  • *What are the requirements for reaching the next maturity level?*

Figure 13. Conceptual Framework of this thesis.

First, *data management* is defined. Data and its various forms must be described as a part of data management, as data is a fundamental part of data management. Data is a part of the foundation for the thesis as data is the core element of the outcome of the thesis. In addition, the *Data Management Plan (DMP)* as a definition must be defined so that the structure and content of the plan is clearly identified. These elements support the creation of the DMP. Gordon's (2013) wheel of activities and deliverables of data management, which is presented in this section, is relevant for the outcome as it is responding to several weaknesses found during the *Current State Analysis*. These weaknesses are:

1. Lack of policies and information flow regarding Data Storage

2. Lack of knowledge and skills regarding Data Reporting

3. Lack of practices and policies regarding Data Sharing

The Five Pillars of data management is also addressed for describing the best practice of the five pillars in the context of the thesis.

Second, data analytics is defined. Data analytics is an essential part of the managing and presenting data. Data analytics is essential for any business to understand and present the business activity. The weakness of "Lack of reporting/reports", as presented in *Table 4*, is directly linked with this section and therefore the section is relevant for the thesis.

Third, *data maturity models* provide the measurement tools for evaluating the current state of data maturity within a business. Understanding the current state is essential for making the right decisions. Data maturity models provide as well clear demands and objectives for the different maturity levels, which is of great help when creating the data management plan and developing it in the future. A data maturity model will be used in the DMP as measurement of the need for actions, which means that actions in the action plan will be based on requirements for reaching the desired maturity level.

As a summary, for building the Data Management Plan in Section 5 these concepts and tools were necessary to scrutinize: (1) what constitutes data and data management; (2) how to use data analytics; and (3) how to use data maturity models when building the Data Management Plan.

The existing knowledge findings will guide the *Data Management Plan* in the proposal building phase in the next Section of this thesis.

5 Building Proposal for the Data Management Plan

This chapter of the thesis focusses on the *proposal building stage* of the *Data Management Plan*. The *proposal building stage* is based on the findings from the current state analysis and the existing knowledge conceptual framework, which were presented in the previous sections of the thesis. This chapter is divided in two sections which include the overview of the proposal building stage, and findings from Data 2. The initial proposal is attached in the thesis as *Appendix 6*.

5.1 Overview of the Proposal Building Stage

This section presents the steps used during the proposal building stage of this study. The Proposal, the *Data Management Plan*, will address the weaknesses found during the Current State Analysis stage. The main weaknesses were related to the lack of internal policies & practices, skills, and reports.

The best practices, described in the previous section, were related to what is necessary to create a data management plan, what the identified five pillars should contain, as well as what data analytics means and how its maturity is measured. All these parts were relevant for building the proposal.

First, the initial master *Data Management Plan (DMP)* document was created together with the necessary templates listed in the DMP appendices. This was necessary as the master document was completed during the following steps of the proposal building stage. The templates were also tested and used by the organization as basis for providing feedback and input for the proposal during the building stage.

Second, the *Data 2* data collection was conducted. *Data 2* contained two internal stakeholder interviews which were: the SGTS Director and the PC Factory Team Leader. The interviews were important for the proposal building stage as each interviewee had to present their expectations and requirements for the content of the *Data Management Plan*.

Third, the initial master document was updated with the information and data collected during data collection *Data 2*. This step ensured that the feedback collected during the *Data 2* data collection stage was taken in consideration and implemented in the proposal.

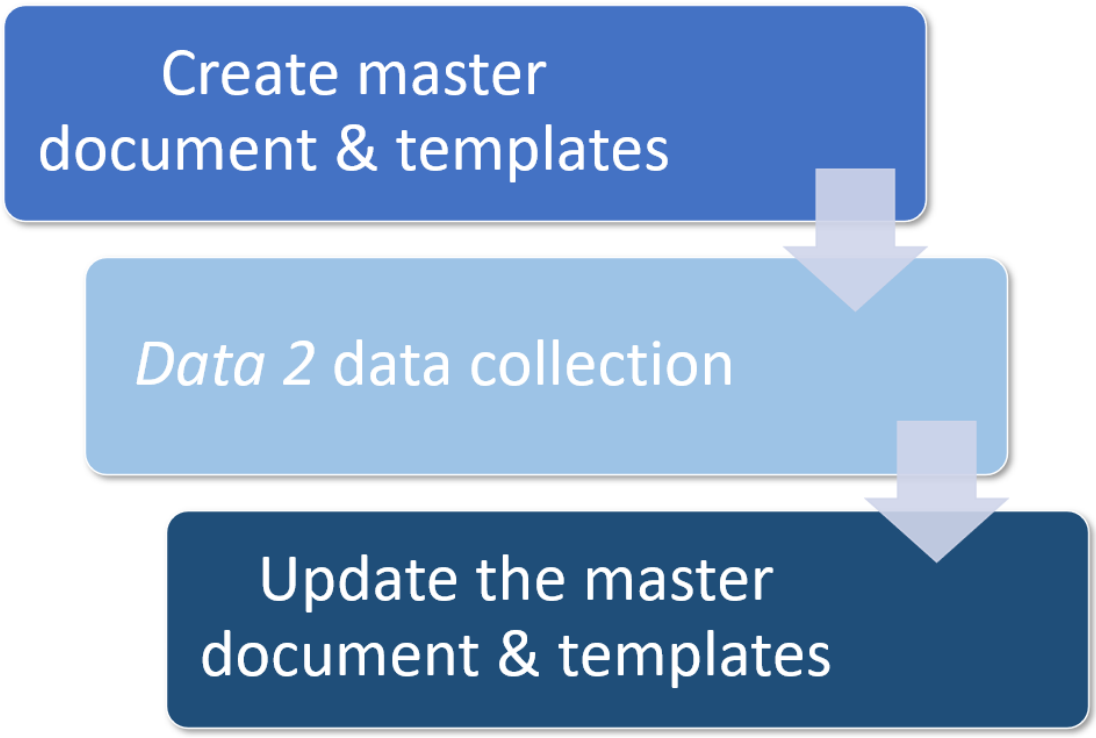


Figure 14. Building the Proposal rationale.

The rationale of the building the proposal is presented in *Figure 14*. The three steps present a logical approach to building the proposal and ensured that the stakeholders were heard, and their requirements & proposals were included in the proposal.

Building the proposal, in the context, was a process where the stakeholders, also validators of the outcome, were a part of the creation process. The stakeholders' input and feedback were asked when the initial proposal had been created. The input and feedback were discussed with each one stakeholder providing feedback individually and documented in field notes. All input and feedback were taken in consideration to make the data management plan as suitable as possible for the case company. The involvement of the stakeholders in the building phase meant that the final validation would be a short process as the validation would be a part of the build phase.

5.2 Findings from Data 2

The initial proposal, document version 0.1, was sent to the stakeholders listed in the *Data 2* data collection. The stakeholders were asked if the document is easy to read, if the stakeholder has any input or feedback, and if there are any other comments the

stakeholders wanted to share about the initial proposal. Each stakeholder provided input and feedback which was tracked, analyzed, and reviewed together with the stakeholder.

The feedback was very positive, and the work was appreciated. The structure of the document and content was clear and understood by the stakeholders. The input and actions the input generated is presented in *Table 6*.

Table 6. Data 2 input and actions.

Input	Action
Easiness of understanding the DMP document text. Abbreviations were asked to be put in a glossary.	Addition of Chapter 8 - Glossary
Clarification on GDPR compliance	Added an action point in the *Data Management Policy* where the GDPR compliance has to be mentioned. The DMP itself do not use nor treat personal data.
Clear statement of *Why DMP?*	The Introduction was updated with a clearer description of the objective of a DMP.
Figure 5 should be updated with the Data warehouse	Figure 5 was updated.
Structure of document could be clearer	Not a part of the thesis as the document template is used by the case company
Typo correction on page 4	Corrected
Unclear statement on page 8, chapter 4.2 and point 1b. Should state *What data is being used within the organization?*	The header of point 1b on page 8 corrected.

The actions, generated by input and feedback, were documented in the field notes of the interviews (Appendix 5), and implemented in the document. The stakeholder input and feedback were written in the 0.2 and 0.3 versions of the document. Document version 0.2 was updated with the feedback from the *SGTS Director* and document version 0.3 was updated with the feedback from the *PC Factory Team Leader*. The document version 0.3 was finally sent to the stakeholders for final approval. Document version 0.3 was converted to the official version 1.0 after the validation and approval of the document as the final version. The documentation of the approval was done in the *Data 2* field notes by adding question 5 "*Do you approve document version 0.3 as the official version of the DMP?"* and completing the answers when approved.

The initial proposal, document version 0.1, is added in the thesis as *Appendix 6* and the final version, document version 1.0, is added in the thesis as *Appendix 8*. The validation of the proposal is discussed in the following section.

6 Validation of the proposed Data Management Plan

The validation of the proposal is presented in this chapter of the thesis. This chapter discusses how the proposal was perceived and what kind of feedback was received from the key stakeholders.

6.1 Validation as a part of the build phase

The validation of the proposal was partially included in the building the proposal section of the thesis. The key stakeholders, the SGTS Director and PC Factory Team Leader, were involved in building the proposal through interviews, where they provided feedback about the initial proposal and evaluated the proposal during the build phase. The content of the initial proposal was updated, along with new version numbers, based on the key stakeholders' feedback and evaluation. The key stakeholders approved document version 0.3 as the outcome of the thesis. Document version 0.3 was converted to the first official version of the document, version 1.0.

6.2 Summary of the Final Data Management Plan

The validation of the data management plan, data collection *Data 3*, was finally a discussion about the implementation of the thesis outcome, the *Data Management Plan*. Possible consequences and impact on the organization were also discussed and speculated, as similar initiatives haven't been implemented before within the case company and experience in data management is limited.

The feedback on the outcome of the thesis was positive and value well understood. The outcome was identified as a very technical and detailed document which implementation require time and effort. It was considered that the case company will raise the professionalism within its activity by following the data management plan, and by doing so get in control of and utilize the data for the benefit of the company.

Table 7. Data 3: Feedback summary

Function	Reaction	Description
SGTS Director PC Factory Team Leader	Positive	"A good initiative as we have much data." "This is a professional approach that is supporting our activity." "This is a process that can be implemented in other parts of our organization as well, if the initial implementation, test that is, is successful." "It is great to get a detailed study and facts with best practice behind the DMP so that it isn't just someone's thought about what is good to do."
	Neutral	"People and willingness to change work methods is the only thing that comes to mind."
	Constructive critisism or feedback	"The document is very clear but very technical." "The challenge with implementation might be related to the change, which will require change management" "Implementation is one thing but the operations and maintaining the process afterwards is the challenge."

As seen in *Table 7*, the key stakeholders seemed satisfied with the outcome and had only a few constructive or feedback comments. The positive feedback was about raising the professionalism within the organization, getting in control of data, and the fact that the study is backed up with facts and existing knowledge. The neutral comment is a comment about the change that the implementation of the outcome will present. The constructive criticism or feedback was related to the technical content of the document, challenges with change and the continuous maintenance of the documentation after the implementation. The outcome was validated according to the plan, through *Data 3*, and is considered successful.

7 Conclusion

This chapter of the thesis summarizes and evaluates the study from the objective definition to the outcome.

7.1 Executive Summary

The Saint-Gobain Technology Services Nordic & Baltic (SGTS NB) PC Factory was established in 2018 with a focus on delivering excellent PC life-cycle services for the Saint-Gobain companies in the Nordic & Baltic region. The PC Factory activity was established from scratch, as the service was previously outsourced to an external supplier. The creation of the activity took time, and the main effort was put in the service delivery, and data management wasn't in focus. The need of data management emerged as the activity grew as issues were identified in the daily operations. Missing insight on the activity along with customer reporting was lacking as well as internal policies, and skills to manage the data populated around the PC Factory services. The objective of this thesis was to create a data management plan, which would address the issues related to data management within the PC Factory activity.

First, the objective of this study was to identify the current situation within the PC Factory when it comes to data management. The aim was to identify strengths and weaknesses in the current data processes and operations. The thesis research used both qualitative and quantitative methods. The qualitative research consisted of in-depth interviews with relevant stakeholders while the quantitative research consisted of a customer survey, which was addressed to the IT-decision makers of the Saint-Gobain companies in the Nordic & Baltic region. Second, the study included an analysis of existing knowledge, including best practices, about data management suitable to be implemented in the PC Factory. Third, the proposal of the data management plan was built together with stakeholders within SGTS NB and the outcome was finalized when the stakeholders approved the document. The overall process, including analysis and summary, took approximately a year to complete.

