Expertise
and insight
for the future

Bruno Pasanen

Developing a Supportive Prototype Tool for Resource Planning Management

Metropolia University of Applied Sciences

Bachelor of Engineering

Industrial Management

Bachelor’s Thesis

6 May 2021

Metropolia
University of Applied Sciences

Preface

It is about nine months since I was offered a summer trainee position at Nokia and less than three years since I started the studies of my dreams at Metropolia UAS to become an Engineer. I can say that time is definitely flying. In the beginning of both journeys, I had no idea where I had stepped, and I had hardly any knowledge of the field of business ICT or the operations of the global organization, but now I have learned a lot.

I want to thank the case organization at Nokia very much for this amazing opportunity to be a part of the global organization as a trainee and at the same time pursue the last year of my bachelor's degree during this special time in 2020 and 2021. The right mix of studies and work have given significant support to me. I am really glad and grateful for all the stakeholders, colleagues, meeting participants and contributors of the case company. Special thanks to Aki, Coffy, Sami, Oanh, Marika, Zoe, Mikko, Ove, Anne, Raija and Mervi for all the support and help for my study.

Thank you Rakel Peltola Lic.Phil., my thesis instructor, for the excellent guidance and Sonja Holappa M.A. for steering the thesis process and providing help with reporting language. I also want to thank all the awesome lecturers of the Industrial Management program and fellow students for this valuable and meaningful time of bachelor's studies.

Thank you very much Phuong, Netta, and Iines for supporting my thesis work and commenting on my writing during this project.

Looking forward to future projects,

Bruno Pasanen

Helsinki

6 May 2021

Author Title Number of Pages Date	Bruno Pasanen Developing a Supportive Prototype Tool for Resource Planning Management 67 pages 6 May 2021
Degree	Bachelor of Engineering
Degree Program	Industrial Management
Professional Major	ICT Business
Instructors	Sami Kivilaakso, Transformation Manager M.Sc. Rakel Peltola, Senior Lecturer Lic.Phil. Sonja Holappa, Senior Lecturer M.A.

The objective of the study was to propose a prototype tool which combines personnel resources and operating expenses data, including product and growth data for monitoring and planning purposes. This thesis aimed for a more accurate and simplified business decision-making process and to reduce manual work by developing the tool. Besides this, the thesis did not focus on applying changes to the case company's organization data or deploying the tool.

The research approach of the study is action research, and the study contains four phases. The first phase was the current state analysis for defining the key strengths and weaknesses of the current resource planning data. The second phase covered the literature and best practises in relation to the CSA key strengths and weaknesses, leading to the conceptual framework of this study. The third phase consisted of co-creating the prototype tool, and it was divided into data set preparation and data modelling parts, utilizing the outcomes of the earlier phases. The fourth phase was the validation of the proposed tool by the key users, leading to the outcome of this thesis.

Multiple current resource planning data strengths and weaknesses were identified in this study. One key finding was that the current confidential resource planning data is not mapped to the organization data of the case organization. The current resource planning process also included a large amount of manual work.

The outcome of this study is the final proposed prototype Power BI tool that supports the personnel resources and operating expenses planning. The proposed Power BI tool makes it possible to visualize the big picture of operating expenses in the case organization, and it significantly eases and automates the resource planning data collection. Therefore, less manual work has to be done, and the tool increases the accuracy in the business decision making process. The recommendations for the next steps are also included in this study. Ultimately, the resource planning data quality and usability will be increased after implementing the recommended next steps, also in other organizations of the case company.

Keywords	Power BI, Tool development, Data Mining, Resource Planning

Tekijä Otsikko Sivumäärä Aika	Bruno Pasanen Työkaluprototyypin kehitys organisaation resurssisuunnittelun tueksi 67 sivua 6.5.2021
Tutkinto	Insinööri (AMK)
Tutkinto-ohjelma	Tuotantotalous
Ammatillinen pääaine	Kansainvälinen ICT-Liiketoiminta
Ohjaajat	Sami Kivilaakso, Transformation Manager DI Rakel Peltola, Senior Lecturer FL Sonja Holappa, Senior Lecturer FM

Tämän insinöörityön tavoitteena oli kehittää prototyyppityökalu henkilöresurssi- ja liiketoimintakulun suunnittelua varten. Työkalun tuli koota henkilö- ja liiketoimintakulutiedot monesta eri lähteestä. Lisäksi työkalun tuli sisältää tuote- ja kasvutiedot. Nämä tiedot tuli saada yhdeksi kokonaisuudeksi kyseisen organisaation henkilö- ja liiketoimintakulun seuraamista ja suunnittelua varten. Työn tavoittelema lopputulos oli kehitetyn työkalun avulla vähentää manuaalista työtä ja lisätä tarkkuutta liiketoiminta päätöksenteossa. Työn tarkoitus ei ollut muuttaa kyseisen organisaation tietolähteitä eikä suunnitella työkalun käyttöönottoa.

Tutkimus tehtiin toiminnallisena tutkimuksena, joka koostuu neljästä osasta. Ensimmäinen osa käsitteli nykytilan analyysiä. Nykytilan analyysissä määriteltiin tämänhetkisen resurssisuunnittelutiedon tärkeimmät vahvuudet ja heikkoudet. Toisessa osassa oli teorian opiskelu nykytilan analyysin löydöksien perusteella. Kolmannessa osassa käsiteltiin työkalun kehitystä. Tämä vaihe koostui tietojoukon valmistelusta ja tiedon mallinnuksesta. Viimeisessä osassa käsiteltiin tiedonlähteiden ja työkalun tuottaman tiedon validointia. Viimeinen osa johtaa tämän insinöörityön lopputulokseen.

Useita nykyisen resurssisuunnittelutiedon vahvuuksia ja heikkouksia määritettiin tämän työn aikana. Huomattavin heikkous oli se, että nykyistä resurssisuunnittelutietoa ei pystytä yhdistämään muuhun organisaatiotietoon. Nykyinen resurssisuunnitteluprosessi sisälsi myös suuren määrän manuaalista työtä.