The information collected during the study was done with the help of three data collection rounds (Data 1 – 3). The first round (Data 1) collected information about the current situation within the PC Factory through key stakeholder interviews, a customer

survey, and an internal team workshop. It was discovered that data management policies and skills were missing which led to missing reports and excessive manual work. All these points contributed to the data management inefficiency experienced within the PC Factory. The second round (Data 2) collected information about the proposal and its development through interviews of key stakeholders. This data collection round as aimed to collect feedback about the first draft of the data management plan to ensure that key stakeholder feedback and input was received, analyzed, and needed changes applied in the final version of the data management plan. The third (Data 3) and final round collected information about the validation of and thoughts about the final version of data management plan through interviews of key stakeholders. The analysis of existing knowledge revealed that existing models for data management and data maturity assessment exist, and suitable models were chosen to be implemented in the outcome of this thesis.

The result of this thesis, the data management plan, is targeting a specific business problem which is causing ineffectiveness and lack of control within the activity. The data management plan is written in a way which makes it possible to replicate, with modifications, to other parts of the organization, which makes the data management plan valuable for the organization.

The data management plan addresses the six activities discovered during the analysis of the existing knowledge. By following the steps in the data management plan, the organization can get in control of the data, start utilizing the data more efficiently, and build the foundation on which the following developments around data management can be built. The six activities, which is the foundation for data management in this thesis, are: education, data management policies and strategy, data definitions, data management tools, system development support, and information service. The final data management plan contains a detailed action plan, which is targeted for the PC Factory, to ensure a structured implementation of the proposal. The implementation requires effort and resources, but the benefits of implementing the data management plan are obvious, as discovered during the data 3 interviews. The organization can assess the current data maturity with the help of the data maturity assessment as well as understand the requirements of reaching the following level, which is essential for further development of data management and moving toward a data-driven activity.

The proposal was built together with stakeholders, who validated the versions of the proposal, and the proposal was finally approved in the end of the build phase. This means that the validation of the final version of the data management plan was done by people who were a part of the creation process. The data management plan is, considered by key stakeholders, to be ready to be implemented within the PC Factory as such. The feedback from the PC Factory team leader is positive and the value is identified. It is also agreed that the data management plan will be implemented according to the action plan defined in the document. Some adjustments in the internal processes are required to ensure the continuous development of data management within the organization.

As a conclusion, a clear data management plan is needed for growing organizations as the data populated in the various data sources are difficult to manage. The data management plan helps the organization to identify, structure, and utilize data efficiently while data maturity assessment is assisting the organization to identify the current state of data maturity in addition to understanding what it requires to get to the next step of data maturity.

7.2 Next Steps and Recommendations

There are some points to take in consideration before continuing with the implementation of the *Data Management Plan*.

First, ensure proper resourcing for the facilitation of the project. Project management and facilitation efforts are needed and should be taken in consideration when planning the implementation. Necessary resources, especially human resources (time), must be allocated to guarantee a successful implementation of the data management plan.

Second, the data management plan doesn’t give the answer to all questions within the organization around data management. This means that additional planning work is needed even if the tasks are listed in the action plan. In addition, the working environment is constantly changing which means that data sources, stakeholders, and report needs should be re-assessed during the project.

Third, the data management plan is created for correcting the current weaknesses. Delaying the implementation might lead to a mismatch between weaknesses identified during the current state analysis and current weaknesses, which means that the weaknesses should be re-evaluated before the initialization of the project.

The implementation of the data management plan is possible, and should be a success, if the points above are taken in consideration before the initiation of the project. Following the action plan will lead the organization toward a data-driven activity, but the project will impact the organization during the implementation phase. Maintaining the momentum after the implementation of the data management plan, continuing the development of data maturity, and driving the change within the organization will require strength and focus from the leaders but all of this is possible to achieve.

7.3 Thesis Evaluation

The objective of the thesis was to create a *Data Management Plan* for the Saint-Gobain Technology Services Nordic & Baltic (SGTS NB) PC Factory. The outcome of the thesis was a concrete data management plan, tailored for the case company. The objective and the outcome of the thesis are in line which means that the outcome of the thesis met the objective of the study.

To ensure the *construct validity* of the thesis, several methods of data collection (e.g., triangulation) were used based on several data sources. The data collection methods used were interviews, a questionnaire, and a workshop. The internal validity was validated by the case company and the validity was positive as the data management plan is considered applicable within the case company. In this study the final data management plan can be utilized in other domains of the organization and therefore the external validity can also be considered as positive. All these measures were aimed to ensure the validity of the study.

Reliability, according to Yin (2009: 40) requires that the research steps in a study can be repeated with the same results. In this study, the conceptual framework is based on the current best practice and data collected from current employees and stakeholders. The study is considered reliable at the time of conducting the study, but it will be outdated in the near future due to the emerging development of IT in general.

The study provides the case company with a concrete plan to implement best practice data management. The data management plan is rather easy to replicate in other organizations, as the plan is formulated according to the best practice structure, even if written for the specific business problem. The data collected and information gathered for this study were from reliable sources and stakeholders, which makes this study credible and reliable.

7.4 Closing Words

The final section of this thesis is dedicated to my own reflection on thesis and process of writing the thesis.

The problem in the business case isn't specific only to the case company. The world is moving toward an environment full of data, which means that the need for a structured way of managing data is needed within all organizations. It is evident that a lot of companies, large and small, haven't reached or even started the journey toward this data-driven world. The case company had identified the need, but the tools and means were missing so something had to be done.

The outcome of the thesis, and the feedback within the case company, is very satisfying. The concrete outcome, in the sense of a concrete and tangible plan that generate value for the business, isn't an everyday production. The process of writing the thesis was long and even painful at times. I had to use a lot of time in identifying the right sources for best practices and translating these to something that could be used within the case company.

I have personally learnt a lot during the master's thesis writing process. I've realized that the structure of the thesis can be used in daily work as the gate model can be implemented in any development situations. In addition, I've become an expert in a topic that is currently changing the world. I feel really privileged to have had the opportunity to write a thesis of this topic.

Finally, my employer will benefit of the outcome of the thesis which will hopefully have a long-lasting effect on how the company manage data. I believe that this thesis is of benefit to all parties involved.

References

Adams, J., Khan, H. and Raeside, R. (2014). *Research Methods for Business and Social Science Students.* 2nd Ed. New Delhi: Sage Publications. Available from: https://ebookcentral.proquest.com/lib/metropolia-ebooks/detail.action?docID=1698991 (Accessed 3 January 2021).

Athuraliya, A. (2021). *The Easy Guide to Developing an Effective Action Plan*. Available from: https://creately.com/blog/diagrams/how-to-write-an-action-plan/ (Accessed 2 October 2021).

Berson, A. and Dubov, L. (2011). *Master Data Management and Data Governance.* 2nd Ed. McGraw-Hill Companies Inc. Available from: https://learning.oreilly.com/library/view/master-data-management/9780071744584/cover.html (Accessed 17 May 2021).

Danciu, T. and McKibben, T. (2017). *Is reporting dead?.* O'Reilly Media Inc. Available from: https://learning.oreilly.com/library/view/is-reporting-dead/9781491990667/ch01.html (Accessed 14 November 2021).

De Onis, T. (2016). *The Four Stages of the Data Maturity Model*. Available from: https://www.cio.com/article/3077871/the-four-stages-of-the-data-maturity-model.html (Accessed 8 May 2021).

Delen, D. (2015). *Real-World Data Mining: Applied Business Analytics and Decision Making.* New Jersey: Pearson Education Inc. Available from: https://learning.oreilly.com/library/view/real-world-data-mining/9780133551150/title.html (Accessed 17 May 2021).

Finereport. (2021). FanRuan Software Co., Ltd web-site. *What is Reporting? Meaning & Examples.* 2nd Ed. Academic Press. Available from: https://www.finereport.com/en/reporting-tools/what-is-reporting.html (Accessed 14 November 2021).

Inmon, W.H., Linstedt, D. and Levins, M. (2019). *Data Architecture: A Primer for the Data Scientist.* 2nd Ed. Academic Press. Available from: https://learning.oreilly.com/library/view/data-architecture-a/9780128169179/Cover.xhtml (Accessed 14 May 2021).

Johnson, R. (2020). *What Is a Data Source?.* Available from: https://www.lifewire.com/what-is-a-data-source-3969346 (Accessed 15 November 2021).

Gordon, K. (2013). *Principles of Data Management: Facilitating information sharing.* 2nd Ed. Swindon: BCS. Available from: https://learning.oreilly.com/library/view/principles-of-data/9781780171845/01_frontcoverimage.xhtml (Accessed 14 May 2021).

Howson, C. and Duncan, A D. (2015). Gartner web-site. *ITScore Overview for BI and Analytics*. Available from: https://www.gartner.com/en/documents/3136418/itscore-overview-for-bi-and-analytics (Accessed 9 May 2021).

Leavy, P. (2014). *The Oxford Handbook of Qualitative Research.* New York: Oxford University Press. Available from: https://ebookcentral.proquest.com/lib/metropolia-ebooks/detail.action?docID=743724 (Accessed 3 January 2021).

Logianalytics. (2021). Logianalytics web-site. *Reporting in BI.* Available from: https://www.logianalytics.com/resources/bi-encyclopedia/reporting-bi/ (Accessed 14 November 2021).

Microsoft. (2021). *Data Sources.* Available from: https://docs.microsoft.com/en-us/sql/odbc/reference/data-sources?view=sql-server-ver15 (Accessed 15 November 2021).

PAT. (2021). Predictive Analytics Today web-site. *11 key capabilities of business intelligence software.* Available from: https://www.predictiveanalyticstoday.com/key-capabilities-of-business-intelligence-software/ (Accessed 3 October 2021).

Strengholt, P. (2020). *Data Management at Scale.* Sebastopol: O'Reilly Media Inc. Available from: https://learning.oreilly.com/library/view/data-management-at/9781492054771/titlepage01.html (Accessed 17 May 2021).

Saint-Gobain. (2020). Saint-Gobain Group web-site. Available from: https://www.saint-gobain.com/en/group/ (Accessed 29 December 2020).

SAS. (2014). *Five Steps to Analytical Maturity.* Available from: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/five-steps-to-analytical-maturity-106929.pdf (Accessed 8 May 2021).

SCDS. (2021). Support Center for Data Sharing web-site. *What is data sharing?.* Available from: https://www.logianalytics.com/resources/bi-encyclopedia/reporting-bi/ (Accessed 17 November 2021).

Simons, H. (2014). *Case Study Research in Practice.* Cornwall: SAGE Publications. Available from: https://eudatasharing.eu/what-data-sharing (Accessed 3 January 2021).

Sisense. (2021). Sisense web-site. *Business Requirements Gathering*. Available from: https://support.sisense.com/kb/en/article/business-requirements-gathering (Accessed 14 November 2021).

Sreejesh, S., Mohapatra, S and Anusree, M R. (2014). *Business Research Methods.* Switzerland: Springer International Publishing. Available from: https://link-springer-com.ezproxy.metropolia.fi/book/10.1007%2F978-3-319-00539-3 (Accessed 3 January 2021).

McKnight, W. (2020). TDWI web-site. *The Importance of Data Maturity Modeling*. Available from: https://tdwi.org/articles/2020/03/06/ba-all-importance-of-data-maturity-modeling.aspx (Accessed 8 May 2021).

Williamson, K. and Johanson, G. (2018). *Research Methods.* 2nd Ed. Cambridge: Chandos Publishing. Available from: https://metropolia.finna.fi/Record/nelli15.4100000001474165 (Accessed 3 January 2021).

Talend. (2020). Talend Inc web-site. *What is a Data Source?.* Available from: https://www.talend.com/resources/data-source/ (Accessed 29 December 2020).

Talend. (2021). Talend Inc web-site. *What is a Data Management?.* Available from: https://www.talend.com/resources/what-is-data-management/ (Accessed 16 May 2021).

Ulagaratchagan, A. (2021). Microsoft PowerBI web-site. *Microsoft named a Leader in the 2021 Gartner Magic Quadrant for Analytics and BI Platforms.* Available from: https://powerbi.microsoft.com/en-us/blog/microsoft-named-a-leader-in-2021-gartner-magic-quadrant-for-analytics-and-bi-platforms/ (Accessed 3 October 2021).

Yin, R. K. (2009). *Case Study Research: Design and Methods*. 4th ed. Thousand Oaks: SAGE Publications.

Interviews

- Paavola, Tuomas 2021. PC Factory Computer Support Specialist. Interview, 22/03/2021. Interviewer: Hans Kanerva

- Knuutinen, Jukka-Pekka 2021. PC Factory Junior Computer Support Specialist. Interview, 22/03/2021. Interviewer: Hans Kanerva

- Hallanmaa, Teppo 2021. SAP FICO Team Leader. Teams interview, 24/03/2021. Interviewer: Hans Kanerva.

- Rohlen, Lars 2021. SGTS Nordic & Baltic Director. Teams interview, 25/03/2021. Interviewer: Hans Kanerva.

- Kallioniemi, Jere 2021. PC Factory Team Leader. Teams interview, 29/03/2021. Interviewer: Hans Kanerva

- Liberg, Jörgen 2021. SGTS Nordic & Baltic Director. Teams interview, 16/04/2021. Interviewer: Hans Kanerva.

Data 1 - Interview Field Notes

Example 1: Interview of Teppo Hallamaa, SAP FICO Team Leader

Date & Place:	24.03.2021 / Teams video conference Started: 15:00 – Ended: 15:40
Recording Link	https://web.microsoftstream.com/video/08bafa10-38ed-4963-84a0-53ce7a1af02c
Interviewee:	Hans Kanerva
Interviewee:	Teppo Hallamaa, FICO Team Leader

1. What is Your role within the organization?	**I am the SAP Competence Center FICO team leader. I have two subordinates, and both are located in Russia. One employee is allocated to the FI module and the other is allocated to the CO module.**

2. About SAP data. In the PC Factory we have a need for extracting SAP data (materials purchasing & inventory, invoicing, delivery) to get a getter overview of our activity. **Is it possible to extract the data from SAP?**	**It is possible but it will require work. Profit & Loss is shown on a single row in the SAP accounting, and this might cause issues in reporting.** **It is possible to collect different pieces of data and collect them on reports. The wanted data has to be identified so that the tailor-made report can be generated.** **The daily stock status can be easily reported through a .csv or through a Webfocus report. The report can be run manually or scheduled to be run automatically. The report can be sent to an e-mail address or be placed on a network drive.**

4. Can we decide what data is exported (e.g. columns)?	SAP Query tool is forbidden within the Saint-Gobain Group due to security reasons and standard reports in SAP can't be modified. Webfocus or SAP connector to PowerBI is needed for automatic reports. If you deliver the required data column needs for us then we'll be able to investigate if we can export all necessary data.

5. How often can the data be exported from SAP?	I believe that the data export can be done daily. I just need to find the tool meant for this. The data can be exported to a network drive or e-mail.

6. Can the export be automated, or do we have to do it manually each time?	The export can be automated e.g. through a .csv export to a default network location.

7. Are there any ready data visualizations (reports) available already now in webfocus?	Yes, some are. Reports do not need specific access rights if they are automated. PCF details can be filtered with parameters and relevant data transmitted.

8. How shall we proceed to get the data export process running?	Creation of ticket in ServiceNow is needed to get the project going. The ticket will be assigned in the end to Bengt and me. Place the PowerBI connector request in the request as well to get the process ongoing.

9. Is there anything else we need to know about the data export?	**PFC data is clear and should be easy to export. Receivables, shipping cost & product invoicing breakdown can be of interest for you. Additional information fields might require identifying on physical invoices. It isn't impossible but require some manual work.**

Example 2: Group interview of Tuomas Paavola and Jukka-Pekka Knuutinen

Date & Place:	22.03.2021 / Vantaa Office
Interviewer:	Hans Kanerva
Interviewee:	Tuomas Paavola & Jukka-Pekka Knuutinen (JP)

1. What are Your roles within the organization?	Tuomas: I work as a Computer Support Specialist in the PC Factory and taking care of the daily operations. JP: I work as a Junior Computer Support Specialist in the PC Factory and I'm also working with taking care of the daily operations.

2. About data and how you use it today. **Do you know which all data sources we have available today and how to access the data?**	Tuomas: Yes, I know which are the data sources and where the data is stored. JP: Not exactly, I know some but not all. I have some difficulties in finding the storage locations.

3. Do you know how we're receiving the data?	Tuomas: No, I don't know. I don't know how the ServiceNow data extract is done. JP: Partially but some sources are unclear.

4. What kind of information or data are you missing in your daily work?	Tuomas: All data is available, but some must be handled and processes manually. AssetID and serial numbers are marked on paper and it makes the process more difficult and less agile. JP: Same as Tuomas. The manually handled data could be managed better.

5. How would you like the data to be presented for you?	Tuomas: There is a need for visual reports for fast review and database access for manual intervention. JP: I have the same need as Tuomas.

6. What are the greatest challenges today when it comes to data and its availability in the PC Factory activity?	Tuomas: There are too many data sources and storage locations which spread around the network. JP: I agree with Tuomas.

7. Other comments?	Tuomas: The AssetID data management needs to be prioritized to minimize the manual work. JP: There is a need to have all data in one place.

Data 1 - Customer questionnaire & answers

Date & Place:	15.04.2021 / MS Forms
On behalf of:	SGTS NB PC Factory
Audience:	Saint-Gobain IT-managers in the NB region

1.	Are You currently satisfied with the reports you can get from the PC Factory? **(Yes/No)**
2.	If you answered "NO" in question 1: What is the reason for Your dissatisfaction?
3.	Do You know where or how You can get the reports related to Your business' PCs?
4.	Are You satisfied with the level of data (e.g. PC fleet information) provided for you about your business?
5.	Are You missing any data that you consider important for You and Your business?
6.	If You answered "Yes" in question 5: What data are You missing?
7.	Do You consider that PowerBI, current data source for SGTS service data, is sufficient for You to retrieve the necessary data and reports?
8.	Do You have any further comments or wishes about the PC Factory service and/or provided service data?

Questionnaire Analysis			Satisfiaction Statistics		
Recipients Addressed	Replies Received	Answering Rate	Satisfied	Dissatisfied	Satisfaction Rate
44	9	20 %	8	1	89 %

Findings	
#	Main Findings
1	Low answering rate - 20%
2	Recipients generally satisfied (89%), only one recipient dissatisfied (11%)
3	Dissatisfaction about response times and not about reports, data, or services.
4	33% of receipients don't know where to find the data or reports.
5	PowerBI is considered to be the preferred reporting tool.

ID	Start time	Completion time	Email	Name	Are You currently s
1	4.15.21 16:11:22	4.15.21 16:12:13	anonymous		Yes
2	4.15.21 16:48:33	4.15.21 16:50:20	anonymous		Yes
3	4.15.21 17:03:13	4.15.21 17:07:58	anonymous		Yes
4	4.16.21 8:10:55	4.16.21 8:11:56	anonymous		Yes
5	4.19.21 8:47:00	4.19.21 8:48:28	anonymous		No
6	4.19.21 13:34:51	4.19.21 13:35:43	anonymous		Yes
7	4.20.21 11:10:44	4.20.21 11:11:31	anonymous		Yes
8	4.20.21 16:51:41	4.20.21 16:56:43	anonymous		Yes
9	4.26.21 10:11:56	4.26.21 10:14:51	anonymous		Yes

If you answered "N	Do You know where	Are You satisfied w	Are You missing an	If You answered "Y	Do You consider th	Do You have any fu
	Yes	Yes	No		Yes	No :-)
	Yes - I ask SGTS USS Tina	Yes	No		Yes	Poor delivery time - not due to you but Lenova / corona. But very finde "heads up" on this issue in SGTS Weekly News
	Yes	Yes	No		Currently yes	
	Yes, Lansweeper	Yes	No			
ny things takes too lon	Yes	Yes	No		Yes, with som adjustmen	-
	no	Yes	No		I don't have access	
	Yes	Yes	No		Yes	
	no	Yes	No		yes	
	No					

SGTS Nordic & Baltic
PC Factory Reporting Questionnaire

9
Responses

02:09
Average time to complete

Active
Status

View results

Open in Excel

1. Are You currently satisfied with the reports you can get from the PC Factory?

More Details

Yes 8
No 1

2. If you answered "NO" in question 1: What is the reason for Your dissatisfaction?

More Details

1
Responses

Latest Responses

3. Do You know where or how You can get the reports related to Your business' PCs?

More Details

9
Responses

Latest Responses
"No"
"no"
"Yes"

4. Are You satisfied with the level of data (e.g. PC fleet information) provided for you about your business ?

More Details

Yes	8
No	0

5. Are You missing any data that you consider important for You and Your business?

More Details

Yes	0
No	8

6. If You answered "Yes" in question 5: What data are You missing?

More Details

0
Responses

Latest Responses

7. Do You consider that PowerBI, current data source for SGTS service data, is sufficient for You to retrieve the necessary data and reports?

More Details

7
Responses

Latest Responses
"yes"
"Yes"

8. Do You have any further comments or wishes about the PC Factory service and/or provided service data?

More Details

3
Responses

Latest Responses
""

Data 1 – Workshop Field Notes

PC FACTORY DATA MANAGEMENT
WORKSHOP DETAILS

- ❑ Started: 22.3.2021 @ 13:05
- ❑ Ended: 22.3.2021 @ 14:43
- ❑ Participants: Hans KANERVA, Jere KALLIONIEMI, Tuomas PAAVOLA, Jukka-Pekka KNUUTINEN
- ❑ Location: Vantaa office & Teams
- ❑ Recorded in Teams: https://web.microsoftstream.com/video/b058626a-7d7c-47b0-a315-d553effc8e7c

DATA MANAGEMENT Nordic & Baltic SAINT-GOBAIN

PC FACTORY DATA MANAGEMENT
WHY A WORKSHOP?

A fool with a tool is still a fool

- Grady Booch (famous software engineer)

3 DATA MANAGEMENT Nordic & Baltic SAINT-GOBAIN

PC FACTORY DATA MANAGEMENT
WHY A WORKSHOP?

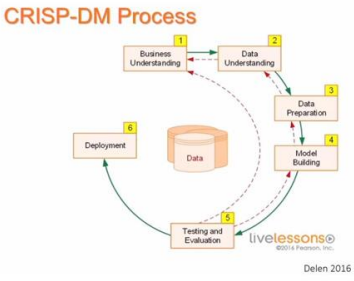
- Current Status
  - Data is widespread and not used
  - Manual work to get an overview
  - Decisions made on assumptions not on facts
  - Not in 100% control of the delivered service
  - Reporting (internal/external) missing
- Need to take actions on getting on track with utilizing the data available!

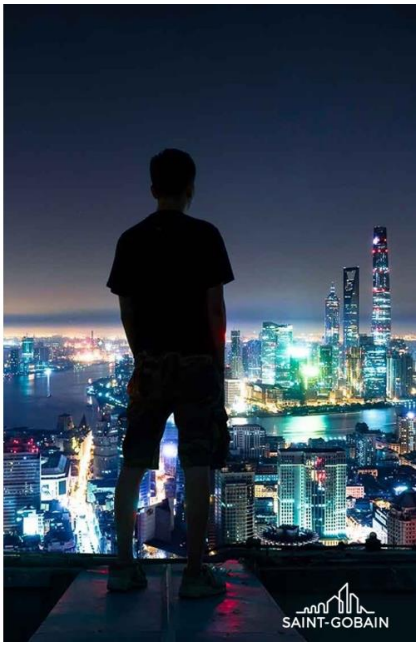
PC FACTORY DATA MANAGEMENT

WORKSHOP - WHAT NOW?

In this exercise we'll brainstorm and identify:

1. **What data sources do we have?**
2. **For what shall we use the data?**
3. **How shall we present the data?**





WORKSHOP - WHAT DATA SOURCES DO WE HAVE?

- **Tracking of data sources in the data card template created - DONE**
- From where do we get data? (DHL, Lenovo, SAP, etc.)
- In which format are we receiving the data? (Excel, .csv, e-mail, etc.)
- What information is the data source containing?
- Where is it stored today?

WORKSHOP – FOR WHAT SHALL WE USE THE DATA?

- What data shall we visualize?
  - External
    - Delivery information & time per order
    - Stock status & Lenovo batch delivery time
    - Recycling report/list
    - Overall statistics – PC Factory
  - Internal
    - Lenovo – Expected Delivery Dates
    - ServiceNow – Open orders & ordered equipment (extracted from ticket data)
    - Order quantities per device (PCs, docking stations, mice, keyboards, etc.) (stock management)
    - DHL – Shipment trends per countries
    - T1A – Recycled products & quantities (can it be tied to LanSweeper report for BU information?)
    - T1A – Recycled products: Scrapped vs second life

WORKSHOP – FOR WHAT SHALL WE USE THE DATA?