Työn lopputulos on prototyyppi Power BI -työkalu henkilöresurssi- ja liiketoimintakulun suunnitteluun. Työkalu mahdollistaa kokonaiskuvan visualisoinnin ja kerää henkilöresurssi- ja liiketoimintakulutiedot monesta eri lähteestä. Työkalu vähentää huomattavasti määrän manuaalista työtä ja lisää suunnitteluun käytettävän tiedon luotettavuutta. Jatkotoimenpiteet ovat myös osa tätä työtä, ja niiden käyttöönotto parantaa resurssisuunnittelutietoa ja sen käyttämistä mahdollisesti myös muissa kohdeyrityksen organisaatioissa.

Avainsanat	Power BI, Työkalun kehittäminen, Tiedonlouhinta, Resurssisuunnittelu

Contents

List of Figures

List of Abbreviations

FTE	Full-Time Equivalent is used to indicate working capacity of the employee. FTE is used because the actual number of working hours in a global organization varies between the countries.
CA	Competence area is used to describe a technical subarea requiring specific knowledge and skills. CA may span over multiple products and teams. The lifespan is permanent, not just during a specific project.
MF PPM	Micro Focus Program and Product Management service is a service for increasing productivity in R&D organizations. MF PPM supports program and project budget planning, organization capacity and project resource allocation planning, and helps in monitoring and storing the data.
OPEX	Operating expense covers the running expenses to function the business, product or system. OPEX includes costs such as salaries and funds allocated to research and development.

1 Introduction

The analysis and collection of data for resource planning has been playing a significant role in corporations and organizations for a long time. Collecting data has become easy for businesses regardless of size (Stobierski 2019). The collected data might not be suitable for uses such as business decision making or forecasting without further processing. This has created an increasing trend for data science discipline and data scientists to build, clean and organize raw data into something that can be easily understood, accessed and used in business decision making (Stobierski 1 2021).

According to Redman (2013), quality is the main point of managing organization data and information assets. Organization data, such as data for monitoring operating expenses, must be correctly created and maintained as well as easily provided for the users to access, find and understand. Well maintained organization data allows its powerful use and helps the organization to trust the data it maintains. Redman (2013) also underlines the importance of verified and secure data sources in order to prevent data from being used in an improper way. (Redman 2013:1-8.35-36)

This thesis focuses on data mining and building a prototype tool for the case organization’s Resource Planning Management.

1.1 Context

Nokia is a global 5G technology leader. Nokia operates in over 100 countries and by more than 160 nationalities. Nokia creates technology that helps the world act together, such as critical networks, machines, devices as well as sensing and acting in real time on a massive scale. (Nokia 1 2021) Nokia has a history of 155 years, and during its history the company has constantly adapted to the changing world. Nokia transformed its primary focus on telecommunication in 1990 and continues its journey towards a healthier and more sustainable world with technology-enhanced life. (Nokia 2 2021)

This thesis was carried out in Nokia Networks in a global research and development unit employing 2000 people, in a team whose key responsibilities are to deliver the unit’s annual plan and supervise its execution as well as deliver location, resource and budget planning. These key functions are ensuring execution of the strategy and goals of the unit as a part of the global organization of Nokia Networks.

1.2 Challenge, Objective and Outcome

A challenge of this study is that the Resource Planning Managers of the case organization have made use of various tools such as internal business specific tools, manual Excels, and Power BI reports to support their decision-making process for personnel resources and operating expense planning. Having multiple tools to conduct the planning process is very time consuming and exposes the managers to possible errors.

The Resource Planning Managers have given an assignment to learn business aspect, gain organization data understanding, collect and observe initial resource planning data, prepare the data in a form that supports the decision-making process and build a working prototype tool to support the planning process of the operating expenses. This topic was chosen due to the importance of collecting evidence-based data, the need to reduce manual work and the aim for more accurate and simplified business decision-making process with continuously maintained and validated data.

The objective of this study is to propose a prototype tool that combines personnel resources and operating expenses data including product and growth data, for monitoring and planning purposes. The expected outcome of this study is a proposal for a prototype tool that allows the Resource Planning Managers to have one supporting tool for personnel resources and operating expense planning which collects the data from multiple reliable sources. The tool must be easy to use and possible to be updated frequently without changing the current way of storing the organization data. The proposed tool must be able to deliver the data by using visuals and numbers.

1.3 Thesis Outline

This study focuses on building a working prototype tool for supporting personnel resources and operating expense planning process and addressing the business challenge introduced in the previous chapter. Data validation of the proposed prototype tool and further suggestions are included in this study. The research data collection and research design are presented in chapter 2. It was decided to build the tool using Microsoft Power BI desktop, due to the existing knowledge in the case organization’s Resource Planning Managers and the possibilities to create an easily updatable, scalable and user-friendly outcome based on the findings from the current state analysis (CSA) explained in chapter 3. The explored literature and best practices, which are based on the current state analysis findings, is presented in chapter 4. The development of the solution based on the CSA and the literature are described in chapter 5.

This thesis work does not cover the best practices for maintaining Power BI report, publishing report or further development process for the prototype tool. Besides this, the study does not focus on applying changes to or improving the case company’s organization data. The deployment of the prototype tool is not included in this study. The next chapter describes the research design used in this thesis and explains how research data was collected for this study.

2 Method and Material

The business challenge and the outcome of this thesis were introduced in the previous chapter. This chapter describes first the chosen research approach and the research design used in this thesis. This is followed by an account of how research data was collected for this study.

2.1 Research Approach

Kananen (2012) describes a research design as a philosophic broad umbrella of science that contains data collections, analysis of the current state and interpretation practices and methods. The most common division between research approaches is embedded in qualitative and quantitative research. Case and action research are regarded as strategies of research, and they utilize both qualitative and quantitative methods, named as researches with multiple approaches. The main differences between case research and action research is the purpose of the research and the role of the researcher. The purpose of case research is understanding and providing information for the observed issue, whereas action research aims to understand of the problem and deliver a solution. In action research, the researcher participates actively while in case research the researcher is usually an external participant. (Kananen 2012:20-44)

According to Erro-Garcés and Alfaro-Tanco (2020) action research is mainly applied to the operations functions of a company in the field of business and management (Erro-Garcés and Alfaro-Tanco 2020:6). Erro-Garcés and Alfaro-Tanco (2020) describes action research as quoted in their journal:

> Action research is a tool whose implementation ought to be promoted in the business/management field as a way of enhancing relevant, rigorous empirical studies and serving as a framework reference in projects based on research and practice contribution as well as active collaboration between researchers and practitioners. (Erro-Garcés and Alfaro-Tanco, 2020:1)

The main approach of action research is to understand problems and deliver answers. Erro-Garcés and Alfaro-Tanco (2020) describe that being a part of the organization during the study increases authenticity and trustworthiness of the outcomes. A main reason for this is the collaborative manner of action research. (Erro-Garcés and Alfaro-Tanco 2020:1)

Based on a journal by Erro-Garcés and Alfaro-Tanco (2020), action research has shifted from the field of medicine and psychology to the field of business and management in the 1950s. One main ideology of action research is also that the practitioner and researcher are hierarchically on the same level. The authors of the journal mentioned as main features of action research such as researcher interaction, data gathering from multiple sources and cyclical characteristic of study including continuous feedback in all stages. (Erro-Garcés and Alfaro-Tanco 2020:2-4)

The research approach selected for this study is action research. This approach was chosen because of the matching features with objective such as co-creation with stakeholders, cyclical nature, continuous feedback gathering, wide data gathering and the author's active role in the study. Qualitative data modelling approach was also selected for this thesis because of the possibilities that the study includes data modelling and prototype analysis.

2.2 Research Design

This thesis is set out in four phases in order to take actions to achieve the objective. The business problem is the initial motive for this study and each phase leads to an outcome of justified decisions. The research design of this study is shown in the next figure.

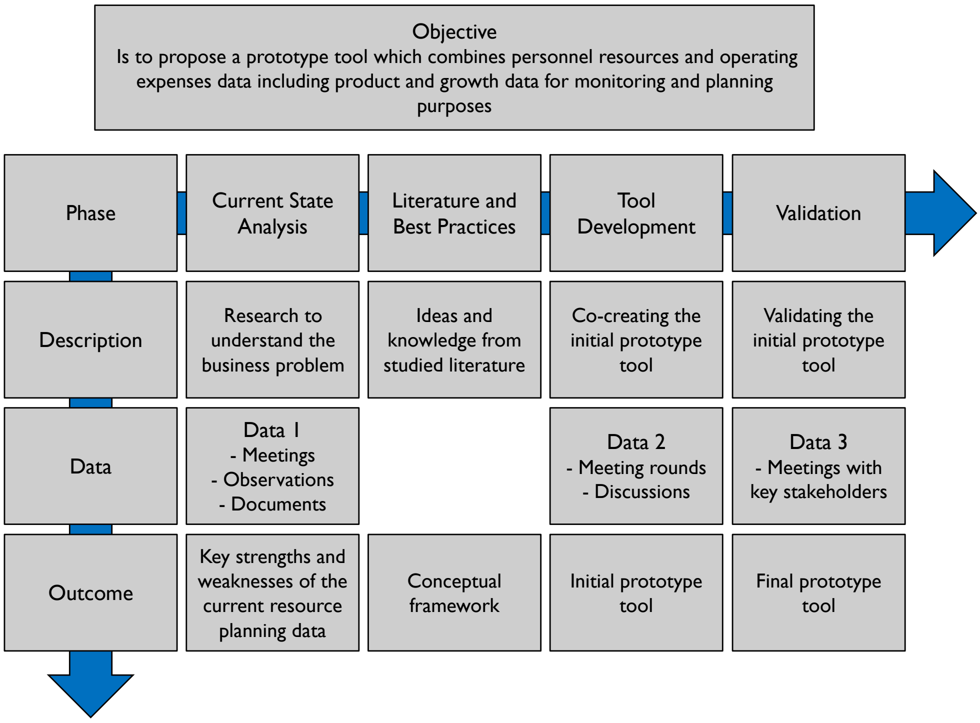

Figure 1. Research Design of This Study

Figure 1 describes the research design of this study. The first phase is current state analysis, which is research of gaining deep understanding of the business problem and understanding. During the current state analysis of this thesis, multiple meetings, discussions were held and participated in. Large amount of current resource planning documents, tools and data were reviewed. The key decision makers of the Resource Planning Management of the case organization participated in the current state analysis. The CSA findings were gathered as a strengths and weaknesses and summarized as a key strengths and weaknesses.

The second phase is the literature and best practice study. The purpose of this phase is to search for relevant literature, based on the CSA key strengths and weaknesses, to achieve the objective. The outcome of this phase is the conceptual framework where the relevant studied literature and the best practices are connected to the CSA key strengths and weaknesses of this study.

The third phase is the tool development of this thesis co-created as a part of the Resource Planning Management team. The tool development of this study is carried out based on the objective, key strengths and weaknesses figured out during the current state analysis and studied literature and best practices. Weekly meeting rounds were organized during this phase and additional meetings were held to build the initial prototype tool of this thesis. The outcome of the third phase is the initial prototype tool.

The fourth phase is the validation of the initial prototype tool and the feedback session. The output data and the data sources of the prototype tool were validated during this phase. The validation was carried out with the key decision makers of the Resource Planning Management team. The outcome of the fourth phase is the validated prototype tool. The recommendations for next steps are also part of this study.

2.3 Data Collection

Research data for this thesis was collected in three rounds from various sources through the case organization. The research data collection steps are data 1, data 2 and data 3. The research data sources were such as internal company documents and tools containing organization data such as full-time equivalent (FTE) numbers, products, costs, project allocations and locations. It is important to notice the difference between the research data collected (e.g. from discussions and observations) for carrying out the thesis and the topic of the thesis, which is organization data related to resource planning (e.g. resource pools, team, location, department, product, FTE and ramp down numbers). Research data were also collected from organized meeting rounds and discussions. Data collection 1 is presented in Table 1.

Table 1. Data Collection 1.