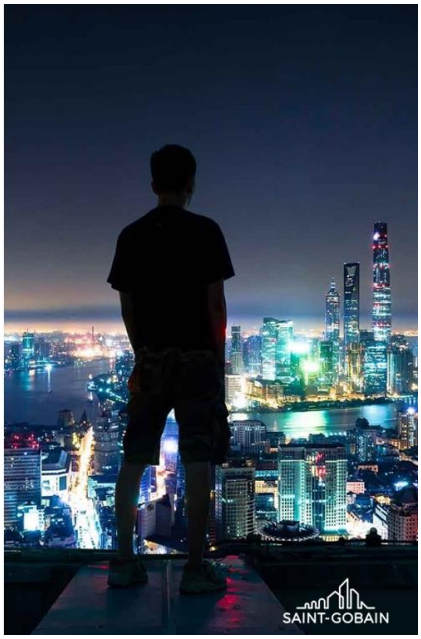
- What is the preferred format of presenting the data?
  - Jere: Visualized PowerBI reports
  - Tuomas: Visualized PowerBI reports but need for access to raw data in Excel
  - JP: Depends on the usage
- Dashboards are nice but a need for data view in list format as well
  - Dates
  - Details

PC FACTORY DATA MANAGEMENT

WORKSHOP - WHAT NEXT?

A summary of the workshop will be compiled, and results analyzed. The results will be taken in consideration during the establishment of the data management plan for PC Factory

Thank you for your contribution!



Data 1 – Workshop Data Source Cards

DATA SOURCE CARD

Data Source: « Lenovo »

What is the format of the data?
(.xls/.csv/E-mail, etc.)

.html (default)
Format can be chosen

How is the data received?
(E-mail, automatic/manual export/import)

Email(automatic)

What information is the data containing?

1. Products ordered
2. Estimated shipment/delivery dates

Where is the data stored?

PC Factory e-mail box → USS Teams - Files

Future: SGTS Universe Teams/SharePoint - Files

Other information / Comments?

Reports configured in the OVP portal and format chosen.

3 DATA MANAGEMENT Nordic & Baltic SAINT-GOBAIN

DATA SOURCE CARD

Data Source: « ServiceNow »

What is the format of the data?
(.xls/.csv/E-mail, etc.)

« .xls»

How is the data received?
(E-mail, automatic/manual export/import)

manual export / scheduled report in e-mail »

What information is the data containing?

1. Customer orders (Order details)
2. Customer details

Where is the data stored?

Not agreed, ServiceNow connector already existing in PowerBI

Other information / Comments?

Need for visualization of order data
Creation of reports in PowerBI by Jere & Tuomas

6 DATA MANAGEMENT Nordic & Baltic SAINT-GOBAIN

DATA SOURCE CARD

Data Source: « LanSweeper »

What is the format of the data?
(.xls/.csv/E-mail, etc.)

« .xls »

How is the data received?
(E-mail, automatic/manual export/import)

« manual import »

What information is the data containing?

1. Asset specific data (incl. Location, warranty, HW details, SW details, latest user)

Where is the data stored?

Connector existing between LS and PowerBI

Other information / Comments?

Data quality not guaranteed, issues with device scanning (Lsagent) → **data collection has to be fixed**
An existing report with necesary data (**to be checked**)

Data 2 – Interview field notes

Example 1

Building the Proposal - Interview

Date & Place:	14.11.2021
Interviewer:	Hans Kanerva
Interviewee:	Jörgen Liberg, SGTS NB Director

1. Is the Data Management Plan (DMP) proposal understandable? Do you understand what you read?	Yes, the DMP is understandable but I have some remarks and comments, but we'll take it later on.

2. Do you consider that the DMP makes sense?	Yes, it does make sense.

3. Do you have any improvement or change suggestions?	1. Please remember to describe at the beginning your abbreviations like (SGTS stands for = Saint-Gobain Technology Services), ITSM, SQL, etc. **HKA: Yes, this is in internal document so the abbreviations should be clear. I will populate a glossary in the end of the document.** 2. What is SGTS Universe? Missing the definition. **HKA: I will add it in the glossary as well.** 3. Maybe a Chapter in the beginning of "Why DMP is important "The why/The Need"" and why you have chosen this for your thesis (Are you talking about Data in general or Master data or even Block chain) – Might be under "Data Definitions". **HKA: The reasoning and justification is in the Master's Thesis document together with the business problem and context.** 4. What about GDPR ? **HKA: We do not populate personal data in the PC Factory. The only personal data is collected from SGTS or Group tools like Lansweeper or ServiceNow**

Building the Proposal - Interview

	which are already listed in Onetrust and no personal data is used for either reporting nor visualization. 5. A chapter of validation of Data (how will you ensure to data are correct and valid?) or this is not described in Maturity model by Howson & Duncan. **HKA: The data accuracy within the PC Factory will be on the data owner's responsibility. Each data source needs an owner and the data source must be documented. The data accuracy and cleanliness must to be validated by the data owner according to data owner responsibilities that are defined in the DMP policy.** 6. Figure 5. I would changes the "PowerBi Server" to "Data warehouse" **HKA: Actually, the SQL database is the data warehouse where the PowerBI is the tool we use for data processing and basic analytics. I will update the graph with data warehouse in brackets.** 7. Chapter 5 – Data Maturity assessment – Very good 8. Chapter 7.1 – Action Plan – Very good

4. Any further comments?	Remember that a DMP is a living document (continues researching). **HKA: yes, this is stated in the DMP as well. This DMP can't ensure the continuous updating and development of the DMP. This has to be considered within the organization and task should be allocated to the PC Factory Team Leader.**

Example 2

Building the Proposal - Interview

Date & Place:	17.11.2021 / Teams Video conferencing
Interviewer:	Hans Kanerva
Interviewee:	Jere Kallioniemi, PC Factory Team Leader

1. Is the Data Management Plan (DMP) proposal understandable? Do you understand what you read?	Yes, the document is understandable but it would be best if the actual topic of DMP would start from *Chapter 1* instead of *Chapter 3*. The setup is understandable though as it is our company template.
2. Do you consider that the DMP makes sense?	Yes, it does make sense and we'll gladly apply the DMP in PC Factory. We just need to review the DMP together in detail before initiating the implementation.
3. Do you have any improvement or change suggestions?	One typo on page 4 --> DATA STORGAE (to be corrected!) Chapter 4.2, point 1b --> ***What data is being*** *used within the organization?*

Building the Proposal - Interview

4. Any further comments?	A new data source has been identified: *The Label Database*. This database contains data of the installed PCs including PC name and PC serial number.

5. Do you approve document version 0.3 as the official version of the DMP?	The version 0.3 is approved by Jere in an e-mail sent on Wednesday the 17th of November.

The Proposal – Data Management Plan – Version 0.1

SAINT-GOBAIN

Data Management Plan

SGTS Nordic & Baltic

PC FACTORY

VERSION 0.1

Table of Contents

1 Document control

1.1 Aim of this document

This document describes the background and action plan for the PC Factory *Data Management Plan (DMP)*.

1.2 Document location

This document is only valid on the day it was printed. The actual document can be found in the PC Factory folder which is placed under the *USS* department folder located in *SGTS Universe* (Sharepoint).

1.3 History

Date of this revision:	**Date of next revision:**

Revision date	**Ver. / Rev.**	**Summary of changes**	**Author**	**Approval date**
22.10.2021	0.1	Creation of document	Hans Kanerva	

1.4 Reviews

Name	**Title**
Jörgen Liberg	SGTS NB Director
Jere Kallioniemi	SGTS NB PC Factory Team Leader

1.5 Approvals

Name	**Title**
Jörgen Liberg	SGTS NB Director

1.6 Distribution

SGTS Nordic & Baltic PC Factory team

2 Introduction

This *Data Management Plan*, DMP from now on, is created on the basis of research conducted by Hans Kanerva (SGTS NB) in the Metropolia University of Applied Sciences Master's Thesis "*Data Management Plan*". The DMP is an actual summary and action plan to correct issues observed during the current state analysis of data management within the PC Factory. The findings of the current state are summarized in *Chapter 4* in this document but the detailed research can be found in the Master's thesis document.

The DMP consists of three main chapters:

Chapter 3: Context and Objective

Chapter 4: Implementation of the Data Management Plan

Chapter 5: The Action Plan

Chapter 3 describes *why* this DMP is needed while *Chapter 4* describes *what* is included and must be considered when facilitating the DMP. *Chapter 5* answer the question *how* through a ready action plan.

Chapter 6 is a summary of the DMP where *Chapter 7* is dedicated for the appendixes.

3 Context and Objective

The DMP objective is to set a framework around data management within the SGTS Nordic & Baltic PC Factory, and to provide an action plan with necessary templates to achieve this. The PC Factory activity was set up in 2018 but the establishment of a proper data management was postponed at the time to the future due to higher priorities. This DMP is aimed at correcting current challenges with data management within the PC Factory activity but also to prepare the PC Factory data management for future data analytics development activities.

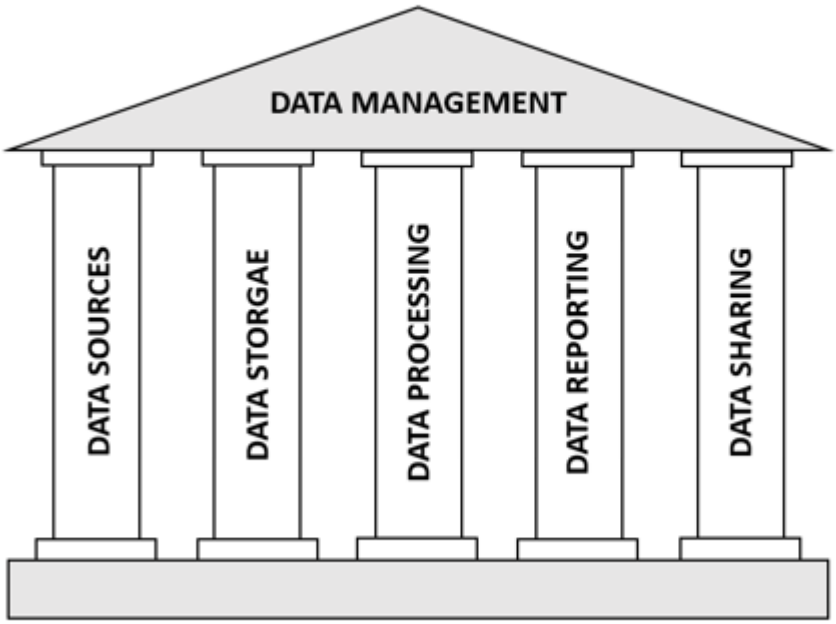


Figure 1. The Five Pillars of the PC Factory Data Management.

The data management in PC Factory is built on five pillars, which are presented in *Figure 1*.

Main challenges with data management in the PC Factory and found during the current state analysis is compiled in *Figure 2*. Main challenges are that the data isn't currently used in an efficient way and necessary reporting and dashboards, descriptive data analytics, is missing. Control of the activity has to be raised through the use of data through the descriptive data analytics means and automation while reducing the need for manual work and possibilities for human errors. There is also a lack of current understanding of the use of data and tools available for managing the data, where clear internal policies around the topic is also missing.

SAINT-GOBAIN

Stage	Challenge	Impact
Data Storage	Data storage location unknown by the staff. No internal agreement, practice, or policy for common data storage practices.	1. Impossible to utilize the data in an efficient way as the staff can't find the data. 2. Loss of staff efficiency due to lost working time.
Data Processing	A manual data processing process.	1. Loss of staff efficiency due to lost working time. 2. Increased risk of human errors.
Data Processing	Lack of data processing knowledge and skills	1. Unutilized system and data. 2. Loss of staff efficiency due to lost working time.
Reporting	Lack of reporting knowledge and skills	1. Missing reports for both internal and external stakeholders. 2. Loss of staff efficiency due to lost working time as the staff use extra time on creating non-relevant reports.
Reporting	Lack of reporting/reports	1. Missing insight to financial performance which has an impact on the decision-making. 2. Customer transparency is missing and well as customer report for supporting the customer decision-making.
Data Sharing	No Internal agreement, practice, or policy for common data sharing practices.	1. Loss of data. 2. Loss of staff efficiency due to lost working time. 3. Impact on decision-making as necessary data is not found or inaccessible.

Figure 2. Key challenges and Business Impact.