Data Collection 1. for the CSA				
	Source	Type	Topic	Documented
1	Team daily and weekly meetings, other meetings	Discussion and observation	Resource planning related operations of the case organization	Field notes
2	Location Plan	Excel	Observation and test of the current planning tool and data sources	Excel document and field notes
3	Corporate Financial System	Corporate Financial System	Average personnel cost data	Field notes
4	SAP / HR Data	SAP	HR data connection to MF PPM	Field notes
5	Project Capacity Allocation	Excel (export from MF PPM)	Project capacity allocation data observation and identifying attributes	Excel document and field notes
6	Line Manager	Discussions and meetings	Discussions of organization data and the planning process and tool	Excel document and field notes
7	Operation Manager	Discussions and meetings	Discussions of the planning process, SAP HR data and MF PPM tool	Field notes
8	Transformation Manager	Discussions and meetings	Discussion of the planning tools and organization data	Field notes
9	Planning Specialist	Discussions and meetings	Discussion of organization data and Power BI data sources	Field notes and Power BI document

Data collection 1 included exploring internal company documents and tools containing organization data and conducting meetings with the Resource Planning Managers and the Planning Specialist. The purpose of Data collection 1 was to obtain information of the current state of resource planning in the case organization by exploring the tools, business processes and organization data that were currently used. The discussions that were held with the Resource Planning Managers provided knowledge about the current state of tools, organization data, planning and business processes and the strengths and weaknesses of those. Multiple other documents were also observed, and meetings were attended to gain an understanding of the bigger picture. Field notes were written down during all the meetings and observations. The next table presents data collection 2. Data 2 was collected during the tool development phase of this thesis, which is explained in chapter 5.

Table 2. Data Collection 2.

Data Collection 2. for Building the Prototype Tool				
	Source	Type	Topic	Documented
1	Team daily and weekly meetings, other meetings	Discussion and observation	Resource planning related operations of the case organization	Field notes
2	Planning Specialist, the Resource Planning Managers and author of the study	Weekly meeting	Progress reports, current state and project steering	Field notes, Excel and Power BI documents
3	Operations Manager	Discussion and meetings	MF PPM data, Ramp Down data, co-creating Power BI reports,	Field notes, Excel and Power BI documents
4	Planning Specialists	Discussion and meetings	Organization data understanding, technical problem solving and Dataflows sources	Field notes, Power BI documents
5	Power BI Dataflows and Excels	The Power BI data sources	Data sources for the prototype tool	Field notes, Excel and Power BI documents
6	Line Manager (Operations)	Discussion and meetings	Product data and co-creating Power BI reports	Field notes, Excel and Power BI documents
7	Line Manager (Release Planning)	Email inquiries, meetings	Growth ratios and product ratios	Field notes, Excel documents

Table 2 presents data collection 2 for the solution development of this study. This data was collected from weekly meeting rounds, meetings, discussions and email inquiries. The purpose of a meeting with the Planning Specialists and Operations Manager was to gain knowledge of the organization data, (e.g. FTE numbers, Competence Areas and Resource Pools), data set maintenance and data set mapping using Power BI. The current status of the development process, feedback and ideas carried out were revived and validated during weekly meetings. The chosen data sets for the tool development are explained in more detail in chapter 5. The meetings and email inquiries with the Resource Planning Managers played a significant role in collecting and testing reliable data source formats for the prototype tool and gaining knowledge of the operations of the case organization as well as other possible other use cases for the prototype tool. The required attributes for the prototype tool were decided with the Resource Planning Managers, explained in more detail in chapter 5.

The purpose of the last data collection round was to gain feedback and validation from the Resource Planning Managers. Data collection 3 is displayed in table 3.

Table 3. Data Collection 3

Data Collection 3. for Validating the Prototype Tool				
	Source	Type	Topic	Documented
2	Line Manager (Operations)	Discussions and meetings	Validating product and growth numbers and review of the know issue	Field notes, Power BI document
3	Operations Manager	Discussions and meetings	Validating ramp down data and defining the maintenance logic	Field notes, Power BI document
4	Transformation Manager	Discussions and meetings	Validating the big picture and review of the known issue	Field notes, Power BI document
5	Planning Specialist	Discussions and meetings	Internal data validation guidance	Field notes, Power BI document

The last data collection round included validation of the proposed prototype tool, comparing the data with validated data sources, and validating the methods of maintaining the tool and analyzing the known issue. The data collection 3 led to the final proposed prototype tool. The next possible steps were also discussed during the meetings. Data collection 3 was collected from multiple discussions and meetings with the Resource Planning Managers. The next chapter of this study covers the current state analysis of the study.

3 Current State Analysis

This chapter describes the current state analysis (CSA) of planning personnel resources and operating expenses in the case organization. The purpose of the current state analysis was to describe and understand the current planning procedures, tools and utilized resource planning data as well as identify the strengths and weaknesses. The information for the analysis was collected at multiple meetings and discussions with the key decision makers responsible for Resource Planning Management of the case organization. Multiple clarifying questions were asked during the discussions, and field notes were written during each meeting. The data collection for the current state analysis started at the kick-off meeting, at which the first observations of the current personnel resources and operating expenses planning procedures were done and documented with the Resource Planning Managers. After the kick-off, multiple individual meetings were organized, where the currently used tools, resource planning process and data were carefully reviewed. The key strengths and weaknesses were combined, based on the field notes to the summary of CSA.

Data collection 1 is presented in table 1. The next sections of the current state analysis chapter first include a description of the Resource Planning Management of the case organization. Secondly there is the explanation of the current tools used in the resource planning process. The third subsection combines the current resource planning data and the strengths and weaknesses. The last one is a summary of the current state analysis of this study and the key strengths and weaknesses of the current resource planning data, which leads to the next chapter in literature and best practices studied for this thesis.

3.1 Resource Planning Management

This chapter provides an overview of the Resource Planning Management of the case organization. Delivering location, resource and budget plans are strategic activities in the case organization and the main responsibilities of the Resource Planning Managers. The case organization consists of roles such as Operation Managers, Transformation Managers, Operations Specialists, Planning Specialists and Line Managers. The main responsibility of the Resource Planning Management team is to deliver accurate planning outcomes that enable maximized business and development outcomes in accordance with business and technology evolution in the case R&D unit.