The PC Factory is currently using data from six (6) different sources, which are listed in *Figure 3*. The data from these sources are spread around the IT environment and used only on ad-hoc basis. The DMP will assist in orchestrating the data from these sources and make the use possible for the organization.

Data Source	Data Format	Alternative Data Format	Data Delivery	Data Storage	Format Approved	Delivery Approved	Storage Approved
DHL	Excel	N/A	E-mail	Network drive	Yes	Yes	Yes
Lenovo	HTML	Excel	E-mail	E-mail	Yes	Yes	No
Tier1Asset	Excel	N/A	E-mail	E-mail	Yes	Yes	No
ERP	Excel	N/A	Manual Export	PC	Yes	No	No
ServiceNow	Excel	N/A	Manual Export	PC	Yes	No	No
LanSweeper	Excel	N/A	Manual Export	PC	Yes	No	No

Figure 3. Data Sources used by PC Factory.

4 Implementation of the Data Management Plan

The implementation of the DMP is divided in six activities that were identified in the research. The six data management activities are adjusted to the PC Factory activity and described in the following section of the document.

4.1 Education

Education of concerned people is considered essential for a working data management. The lack of skills and knowledge in both *Data Processing* and *Reporting* were identified as challenges in the research of the current state of data management within the organization. There is a need to educate concerned people in:

1. Internal data policies and practices
2. Internal data management tools
3. Internal data management responsibilities

The training within the team can be conducted either through a *seminar* with the appropriate trainer or by *reading* internal documentation. The latter poses a risk as it is difficult to control people's reading and therefore the first alternative of ***seminar* is the proposed as the primary mean of education** in this context. The seminar can be conducted either as classroom training or online.

Education preparations are required and the following missing points must be prepared and finalized before the seminar:

1. Formalize the internal *Data Management Policy*
2. Formalize data sources in data cards, incl. data ownership
3. Formalize the *Information Service* -service description
4. Grant access to the data management tool, *PowerBI*

The proposal for the seminar lecturers are:

1. SGTS NB PowerBI Solution Owner
2. PC Factory Team Leader

Additional external experts can be added as lecturers if it is considered needed.

4.2 Data Management Policies and Strategy

The data management policies and strategy defines the basic common rules of how to deal with data. The policy should define how the concerned staff should use data. The *Data Management Policy* require a formalized document which is distributed within the organization and finally reviewed together with concerned staff. The policy is a pre-requisite for *Education*.

The *Data Management Policy* should contain the guidelines for the following:

1. The purpose of data within the organization
    a. What is data used within the organization?
    b. Why is data used within the organization?
2. Data location guidelines, incl. *Data Card* location
    a. Where is data located and stored?
3. Data processing guidelines
    a. What is data ownership?
    b. Who can have access to the data?
    c. How the data is accessed and processed?
4. Data management tool guidelines
    a. What tool is used within the organization for the purpose?
    b. How can the tool be accessed?
5. Data backup & restore guidelines
    a. What data is backed up?
    b. How is the data backed up?
    c. How is the data restored in case of data loss?
6. Data sharing and reporting
    a. What data should be shared with internal and external stakeholders?
    b. How is data shared with concerned stakeholders?

Preparations for the *Data Management Policy* are:

1. Creation of the initial document
2. Review by relevant stakeholders:
    a. SGTS Director – general data management rules
    b. SGTS PowerBI Solution Owner – data location, data processing, data management tool(s)
    c. SGTS Head of Infrastructure – data backup & restore
3. Formalization & publication of the document
    a. Placement in SGTS Universe (Sharepoint)
    b. Publication SGTS Intranet in Yammer

The proposal for the document creator is the *SGTS NB User & Site Support Director*.

4.3 Data Definitions

The *data definitions*, in this context, is the appropriate documentation of the data sources. The *Data Card* template, appendix *A*, should be used for as a documentation template per data source. The templates include all necessary information for proper documentation of a data source.

The data definition ensure that each data source is tagged with an appropriate name, owner and other criteria that is necessary for future management and usage of the data source. The *Data Card* will also ensure that documentation of data sources exist which means that the organization is less dependent on staff knowledge and persons memory.

It is advised that a summary of existing and/or future data sources are identified in a list (example: *Figure 3*) and then a *Data Card* is created for each data source. The storage of the *Data Cards* should follow the guidelines communicated in the *Data Management Policy*.

The *Data Card* should be filled and completed by the identified *Data Owner*. It is suggested that the *PC Factory Team Leader* initiates the kick-off or launch of completing the *Data Cards*.

4.4 Data Management Tools

The data management tool used within SGTS Nordic & Baltic is *Microsoft PowerBI*. This is the only tool that should be used for the data management purpose and is chosen and validated as the standard by the organization management. The data management tool guidelines are communicated in the *Data Management Policy*. The organization staff can get access to PowerBI upon request to the *PowerBI Solution Owner*.

The PowerBI topology with data source types are presented in *Figure 5*. Three types of data sources can be used for producing reports and visualization in PowerBI:

1. An Excel file with columns and rows
2. Data from the SQL database
3. Data through an application integration

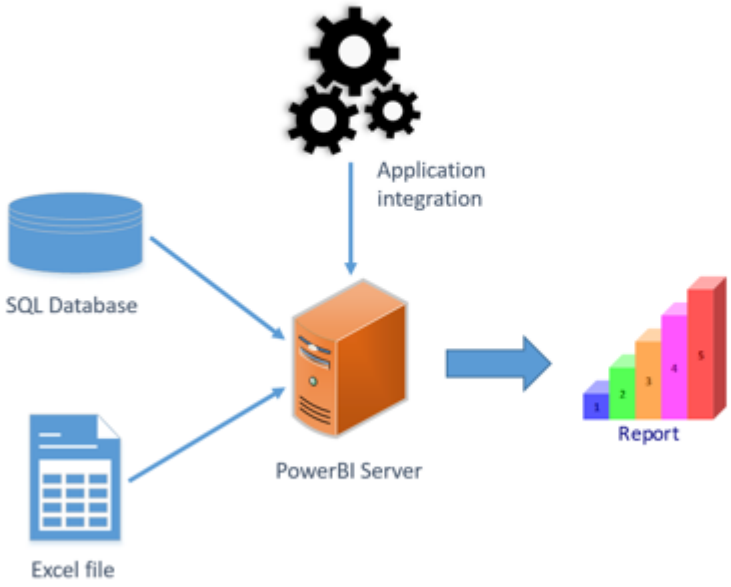


Figure 5. SGTS NB PowerBI Topology

Current existing application integrations are:

1. Lansweeper
2. ServiceNow

Current and new application integrations are managed by the *PowerBI Solution Owner* and each need for a new integration must be addressed by raising a request in the ITSM tool.

The *SQL database* can be used as a data source if the data is first imported to the database. Database imports are managed, as well, by the *PowerBI Solution Owner*.

Excel-files can be used as such as data sources and they are managed by the data owner defined in the *Data Card* for the specific data source. The storage guidelines for the Excel-files are defined in the *Data Management Policy*.

The action needed in this section is related to *education* of concerned staff and raising awareness of the tools used among concerned users.

4.5 System Development Support

The *system development support* in this context refer to the activity where the data owner must ensure that the data quality is good and ensure well-coordinated and -facilitated changes if such are concerning the data or data source. The data owner must be the primary contact for any changes related to any aspect of the data itself or the data source.

It is essential that a *data owner* is identified for each data source, documented in a *Data Card* and that the data owner responsibilities are formalized in the *Data Management Policy*. This will ensure that the data owners are informed about the responsibility during the *education*. Education will also ensure that the concerned people knows about the data management policy and data ownership model implemented within the organization as well as the location of relevant documentation related to the data sources.

4.6 Information Service

The *information service* describes how information and data is shared within the business in an efficient way. The information service and its content, including data and report inventory, must be formalized in a document or system. The document can be an internal service description, that is used for other services within the organization. The system can be an existing system like the ITSM tool Knowledge Base, ServiceNow, in this case.

It is recommended to start with a service description document, as the current ITSM system Knowledge Base isn't used by all concerned stakeholders. The usage of a service description document will require a periodic review and it is recommended that the document is reviewed at least twice per year.

The service description document shall contain at least the following details:

- Data & information sharing tool details
- Data & information sharing tool access details
- Data & data owner inventory
- Report & report owner inventory

It is recommended that the *Information Service* -service description is created and maintained by the *PC Factory Team Leader*. The document doesn't have to contain all information but information references can be used instead.

The *Information Service* -service description shall be completed with the creation of a *Report Card* per report. The *Report Card* will ensure that each report is defined and documented with necessary details including an appropriate report name & creator, significance statement, used data sources and other criteria that is necessary for future management and usage of the report.

5 Data Maturity Assessment

Data maturity assessment assist the organization with understanding the current data maturity level in addition to identifying the necessary requirements for reaching the following maturity level. The data maturity assessment must be done in the beginning of the development and reassessed during the development journey.

The data maturity model used in this context is the model created by Howson & Duncan at Gartner in 2015, see *Figure 6*. An alternative model was used in the research, in addition to the in one in *Figure 6*, to validate the relevancy of the data maturity

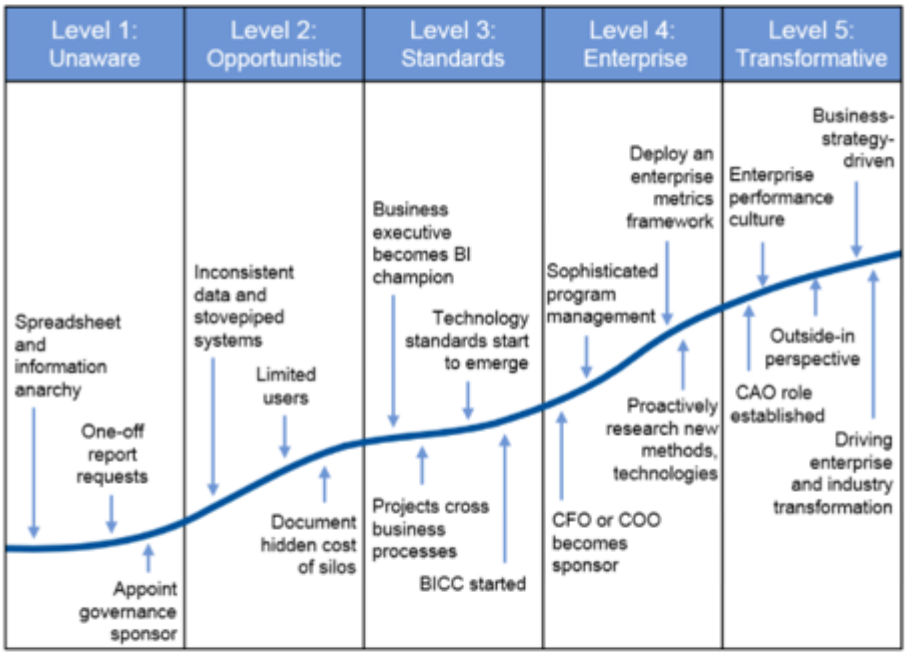


Figure 6. The BI and Analytics Maturity Model by Howson & Duncan (https://www.gartner.com/en/documents/3136418/itscore-overview-for-bi-and-analytics)

It was assessed during the current state analysis that the **current data maturity level is on the lowest level (Unaware)** as awareness and necessary coordinated activities are currently missing.

The descriptions of the first two levels are as following:

Level 1: *Unaware* – The demand for business intelligence and analytics is ad hoc. The decision-making is unstructured and lacking processes and practices. There is no information infrastructure, processes, or performance metrics.

Level 2: *Opportunistic* – Business units manage their own business intelligence and analytics projects. These business units have their own tools and information infrastructure. The business units have dedicated IT organization. The process modeling is usually missing, and the usage of the existing tools isn't efficient. Analytics outcome are mainly reporting, and dashboards which contain business or domain specific content.

The objective for the organization should be to reach Level 2 – Opportunistic, in order to be able to develop further in the future.

6 The Action Plan

The *Action Plan*, appendix 1, is prepared based on the recommendations in chapter 4. The *Action Plan* is based on a project Work Breakdown Structure (WBS) including all necessary elements for an *Action Plan*. The necessary tasks and preparations, including creation of policies and documents, are included in the WBS. The *Action Plan* will not ensure the continuous maintenance and improvement of data management within the organization, which means that the operational part must be validated and agreed internally after the action plan tasks are completed.

It is required that timing and resourcing is evaluated for each action point in the beginning of the project

It is recommended that a *Project Manager* (PM) is assigned to guarantee the completion and coordination of the tasks in the *Action Plan*.