Weekly tasks of the Resource Planning Management were observed at daily and weekly meetings and task-specific meetings. Each meeting had its own agenda. The purpose of the daily and weekly meetings was to review and process current work tasks and share information within the team, however, the weekly meetings had a more detailed agenda. The tasks-specific meetings included tasks such as monitoring, planning and managing certain parts of the resource planning process and organization data. Clarifying discussions of the planning process, recruitment and organization data maintenance were held with the key stakeholders to gain more knowledge and understanding of the personnel resources and operating expenses planning.

The personnel resources and operating expenses planning tasks start from the need of having the right amount of employee capacity for a certain product or program release. A sponsor of the organization receives a budget, which is allocated into each part of the organization based on capacity plans and needs. The Resource Planning Managers negotiate with different teams and locations of the case organization about capacity changes between products, competence areas and recruitment needs. The plans also take higher organization plans into account and define the competence of the teams in the decision-making process based on customer needs. In this case, the customers are large telecom operators. The Line Manager mentioned during the CSA that the current responsibility of maintaining the organization data is managed centralized in the Resource Planning Management team.

3.2 Current Tools Used in Resource Planning Process

According to the documents reviewed and discussed with the key decision-makers of the organization, the Resource Planning Managers have used multiple tools in the planning process. The tools used are SAP, Power BI reports, complex custom-made Excel tool, Excel exports from Corporate Financial System, and MF PPM tool on the planning process as well as task-specific user inputs. The majority of these tools are updated daily and maintained by data owners of the organization. The planning process occurs biweekly or more often due to some urgent situation.

Based on internal sources of the case company, The Micro Focus Program and Product Management service (MF PPM) is a service for increasing productivity in R&D organizations. MF PPM supports program and project budget planning, organization capacity and project resource allocation planning, and helps in monitoring and storing the data. Corporate Financial System and SAP are used to fetch resource capacity and cost data to MF PPM and the Excel tool. Sensitive resource planning data is collected manually for planning purposes and processed with data from the tools mentioned above. Power BI reports are used to observe past, current and predicted statuses of multiple different aspects of operations in the organizations.

A major part of organization data maintenance in the case organization is done by Planning Specialists and Operations Manager in MF PPM. MF PPM contains relevant data for this thesis such as work and project resource allocations and FTE numbers. The Planning Specialists are responsible for maintaining Competence Areas and Resource Pools. Corporate Financial System data is maintained by the Financial Department of the case organization and SAP data by the HR unit. The complex Excel tool is used for the personnel resources and operating expenses planning, maintained by Line Managers in close cooperation with the Operations Manager, Planning Specialists, and related teams of the case organizations. Power BI reports, that are used for planning purposes, are maintained locally in teams for more accurate outcomes.

3.3 Current Resource Planning Data

The purpose of analyzing the current resource planning and organization data was to gain understanding and logic and find strengths and weaknesses. The key strengths and weaknesses of the current resource planning data is shown in table 4.

During the current state analysis, multiple internal documents, tools and Power BI reports were reviewed and observed. Most of the research data for the thesis was collected as Excel documents and field notes, during meetings with the key decision makers of the organization. At the first CSA meeting with the Resource Planning Managers, the current process of personnel resources and operating expenses planning was clarified. The manual data collection parts of the process were reviewed with care, and the data in different tools was observed. The process was carried out using a complex and large manual Excel tool called Location Plan. The Excel tool has been known to be outdated but kept in use because the tool includes manually maintained sensitive resource planning data that does not exist in the other tools with non-sensitive organization data. In the current state analysis meeting, one participant mentioned as follows:

> The sensitive resource planning data is completely collected manually to the Excel tool from different sources. (CSA participant)

Manual work causes a great deal of waste work and possible data errors. The biggest reason for maintaining the manual Excel tool was the lack of mappings between the MF PPM data and confidential resource planning data, that were accessed with limits in the case organization. The risk of possible data errors was also a concern of using the current tool. The CSA meetings also revealed that visualizing the data was seen as an issue due to limited capabilities and manual maintenance of the Excel tool. A CSA participant mentioned that being able to see the big picture as a visual is not possible with the current tool. Building visualizations using the Location Plan Excel tool was seen not possible.

The cost data was reviewed as a part of the manual Excel tool. In the current planning process, the cost data was used in the same format as the Excel export from the Corporate Financial System, and the maintenance of cost data was seen to be working. The cost data was provided as half a year time frame and the average of 12 months. The planned growth data was reviewed as a part of the manual Excel tool along with product and team attributes, that were also part of the manual Excel tool. The Line Manager mentioned as follows during the CSA discussion:

> This planned growth and product data per country currently exists only in the Location Plan Excel tool. (Line Manager)

Based on the CSA findings the manual Location Plan Excel tool is one of the main tools in the current personnel resources and operating expenses planning process.

The Project Capacity Allocation data, from MF PPM tool, was reviewed as an Excel export with the planning Specialist and Operations Manager during separate meetings. The Project Capacity Allocation data was a non-sensitive internal data set, containing organization data such as FTE numbers, Competence Areas and Resource Pools. An outcome of the reviews was that only a minor part of the entire Project Capacity Allocation data was needed for the resource planning process. However, the products reviewed on the manual Excel tool did not exist in the Project Capacity Allocation Excel export. The Planning Specialists also pointed out a different possibility to access the Project Capacity Allocation data using Power BI Dataflows source. The additional data source was reviewed after the meeting and field notes were written down. Possibilities of accessing multiple other internal non sensitive data sets by using Power BI Dataflows were also covered during the Dataflows review.

During the CSA meetings, the quality of current resource planning data was seen good, except for the possible risk of data errors due to manual data collection for the Location Plan Excel tool. The Project Capacity Allocation data from MF PPM was seen excellent and large. However, the product attributes were missing from MF PPM, and the lack of addressed growth mapping per location is causing manual work that is done in Location Plan Excel. The lack of existing data set mappings were also seen as a possible quality issue between the sensitive and non-sensitive resource planning data. The Location Plan Excel tool was seen causing a large amount of manual work and containing possible data errors. The purpose of analyzing the data quality was to gain understanding and detect possible issues that might occur later during the tool development phase of this thesis. In the following table, the strengths and weaknesses of the current resource planning data are presented.

Table 4. Strengths and Weaknesses of Current Resource Planning Data

Strengths and Weaknesses of Current Resource Planning Data	
Strengths	Weaknesses
1. Frequently maintained and updated data sets available 2. Resource Planning Managers are competent in using Power BI 3. Some Power BI Dataflows sources exist	1. Data is collected manually from multiple tools 2. Lack of mappings between tools 3. Lack of supporting visuals in personnel and operating expenses planning process 4. Lack of required product attributes and planned growth per location in the MF PPM data

The strengths and weaknesses of current Resource Planning data were analyzed, after all the CSA meetings, with the Resource Planning Managers and Planning Specialists. Data understanding was also deepened by reviewing and observing different data sets and reports, using Power BI Desktop and Excel, as well as participating in multiple different data review, planning, and recruitment meetings. In the next subsection, the current state analysis findings are combined in a summary.

3.4 Summary of CSA Findings

The main responsibility of the Resource Planning Management team is to deliver accurate planning outcomes that enable maximized business and development outcomes, in accordance with business and technology evolution in the case unit. The Resource Planning Managers must have a sharp understanding of the product lifecycle, available resources, open communication to the different organization levels and time to market understanding.

The current state analysis revealed that a large amount of frequently maintained tools are used in the resource planning process. The most important tools are MF PPM, Power BI reports and the Location Plan Excel tool. Sensitive resource planning data is collected and processed manually, using the Excel tool. Possible data errors might exist due to manual collection and input of sensitive resource planning data. Requested product names and planned growth per country are maintained manually in the Location Plan Excel and are not part of MF PPM data that contains non-sensitive organization data. The Location Plan document is maintained centralized by the Resource Planning Management team. Based on the CSA findings, there is not an easy way for the Resource Planning Managers to visualize the big picture with the current tools. This is due to a lack of mappings with manually collected sensitive resource planning data. The main finding from the current resource planning data reviewed is that all the required data sets seem to exist, and they are frequently maintained but used mainly for different purposes. The Resource Planning Managers are also competent in using Power BI reports, and the organization already has some Power BI Dataflows data sources available from the MF PPM. The next table presents the key strengths and weaknesses of the current resource planning data.

Table 5. Key Strengths and Weaknesses of the Current Resource Planning Data

Key Strengths and Weaknesses of the Current Resource Planning Data
Data is collected manually from multiple tools.
Lack of mappings between tools.
Lack of required product attributes and planned growth per location in the MF PPM data
Lack of supporting visuals in personnel and operating expenses planning process.
Resource planning management is competent in using Power BI reports.
Power BI dataflows data sources available

Table 5 presents the key strengths and weaknesses of the current resource planning data in the case company. Red represents the weaknesses and green shows the strengths.

It has been decided that the prototype tool will be developed using Power BI Desktop, due to high competence of using Microsoft Power BI reports and the frequently maintained Power BI Dataflows data sets in the organization. The next chapter combines the literature and best practices studied for this thesis based on the current state analysis.

4 Literature and Best Practices

This chapter gathers the relevant knowledge and best practices studied for this thesis. Some theoretical background information was gained from the experts inside the case company, such as internal tools, organization data understanding and business understanding. Best practices and literature were studied based on the key findings of the current state analysis presented in the previous chapter. First, this chapter introduces the Cross-Industry Standard Process for Data Mining (CRISP-DM). Secondly, it describes the theory of studied tools utilized in tool development. Lastly, it presents the summary, the conceptual framework, which includes the most important knowledge to be used when building the prototype tool for the case company in chapter 5.

4.1 CRISP-DM

Deluge of data is a result of advanced technology and modern life (Aggarwal 2015:1-3). Data mining is a field of study of collecting, processing, cleaning, analyzing and obtaining useful insights from different kinds of data. Data mining is a large term that includes different aspects of data processing (Aggarwal 2015:1-3). CRISP-DM, the Cross-Industry Standard Process for Data Mining, was designed independently of any specific tool or kind of data (North 2012:5). Since the release of the process, CRISP-DM has become the most commonly used method for data mining projects (Saltz 2020). The purpose of the process is to convert raw data into an usable format.

CRISP-DM consists of six phases that naturally describe the process of data mining: business understanding, data understanding, data preparation, modelling, evaluation and deployment. These phases are visualized in figure 2.

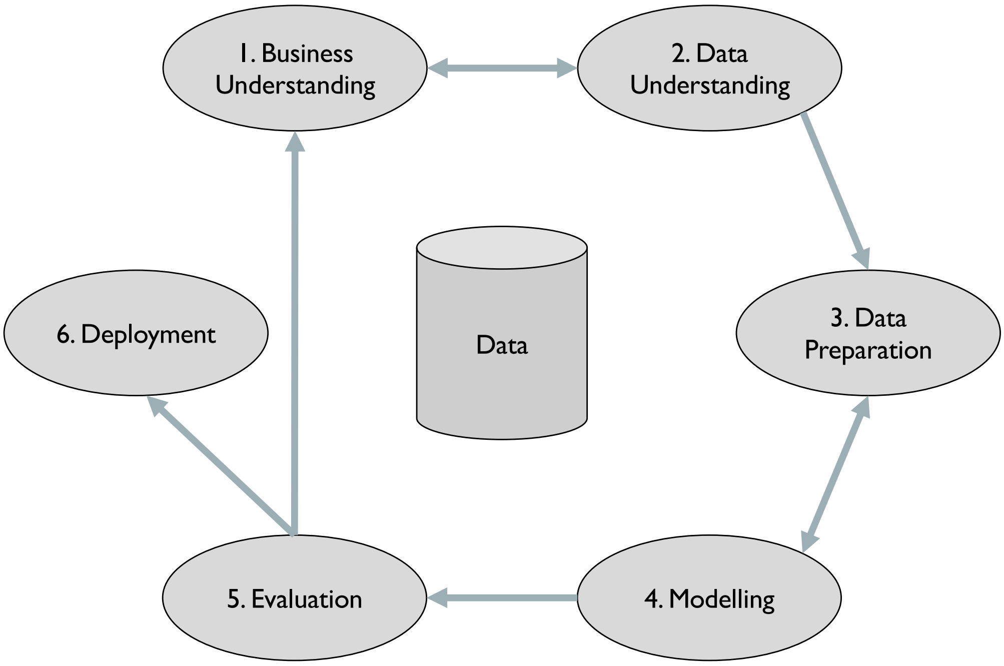

Figure 2. CRISP-DM Model

Additionally, the phases of CRISP-DM can be cycled through adaptively rather than following the process directly. The knowledge learned from previous phases supply and support the following phases. Tasks included in each phase of CRISP-DM can be adjusted based on the nature of business and tools used in the process. However, skipping phases of the process would not make sense because there would not be a model or a goal to evaluate. As a result of using CRISP-DM, proceeding to the next phase can be always reflected to the current phase. (Chapman et al., 1999:6-8)