7 Appendixes

7.1 Action Plan

	UPDATED BY:	"Initials"	DATE:	xx/xx/20xx								
#	Phase	Task	Subtask 1	Expected Outcome	Responsible	Due Date	Status	Progress / Comments	Progress %	Priority	Milestone (Yes / No)	Dependencies #
1	1 - Education	Prepare education seminar	Book & confirm lecturer(s)	Confirmation from lecturer	PC Factory Team Leader		Not Started		0	3 - Low	No	-
2	1 - Education	Prepare education seminar	Book & confirm facility/room	Facility confirmation	PC Factory Team Leader		Not Started		0	3 - Low	No	-
3	1 - Education	Prepare education seminar	Create training materials	Set of training materials	PC Factory Team Leader		Not Started		0	3 - Low	No	-
4	1 - Education	Prepare education seminar	Ensure formalization of Data Management Policy	Proof of formalized Data Management Policy	PC Factory Team Leader		Not Started		0	3 - Low	No	18
5	1 - Education	Prepare education seminar	Ensure formalization of Information Service	Proof of formalized Information Service	PC Factory Team Leader		Not Started		0	3 - Low	No	34
6	1 - Education	Prepare education seminar	Ensure user access to PowerBI	Confirmation from system owner	PC Factory Team Leader		Not Started		0	3 - Low	No	-
7	1 - Education	Hosting seminar	Confirm when seminar is completed	Attendance & completion confirmation	PC Factory Team Leader		Not Started		0	3 - Low	Yes	
8	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about the purpose of data within the organization	Internal data collection & definition of data purpose within the organization	USS Director		Not Started		0	1 - High	No	-
9	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data location guidelines	Internal data collection & definition of data location guidelines	USS Director		Not Started		0	1 - High	No	-
10	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data processing guidelines	Internal data collection & definition of data processing guidelines	USS Director		Not Started		0	1 - High	No	-
11	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data management tool guidelines	Internal data collection & definition of data management tool guidelines	USS Director		Not Started		0	1 - High	No	-
12	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data backup & restore guidelines	Internal data collection & definition of data backup & restore guidelines	USS Director		Not Started		0	1 - High	No	-
13	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data sharing & reporting guidelines	Internal data collection & definition of data sharing & reporting guidelines	USS Director		Not Started		0	1 - High	No	-
14	2 - Data Management Policy	Creation of the Data management Policy	Creation of the Data Management Policy document	Creation of the Data Management Policy document	USS Director		Not Started		0	1 - High	No	8 - 13

15	2 - Data Management Policy	Send document for review	-	Send the document to be reviewed according to the reviewer list in the document	USS Director		Not Started		0	1 - High	No	14
16	2 - Data Management Policy	Update of the Data Management Policy	-	Update the document in case document validator(s) requesting updates	USS Director		Not Started		0	1 - High	No	15
17	2 - Data Management Policy	Approval of the Data Management Policy	-	Final approval of the document	USS Director		Not Started		0	1 - High	Yes	16
18	2 - Data Management Policy	Formalize the Data Management Policy	-	Publication & communication of the document within the organization	USS Director		Not Started		0	1 - High	No	17
19	3 - Data Definitions	Identification of data owners	-	A data owner should be identified for each data source	PC Factory Team Leader		Not Started		0	1 - High	No	-
20	3 - Data Definitions	Creation of a *Data Card* for each data source	-	A data card should be created for each data source and stored according to the data management policy	Data Owners		Not Started		0	1 - High	Yes	19
21	4 - Data management Tools	Preparation of education materials for the seminar	-	Training materials as presentation and/or video for the eduction seminar	PowerBI Solution Owner		Not Started		0	2 - Medium	No	-
22	5 - System Development Support	Organization communication about *Data Owners*	-	Proof of organization communication about *Data Owners*	USS Director		Not Started		0	3 - Low	No	-
23	6 - Information Service	Preparation for the *Information Service* - service description	-	Data collection for: Data & information sharing tool details, Data & information sharing tool access details, Data & report owner inventory, Report & report owner inventory	PC Factory Team Leader		Not Started		0	1 - High	No	-
24	6 - Information Service	Creation of the *Information Service* - service description	-	Creation of the initial document	PC Factory Team Leader		Not Started		0	1 - High	No	23
25	6 - Information Service	Identification of needed reports	-	Creation of an inventory of report needs. From all stakeholders: internal, external (customers, auditors)	PC Factory Team Leader		Not Started		0	1 - High	No	-
26	6 - Information Service	Creation of *Report Cards*	-	Creation and storage of a report card per needed report for definition and documentation purposes	PC Factory Team Leader		Not Started		0	3 - Low	No	25
27	6 - Information Service	Creation & publication of reports in PowerBI	-	Creation and publication of reports and visualizations in PowerBI together with	PC Factory Team Leader		Not Started		0	3 - Low	Yes	26

				the Powerbi Solution Owner								
28	7 - Data Maturity Assessment	Document the initial data maturity assessment	-	Create a Data Maturity Assessment document	USS Director		Not Started		0	1 - High	No	-
29	7 - Data Maturity Assessment	Re-assess the data maturity	-	Create a Data Maturity Assessment document	USS Director		Not Started		0	2 - Medium	No	28
30	8 - Preparation for the Action Plan	Assignment of a Project Manager	-	Official announcement of a Project manager for the coordination and follow-up of the Action Plan	USS Director		Not Started		0	1 - High	No	-

SAINT-GOBAIN

7.2 Data Card

DATA CARD	
Name or Label	
Synonyms	
Significance Statement	
Data Format	
Data File Format	
Value List	
Validation Criteria	
Valid Operations	
Ownership	
Users	
Source	
Comments	
Creation Date	
Author	
Last Updated Date	

Description	
Name or Label	A unique name or label (unique identifier).
Synonyms	Any synonyms or aliases for the data object.
Significance Statement	A description of why the data is relevant.
Data Format	Used data formats.
Value List	A valid value lists describing what kind of values are expected in the data object
Validation Criteria	A valid validation criterion describing what kind of values are expected in the data object.
Valid Operations	Valid operations describing for which operations the data is used.
Ownership	Data ownership information.
Users	A description of for who the data is for.
Source	A data source description.

Source File Format	Description in which file format the data is populated or generated
Comments	Any relevant comments
Date Created	Date when the data card was initially created
Author	Name of author of the datacard
Last Updated	Date of the last update done in the file.

File Naming Convention	
File Naming Convention	*SGTS NB - Data Card - "Label"*

SAINT-GOBAIN

7.3 Report Card

REPORT CARD	
Name or Label	
Meta data	
Significance Statement	
Data Sources (data card names)	
Published in	
Ownership	
Users	
Comments	
Creation Date	
Author	
Last Updated Date	

Description	
Name or Label	A unique name or label (unique identifier)
Meta data	Descriptive words for the report
Significance Statement	A description of why the report is relevant
Data Sources (data card names)	Names of data sources that are used for populating the report
Published in	Tool where the report is published
Ownership	Report ownership statement
Users	List of users or user groups
Comments	Any relevant comments regarding the report
Creation Date	Date when the report card was initially created
Author	Name of author of the report card
Last Updated Date	Date when the report card was last updated

File Naming Convention	
File Naming Convention	*SGTS NB - Report Card - "Label"*

Data 3 – Interview field notes

Example 1:

Validation of the Proposal - Interview

Date & Place:	25.11.2021
Interviewer:	Hans Kanerva
Interviewee:	Jere Kallioniemi, PC Factory Team Leader

1. What is your impression and thoughts about the Data Management Plan (DMP) version 1.0?	A good initiative as we have much data. This is a professional approach that is supporting our activity. I'm very positive to this.
2. Will it be worth implementing the DMP in the PC Factory?	Yes, it is. We can see the value.
3. Do you foresee any issues, problems, or challenges with the implementation of the proposal?	I can see that time will be an issue. Time is limited and the challenges will be related to time allocation for all aspects like training and preparation for the deployment.

Validation of the Proposal - Interview

4. What benefits to the organization do you expect when or if the DMP is implemented?	Professional and standardized approach to data management. Efficiency in data processing will also be improved in the future based on the DMP.

5. Do you foresee any negative effects on the organization with the implementation of the DMP?	The challenge might be related to the change, which will require change management. I do not foresee any further issues and challenges that we can't handle.

6. Do you have any further comments regarding any part of the DMP, its implementation, or impact on any part of the organization?	The document is clear but very technical. A good document.

Example 2:

Validation of the Proposal - Interview

Date & Place:	29.11.2021
Interviewer:	Hans Kanerva
Interviewee:	Jörgen Liberg, SGTS NB Director

1. What is your impression and thoughts about the Data Management Plan (DMP) version 1.0?	This is a very good document and tackling an issue we've had for quite some time. I like it!

2. Will it be worth implementing the DMP in the PC Factory?	Yes, it will be worth the effort.

3. Do you foresee any issues, problems, or challenges with the implementation of the proposal?	Implementation is one thing but the operations and maintaining the process afterwards is the challenge. This will require a change in how we operate and that all employees in the PC Factory buy the idea of the DMP.

Validation of the Proposal - Interview

4. What benefits to the organization do you expect when or if the DMP is implemented?	I see several benefits. We get in control of data which we've been lacking. We can now document both data sources and reports in ways that makes it possible for us track changes and further develop the data management within the team activity. I get personally added value through activity reporting which give me detailed insight of the activity and costs. This is a process that can be implemented in other parts of our organization as well, if the initial implementation, test that is, is successful.

5. Do you foresee any negative effects on the organization with the implementation of the DMP?	People and willingness to change work methods is the only thing that comes to mind. This might have a negative effect in the beginning.

6. Do you have any further comments regarding any part of the DMP, its implementation, or impact on any part of the organization?	All in all, this has been a very good exercise for our organization, and as already said, is targeting an organization wide problem. It is great to get a detailed study and facts with best practice behind the DMP so that it isn't just someone's thought about what is good to do.

The Final Outcome – Data Management Plan – Version 1.0

SAINT-GOBAIN

Data Management Plan

SGTS Nordic & Baltic

PC FACTORY

VERSION 1.0

SAINT-GOBAIN

Table of Contents

1 Document control

1.1 Aim of this document

This document describes the background and action plan for the PC Factory *Data Management Plan (DMP)*.

1.2 Document location

This document is only valid on the day it was printed. The actual document can be found in the PC Factory folder which is placed under the *USS* department folder located in *SGTS Universe* (Sharepoint).

1.3 History

Date of this revision:	Date of next revision:

Revision date	Ver. / Rev.	Summary of changes	Author	Approval date
22.10.2021	0.1	Creation of document	Hans Kanerva	
1.11.2021	0.2	Minor updates in the document based on Data 2 input and feedback (SGTS Director)	Hans Kanerva	
9.11.2021	0.3	Minor updates in the document based on Data 2 input and feedback (PC Factory Team Leader)	Hans Kanerva	
9.11.2021	1.0	Offical version approved	Hans Kanerva	9.11.2021

1.4 Reviews

Name	Title
Jörgen Liberg	SGTS NB Director
Jere Kallioniemi	SGTS NB PC Factory Team Leader

1.5 Approvals

Name	Title
Jörgen Liberg	SGTS NB Director

1.6 Distribution

SGTS Nordic & Baltic PC Factory team

2 Introduction

This *Data Management Plan*, DMP from now on, is created on the basis of research conducted by Hans Kanerva (SGTS NB) in the Metropolia University of Applied Sciences Master's Thesis "*Data Management Plan*". The DMP is an actual summary and action plan to correct issues observed during the current state analysis of data management within the PC Factory. The findings of the current state are summarized in *Chapter 4* in this document but the detailed research can be found in the Master's thesis document.

The DMP consists of three main chapters:

Chapter 3: Context and Objective

Chapter 4: Implementation of the Data Management Plan

Chapter 5: The Action Plan

Chapter 3 describes *why* this DMP is needed while *Chapter 4* describes *what* is included and must be considered when facilitating the DMP. *Chapter 5* answer the question *how* through a ready action plan.

Chapter 6 is a summary of the DMP where *Chapter 7* is dedicated for the appendixes.

2.1 Data Management

Data management is usually needed in medium size or large companies where finding data is time consuming. In small companies the data and applications are close to the users and help for accessing the data is close as well. The data management need arise when the organization grow and the accessibility to data is delayed because of scattered knowledge, delay in the decision-making process, or lack of data ownership.

There are different definitions of data management but one of the simplest definitions is: "data management is a function which ensure that data, information in other words, is managed in a way so that it is suitable for communication, reading, or processing. The definition data management is often associated with enterprise-wide data services, but it hasn't to be so in reality.

Working data management has a lot of benefits for a business but direct costs savings shouldn't be expected. There are two specific areas where business benefits can be expected. The first area is

business related and the second area is IT and systems related. Data availability and data quality across the business is a great benefit as it supports the decision-making process and ensure improvement in the overall business efficiency. Available data ensure that the business gains can be measured through financial figures, which make the benefit tangible and visible. IT and systems related benefits are collected through costs savings in application maintenance and increased development productivity.

Data management can support the business to comply with regulatory, privacy, data protection, safety, and security requirements. Businesses must be able to provide authorities verifiable data about regulatory compliance. The business can be fined in case it fails to comply with these requirements. The business must also comply with privacy and data protection requirements set by authorities. The business and customer data must be protected from unauthorized access and misuse as this can lead to criminal activity as identity thefts, sabotage, breach of immaterial rights.

This document is created for making data management tangible and easy to implement by following the pre-defined action plan in appendix 1.

3 Context and Objective

The DMP objective is to set a framework around data management within the SGTS Nordic & Baltic PC Factory, and to provide an action plan with necessary templates to achieve this. The PC Factory activity was set up in 2018 but the establishment of a proper data management was postponed at the time to the future due to higher priorities. This DMP is aimed at correcting current challenges with data management within the PC Factory activity but also to prepare the PC Factory data management for future data analytics development activities.

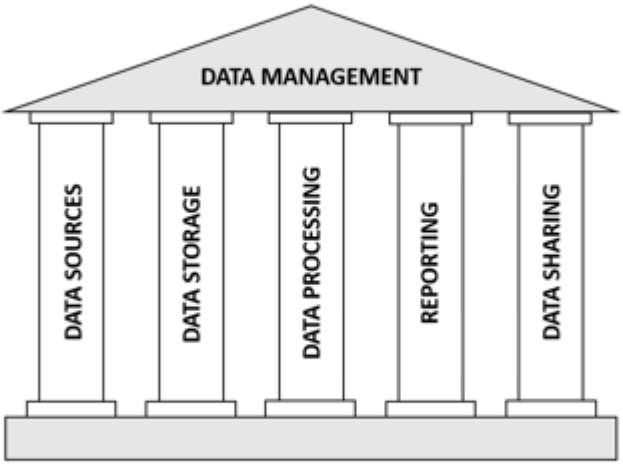


Figure 1. The Five Pillars of the PC Factory Data Management.

The data management in PC Factory is built on five pillars, which are presented in *Figure 1*.

Main challenges with data management in the PC Factory and found during the current state analysis is compiled in *Figure 2*. Main challenges are that the data isn't currently used in an efficient way and necessary reporting and dashboards, descriptive data analytics, is missing. Control of the activity must be raised using data through the descriptive data analytics means and automation while reducing the need for manual work and possibilities for human errors. There is also a lack of current understanding of the use of data and tools available for managing the data, where clear internal policies around the topic is also missing.

Stage	Challenge	Impact
Data Storage	Data storage location unknown by the staff. No internal agreement, practice, or policy for common data storage practices.	1. Impossible to utilize the data in an efficient way as the staff can't find the data. 2. Loss of staff efficiency due to lost working time.
Data Processing	A manual data processing process.	1. Loss of staff efficiency due to lost working time. 2. Increased risk of human errors.
Data Processing	Lack of data processing knowledge and skills	1. Unutilized system and data. 2. Loss of staff efficiency due to lost working time.
Reporting	Lack of reporting knowledge and skills	1. Missing reports for both internal and external stakeholders. 2. Loss of staff efficiency due to lost working time as the staff use extra time on creating non-relevant reports.
Reporting	Lack of reporting/reports	1. Missing insight to financial performance which has an impact on the decision-making. 2. Customer transparency is missing and well as customer report for supporting the customer decision-making.
Data Sharing	No internal agreement, practice, or policy for common data sharing practices.	1. Loss of data. 2. Loss of staff efficiency due to lost working time. 3. Impact on decision-making as necessary data is not found or inaccessible.

Figure 2. Key challenges and Business Impact.

The PC Factory is currently using data from six (6) different sources, which are listed in *Figure 3*. The data from these sources are spread around the IT environment and used only on ad-hoc basis. The DMP will assist in orchestrating the data from these sources and make the use possible for the organization.

Data Source	Data Format	Alternative Data Format	Data Delivery	Data Storage	Format Approved	Delivery Approved	Storage Approved
DHL	Excel	N/A	E-mail	Network drive	Yes	Yes	Yes
Lenovo	HTML	Excel	E-mail	E-mail	Yes	Yes	No
Tier1Asset	Excel	N/A	E-mail	E-mail	Yes	Yes	No
ERP	Excel	N/A	Manual Export	PC	Yes	No	No
ServiceNow	Excel	N/A	Manual Export	PC	Yes	No	No
LanSweeper	Excel	N/A	Manual Export	PC	Yes	No	No

Figure 3. Data Sources used by PC Factory.

4 Implementation of the Data Management Plan

The implementation of the DMP is divided in six activities that were identified in the research. The six data management activities are adjusted to the PC Factory activity and described in the following section of the document.

4.1 Education

Education of concerned people is considered essential for a working data management. The lack of skills and knowledge in both *Data Processing* and *Reporting* were identified as challenges in the research of the current state of data management within the organization. There is a need to educate concerned people in:

1. Internal data policies and practices
2. Internal data management tools
3. Internal data management responsibilities

The training within the team can be conducted either through a *seminar* with the appropriate trainer or by *reading* internal documentation. The latter poses a risk as it is difficult to control people's reading and therefore the first alternative of ***seminar* is the proposed as the primary mean of education** in this context. The seminar can be conducted either as classroom training or online.

Education preparations are required and the following missing points must be prepared and finalized before the seminar:

1. Formalize the internal *Data Management Policy*
2. Formalize data sources in data cards, incl. data ownership
3. Formalize the *Information Service* -service description
4. Grant access to the data management tool, *PowerBI*

The proposal for the seminar lecturers are:

1. SGTS NB PowerBI Solution Owner
2. PC Factory Team Leader

Additional external experts can be added as lecturers if it is considered needed.

4.2 Data Management Policies and Strategy

The data management policies and strategy defines the basic common rules of how to deal with data. The policy should define how the concerned staff should use data. The *Data Management Policy* require a formalized document which is distributed within the organization and finally reviewed together with concerned staff. The policy is a pre-requisite for *Education*.

The *Data Management Policy* should contain the guidelines for the following:

1. The purpose of data within the organization
    a. What data is being used within the organization?
    b. Why is data used within the organization?
2. Data location guidelines, incl. *Data Card* location
    a. Where is data located and stored?
3. Data processing guidelines
    a. What is data ownership?
    b. What are the responsibilities of a data owner, incl. data documentation and validation?
    c. Who can have access to the data?
    d. How is data accessed and processed?
    e. GDPR compliance statement
4. Data management tool guidelines
    a. What tool is used within the organization for the purpose?
    b. How can the tool be accessed?
5. Data backup & restore guidelines
    a. What data is backed up?
    b. How is the data backed up?
    c. How is the data restored in case of data loss?
6. Data sharing and reporting
    a. What data should be shared with internal and external stakeholders?
    b. How is data shared with concerned stakeholders?

Preparations for the *Data Management Policy* are:

1. Creation of the initial document
2. Review by relevant stakeholders:
    a. SGTS Director – general data management rules
    b. SGTS PowerBI Solution Owner – data location, data processing, data management tool(s)
    c. SGTS Head of Infrastructure – data backup & restore
3. Formalization & publication of the document
    a. Placement in SGTS Universe (Sharepoint)
    b. Publication SGTS Intranet in Yammer

The proposal for the document creator is the *SGTS NB User & Site Support Director*.

4.3 Data Definitions

The *data definitions*, in this context, is the appropriate documentation of the data sources. The *Data Card* template, appendix *A*, should be used for as a documentation template per data source. The templates include all necessary information for proper documentation of a data source.

The data definition ensure that each data source is tagged with an appropriate name, owner and other criteria that is necessary for future management and usage of the data source. The *Data Card* will also ensure that documentation of data sources exist which means that the organization is less dependent on staff knowledge and persons memory.

It is advised that a summary of existing and/or future data sources are identified in a list (example: *Figure 3*) and then a *Data Card* is created for each data source. The storage of the *Data Cards* should follow the guidelines communicated in the *Data Management Policy*.

The *Data Card* should be filled and completed by the identified *Data Owner*. It is suggested that the *PC Factory Team Leader* initiates the kick-off or launch of completing the *Data Cards*.

4.4 Data Management Tools

The data management tool used within SGTS Nordic & Baltic is *Microsoft PowerBI*. This is the only tool that should be used for the data management purpose and is chosen and validated as the standard by the organization management. The data management tool guidelines are communicated in the *Data Management Policy*. The organization staff can get access to PowerBI upon request to the *PowerBI Solution Owner*.

The PowerBI topology with data source types are presented in *Figure 5*. Three types of data sources can be used for producing reports and visualization in PowerBI:

1. An Excel file with columns and rows
2. Data from the SQL database
3. Data through an application integration

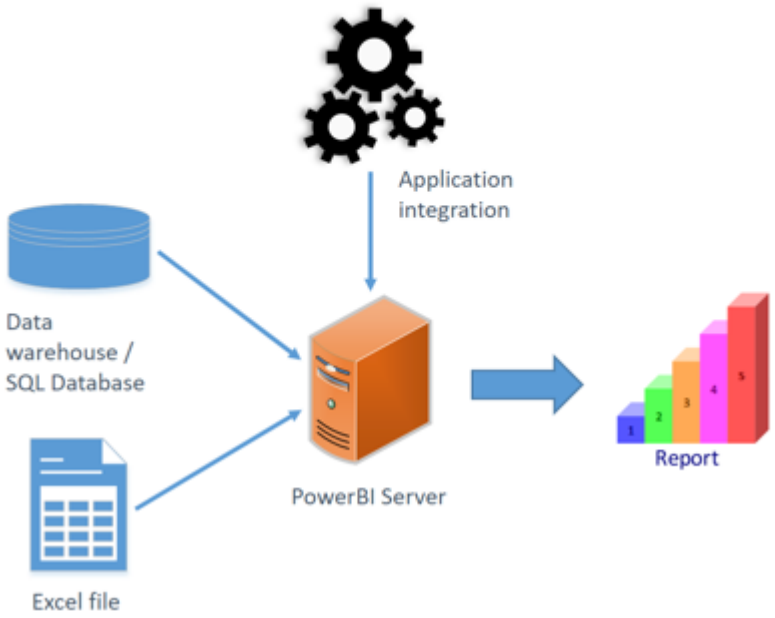


Figure 5. SGTS NB PowerBI Topology

Current existing application integrations are:

1. Lansweeper
2. ServiceNow

Current and new application integrations are managed by the *PowerBI Solution Owner* and each need for a new integration must be addressed by raising a request in the ITSM tool.

The *SQL database* can be used as a data source if the data is first imported to the database. Database imports are managed, as well, by the *PowerBI Solution Owner*.

Excel-files can be used as such as data sources and they are managed by the data owner defined in the *Data Card* for the specific data source. The storage guidelines for the Excel-files are defined in the *Data Management Policy*.

The action needed in this section is related to *education* of concerned staff and raising awareness of the tools used among concerned users.

4.5 System Development Support

The *system development support* in this context refer to the activity where the data owner must ensure that the data quality is good and ensure well-coordinated and -facilitated changes if such are concerning the data or data source. The data owner must be the primary contact for any changes related to any aspect of the data itself or the data source.

It is essential that a *data owner* is identified for each data source, documented in a *Data Card* and that the data owner responsibilities are formalized in the *Data Management Policy*. This will ensure that the data owners are informed about the responsibility during the *education*. Education will also ensure that the concerned people knows about the data management policy and data ownership model implemented within the organization as well as the location of relevant documentation related to the data sources.

4.6 Information Service

The *information service* describes how information and data is shared within the business in an efficient way. The information service and its content, including data and report inventory, must be formalized in a document or system. The document can be an internal service description, that is used for other services within the organization. The system can be an existing system like the ITSM tool Knowledge Base, ServiceNow, in this case.

It is recommended to start with a service description document, as the current ITSM system Knowledge Base isn't used by all concerned stakeholders. The usage of a service description document will require a periodic review and it is recommended that the document is reviewed at least twice per year.