The first phase, business understanding, defines a main need that the organization expects to discover and gain from a data mining project. Being an active part of the organization eases the understanding of business and makes the business reasons and aims for data mining as clear as possible. Documenting the needs, finding available resources, defining goals and success criteria with key decision makers, supports the next phases of a data mining process. (IBM 2012:3-9)

Data understanding is the second phase. In this phase, business goals are turned into data mining goals, and data is explored by using tables and graphs. Data mining goals include methods for model assessment, such as accuracy, needed information and type of data mining problem such as data clustering, manipulation or prediction. Actual numbers for desired data are needed in this phase, as well as producing the desired outcome later in the modelling phase. Defining the success criteria in technical terms is as important as defining success criteria in business terms, to be able to validate the results in the later phases. The data and business understanding with planned goals formulate the foundation of the road map for the data mining project. (IBM 2012:11-15) The discovery of available and possible data sources for analysis is also a part of the first two phases of the CRISP-DM process. The knowledge of strengths and weaknesses of possible data sets helps in proceeding to the data preparation phase. (Chapman et al., 1999:15-19)

Data Preparation is the third phase, which is usually the most time-consuming part of data mining. The contents of the data preparation phase usually vary quite a lot depending on the business case and data mining goals. The preparation phase typically includes selecting, cleaning, constructing, merging, appending, sorting, replacing, manipulating and splitting data, commonly called as normalizing data. Preparation can be divided into even smaller parts, such as including or excluding certain rows or attributes. The outcome of the data preparation phase is a data set that can be used for modelling. (IBM 2012:17-21)

Modelling is the fourth phase after the data has been prepared for the use case. The models can be divided into two main categories; models that predict and models that classify (North 2012:9). The modelling phase is the first time when the success of data preparation and data mining goal will be truly tested. It is important to consider how the results of the model will be tested and validated. The models can be compared with data that is already validated. As previously said, only one test model might not be enough for drawing a clear picture and building trust in the model or the modelling results might not be as expected. Documenting the steps and data used for each model is important for being able to retrace the steps back to the data preparation phase when necessary. When the results of the model are satisfying and they are reaching the data mining goal, it is time to proceed to the next phase. (IBM 2012:23-27)

Evaluation is the main part of testing the effectiveness of the data mining results. The models chosen in this phase have already been tested and validated regarding the data mining goal. The resulting model is tested in relation to business needs. However, the evaluation phase however requires a clear understanding of the business of the organization. It is strongly recommended to involve the key decision makers in the evaluation phase, for determining if some important factors are missing or unclear. The approved models are those, which satisfy the data mining and business goals of the organization. (Chapman et al. 1999:32-35)

Deployment is the implementation or wrap-up process of the validated models, using new insights to make improvements within the organization (IBM 2012:33). The deployment phase also includes reporting, such as a step-by-step process for models, findings, known problems and technical details to be able to carry out the data mining results into use (Chapman et al., 1999:36-40). However, different organizations or industries want different forms of deployments based on business needs. Possible deployment targets are wide. Perhaps, the deployment can be perhaps continuous analysis of incident patterns, use of a developed tool or an implementation plan and a report of the data mining project. (IBM 2012:34-36).

4.2 Tools

This subsection covers the tools that were studied based on the current state analysis key findings. Microsoft Power BI is introduced first, mainly focusing on Power BI Desktop, followed by a brief description of Microsoft Excel.

4.2.1 Microsoft Power BI

Microsoft Power BI is a collection of business analytics tools and connectors that all work seamlessly together. The purpose of Power BI tools are to help users to turn their unrelated data sources into a consistent format. Power BI allows users to connect, visualize, share and discover data. The most commonly known Power BI tools are Power BI service, Power BI Mobile and Power BI Desktop. (Microsoft 1 2021) Power BI Service is a cloud-based tool, mainly used to view, share and control dashboards. The dashboards has been mainly created and published using Power BI Desktop, which is a data visualization tool only for Windows users. Power BI Mobile is a multiplatform mobile application that allows the users to view and interact with reports. (Microsoft 8 2021)

4.2.2 Power BI Desktop

Power BI Desktop is a tool for data analytics and visualization, with a wide range of possible data sources and output visualizations (Stobierski 2 2021). Power BI contains many options for deep diving into data sources and for observing and transforming data. The layout of Power BI Desktop is the same as Microsoft office tools. The following figure displays Power BI Desktop view.