The service description document shall contain at least the following details:

- Data & information sharing tool details
- Data & information sharing tool access details
- Data & data owner inventory
- Report & report owner inventory

It is recommended that the *Information Service* -service description is created and maintained by the *PC Factory Team Leader*. The document doesn't have to contain all information but information references can be used instead.

The *Information Service* -service description shall be completed with the creation of a *Report Card* per report. The *Report Card* will ensure that each report is defined and documented with necessary details including an appropriate report name & creator, significance statement, used data sources and other criteria that is necessary for future management and usage of the report.

Questions that should be addressed to the stakeholders of the reports are:

- Who is going to use the reports?
- What are the report business' needs? (Summarize/ analyze/ monitor)
- When will the report be used? How timely, must the data be?
- Describe success: what is your desired business result?
- What KPIs represents this business result?
- What behaviors or activities will drive desired business success?
- What are the expected intrinsic outcomes?
- What measures represent intrinsic outcomes?
- How would you describe it?

5 Data Maturity Assessment

Data maturity assessment assist the organization with understanding the current data maturity level in addition to identifying the necessary requirements for reaching the following maturity level. The data maturity assessment must be done in the beginning of the development and reassessed during the development journey.

The data maturity model used in this context is the model created by Howson & Duncan at Gartner in 2015, see *Figure 6*. An alternative model was used in the research, in addition to the in one in *Figure 6*, to validate the relevancy of the data maturity

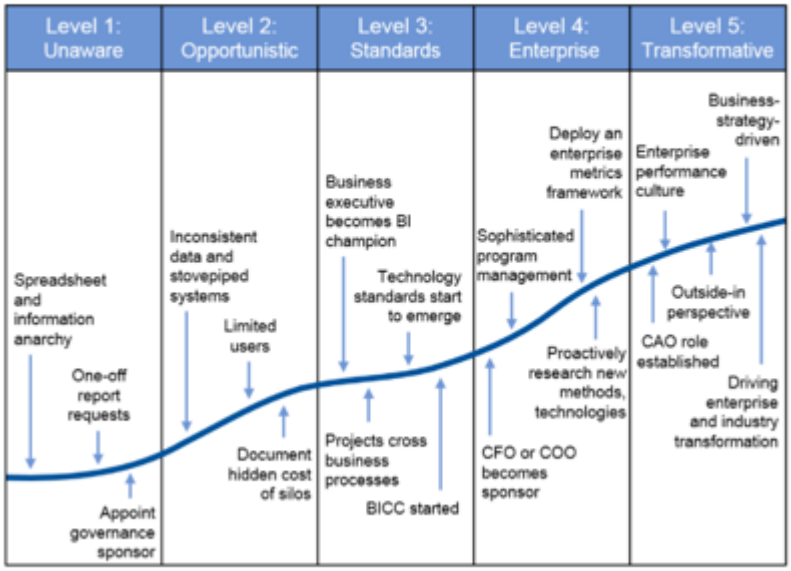


Figure 6. The BI and Analytics Maturity Model by Howson & Duncan (https://www.gartner.com/en/documents/3136418/itscore-overview-for-bi-and-analytics)

It was assessed during the current state analysis that the **current data maturity level is on the lowest level (Unaware)** as awareness and necessary coordinated activities are currently missing.

The descriptions of the first two levels are as following:

Level 1: *Unaware* – The demand for business intelligence and analytics is ad hoc. The decision-making is unstructured and lacking processes and practices. There is no information infrastructure, processes, or performance metrics.

Level 2: *Opportunistic* – Business units manage their own business intelligence and analytics projects. These business units have their own tools and information infrastructure. The business units have dedicated IT organization. The process modeling is usually missing, and the usage of the existing tools isn't efficient. Analytics outcomes are mainly reporting, and dashboards which contain business or domain specific content.

The objective for the organization should be to reach Level 2 – Opportunistic, to reach the next level and be able to develop further to the next level in the future. The requirements for reaching *Level 2 – Opportunistic* are:

1. The case company manage their own BI and analytics projects.
2. Use their own tools and infrastructure.
3. Have a dedicated IT organization.
4. Usage of existing tools is inefficient.
5. Process modeling missing.
6. Analytics outcomes are usually reporting and dashboards.

All these points are covered, and objectives reached by following the *Action Plan*.

6 The Action Plan

The *Action Plan*, appendix 1, is prepared based on the recommendations in chapter 4. The *Action Plan* is based on a project Work Breakdown Structure (WBS) including all necessary elements for an *Action Plan*. The necessary tasks and preparations, including creation of policies and documents, are included in the WBS. The *Action Plan* will not ensure the continuous maintenance and improvement of data management within the organization, which means that the operational part must be validated and agreed internally after the action plan tasks are completed.

It is required that timing and resourcing is evaluated for each action point in the beginning of the project

It is recommended that a *Project Manager* (PM) is assigned to guarantee the completion and coordination of the tasks in the *Action Plan*.

SAINT-GOBAIN

7 Appendixes

7.1 Action Plan

	UPDATED BY:	"Initials"	DATE:	xx/xx/20xx								
#	Phase	Task	Subtask 1	Expected Outcome	Responsible	Due Date	Status	Progress / Comments	Progress %	Priority	Milestone (Yes / No)	Dependencies #
1	1 - Education	Prepare education seminar	Book & confirm lecturer(s)	Confirmation from lecturer	PC Factory Team Leader		Not Started		0	3 - Low	No	-
2	1 - Education	Prepare education seminar	Book & confirm facility/room	Facility confirmation	PC Factory Team Leader		Not Started		0	3 - Low	No	-
3	1 - Education	Prepare education seminar	Create training materials	Set of training materials	PC Factory Team Leader		Not Started		0	3 - Low	No	-
4	1 - Education	Prepare education seminar	Ensure formalization of Data Management Policy	Proof of formalized Data Management Policy	PC Factory Team Leader		Not Started		0	3 - Low	No	18
5	1 - Education	Prepare education seminar	Ensure formalization of Information Service	Proof of formalized Information Service	PC Factory Team Leader		Not Started		0	3 - Low	No	14
6	1 - Education	Prepare education seminar	Ensure user access to ~~Cygate~~	Confirmation from system owner	PC Factory Team Leader		Not Started		0	3 - Low	No	-
7	1 - Education	Hosting seminar	Confirm when seminar is completed	Attendance & completion confirmation	PC Factory Team Leader		Not Started		0	3 - Low	Yes	
8	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about the purpose of data within the organization	Internal data collection & definition of data purpose within the organization	LISS Director		Not Started		0	1 - High	No	-
9	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data location guidelines	Internal data collection & definition of data location guidelines	LISS Director		Not Started		0	1 - High	No	-
10	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data processing guidelines	Internal data collection & definition of data processing guidelines	LISS Director		Not Started		0	1 - High	No	-
11	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data management tool guidelines	Internal data collection & definition of data management tool guidelines	LISS Director		Not Started		0	1 - High	No	-
12	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data backup & restore guidelines	Internal data collection & definition of data backup & restore guidelines	LISS Director		Not Started		0	1 - High	No	-
13	2 - Data Management Policy	Preparation of the Data management Policy	Collect information about data sharing & reporting guidelines	Internal data collection & definition of data sharing & reporting guidelines	LISS Director		Not Started		0	1 - High	No	-
14	2 - Data Management Policy	Creation of the Data management Policy	Creation of the Data Management Policy document	Creation of the Data Management Policy document	LISS Director		Not Started		0	1 - High	No	8 - 13
15	2 - Data Management Policy	Send document for review	-	Send the document to be reviewed according to the reviewer list in the document	LISS Director		Not Started		0	1 - High	No	14

16	2 - Data Management Policy	Update of the Data Management Policy	-	Update the document in case document validation(s) requesting updates	LSS Director		Not Started		0	1 - High	No	15
17	2 - Data Management Policy	Approval of the Data Management Policy	-	Final approval of the document	LSS Director		Not Started		0	1 - High	Yes	16
18	2 - Data Management Policy	Formalize the Data Management Policy	-	Publication & communication of the document within the organization	LSS Director		Not Started		0	1 - High	No	17
19	3 - Data Definitions	Identification of data owners	-	A data owner should be identified for each data source	PC Factory Team Leader		Not Started		0	1 - High	No	-
20	3 - Data Definitions	Creation of a *Data Card* for each data source	-	A data card should be created for each data source and stored according to the data management policy	Data Owners		Not Started		0	1 - High	Yes	19
21	4 - Data management Tools	Preparation of education materials for the seminar	-	Training materials as presentation and/or video for the [redacted] seminar	[redacted] Owner		Not Started		0	2 - Medium	No	-
22	5 - System Development Support	Organization communication about *Data Owners*	-	Proof of organization communication about *Data Owners*	LSS Director		Not Started		0	3 - Low	No	-
23	6 - Information Service	Preparation for the *Information Service* - service description	-	Data collection for: Data & information sharing tool details, Data & information sharing tool access details, Data & report owner inventory, Report & report owner inventory	PC Factory Team Leader		Not Started		0	1 - High	No	-
24	6 - Information Service	Creation of the *Information Service* - service description	-	Creation of the initial document	PC Factory Team Leader		Not Started		0	1 - High	No	23
25	6 - Information Service	Identification of needed reports	-	Creation of an inventory of report needs. From all stakeholders: internal, external (customers, auditors)	PC Factory Team Leader		Not Started		0	1 - High	No	-
26	6 - Information Service	Creation of *Report Cards*	-	Creation and storage of a report card per needed report for definition and documentation purposes	PC Factory Team Leader		Not Started		0	3 - Low	No	25
27	6 - Information Service	Creation & publication of reports in [redacted]	-	Creation and publication of reports and visualizations in [redacted] together with the [redacted] Solution Owner	PC Factory Team Leader		Not Started		0	3 - Low	Yes	26
28	7 - Data Maturity Assessment	Document the initial data maturity assessment	-	Create a Data Maturity Assessment document	LSS Director		Not Started		0	1 - High	No	-

29	7 - Data Maturity Assessment	Re-assess the data maturity	-	Create a Data Maturity Assessment document	USS Director		Not Started		0	2 - Medium	No	28
30	8 - Preparation for the Action Plan	Assignment of a Project Manager	-	Official announcement of a Project manager for the coordination and follow-up of the Action Plan	USS Director		Not Started		0	1 - High	No	-

SAINT-GOBAIN

7.2 Data Card

DATA CARD	
Name or Label	
Synonyms	
Significance Statement	
Data Format	
Data File Format	
Value List	
Validation Criteria	
Valid Operations	
Ownership	
Users	
Source	
Comments	
Creation Date	
Author	
Last Updated Date	

Description	
Name or Label	A unique name or label (unique identifier).
Synonyms	Any synonyms or aliases for the data object.
Significance Statement	A description of why the data is relevant.
Data Format	Used data formats.
Value List	A valid value lists describing what kind of values are expected in the data object
Validation Criteria	A valid validation criterion describing what kind of values are expected in the data object.
Valid Operations	Valid operations describing for which operations the data is used.
Ownership	Data ownership information.
Users	A description of for who the data is for.
Source	A data source description.
Source File Format	Description in which file format the data is populated or generated
Comments	Any relevant comments
Date Created	Date when the data card was initially created
Author	Name of author of the datacard
Last Updated	Date of the last update done in the file.

File Naming Convention	
File Naming Convention	*SGTS NB - Data Card - "Label"*

7.3 Report Card

REPORT CARD	
Name or Label	
Meta data	
Significance Statement	
Data Sources (data card names)	
Published in	
Ownership	
Users	
Comments	
Creation Date	
Author	
Last Updated Date	

Description	
Name or Label	A unique name or label (unique identifier)
Meta data	Descriptive words for the report
Significance Statement	A description of why the report is relevant
Data Sources (data card names)	Names of data sources that are used for populating the report
Published in	Tool where the report is published
Ownership	Report ownership statement
Users	List of users or user groups
Comments	Any relevant comments regarding the report
Creation Date	Date when the report card was initially created
Author	Name of author of the report card
Last Updated Date	Date when the report card was last updated

File Naming Convention	
File Naming Convention	*SGTS NB - Report Card - "Label"*

8 Glossary

Data warehouse = Is a system used for reporting and data analysis and is considered a core component of business intelligence. DWs are central repositories of integrated data from one or more disparate sources. They store current and historical data in one single place.

ITSM = Information Technology Service Management

SGTS NB = Saint-Gobain Technology Services Nordic & Baltic

SGTS Universe = The name of the document and information management tool within the SGTS NB organization. Based on the Microsoft Sharepoint technology.

SQL database = Microsoft's database product which is used within SGTS NB