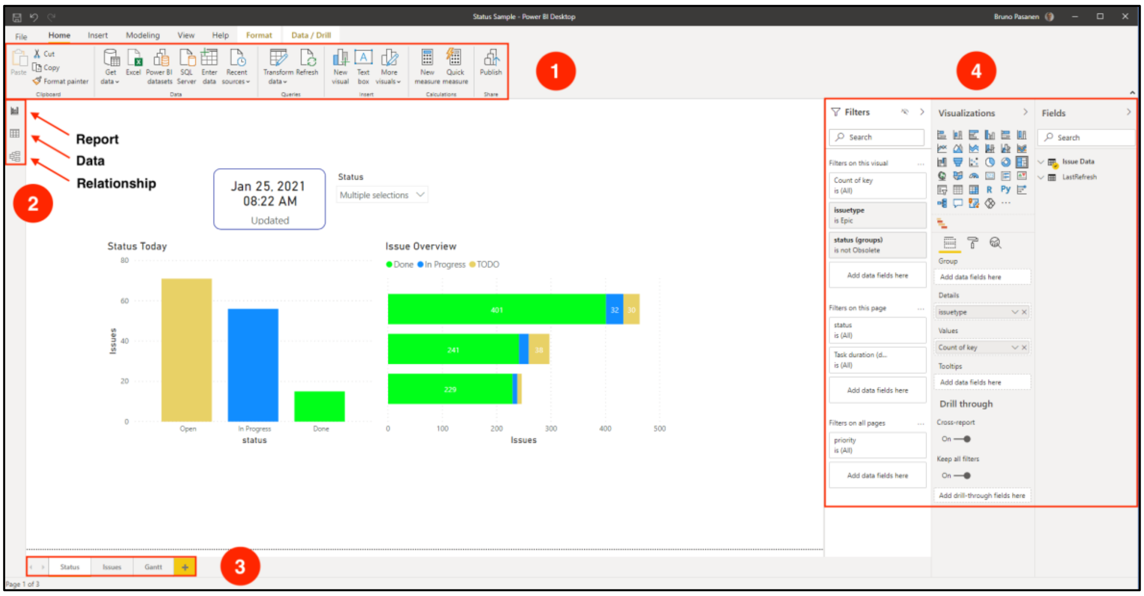

Figure 3. Power BI Desktop View

Figure 3 presents Power BI Desktop view with sample visualizations in the middle. The most commonly used areas are circled in red boxes. Power BI ribbon (1) interacts in relation to the chosen tab above the ribbon. All the tools like sharing, transforming and connecting data can be found from the ribbon. Report, Data and Relationship views (2) change the view of the Power BI Desktop. Currently Report view is chosen in Figure 3. Report view is for building the data visualization and transforming data. Data view is mostly for viewing data and relationship view is for adding relationships between data tables. Pages (3) are used to separate the visualizations, for example, per topic. Filters, Visualizations and Fields (4) are major functionalities for constructing the Power BI report. Fields contain the data sets that are brought into the report. Visualizations are used to choose how the data is presented, for example, a line chart is used for displaying change over time, and a percentual stacked column chart is used for visualizing comparison. Filters are used for filtering the chosen data attributes. Mostly, filters are used to include or exclude chosen data. (Microsoft 2 2020) Power BI also contains slicers as a filter on report, which can be chosen from the visualizations field.

Relationships are for creating related tables. A common use case for relationships would be when the first table contains attributes of sales data. The second table contains department per location data. A unique key exists on both tables when relationship can be created as a relational database. Two major settings apply to the relationships in Power BI, cardinality and cross filtering direction. Cardinality defines the type of the relationship. Cross filtering does the filtering between related tables apply on one or both directions between the tables.

Power BI Desktop has a built-in data query editor called Power Query Editor, in which data transformations can be done, such as renaming columns, unpivoting columns, combining data and constructing data using formulas. Power Query is a data connection and transformation tool, a part of Microsoft tool family designed for Power BI and Excel. (Microsoft 3 2021) Power Query Editor can be accessed by Transform Data button on the ribbon (1). The following figure presents the view from Power Query Editor

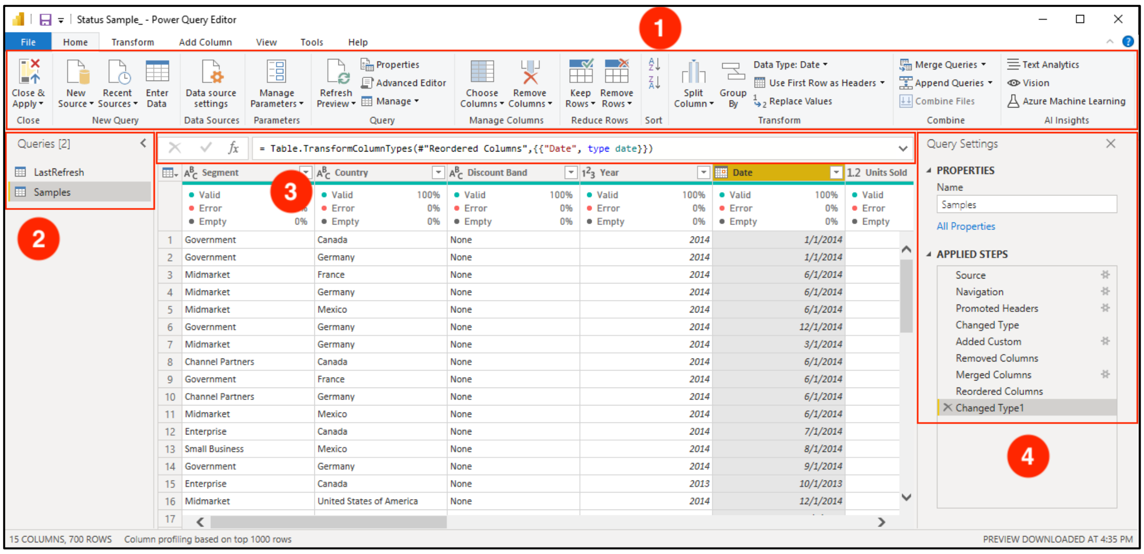

Figure 4. Power Query Editor View

In figure 4, ribbon (1) contains the tool collection, which is interactive based on the chosen tab. Queries (2) are the data sets brought into the report. M Language formula bar (3) currently displays the change of data type of Date column. The formula bar can be used to write and edit formulas. Applies Steps (4) list contains each edit done to the chosen query (2). Every time a transformed query connects to the data source, applied steps are carried out (Microsoft 3 2021).

In Power Query Editor, queries can be combined using merge or append (Microsoft 2019). The logic between these two is that merge is used to combine two or more matching columns into an existing table or as a new table. For each merge, the type of join is also chosen. Append is used for combining two or more tables into a single table. This way the new rows are added when matching columns exist. In both cases, matching attributes are the key to conduct the merge or append in Power Query Editor. Merge and append tools can be found on the tool bar under combine in figure 4. (Microsoft 3 2021)

Power Query formula language is used in the background of all kinds of data mashups in Microsoft Power BI and Excel. The language is more commonly known as M language. The basic concepts of the M language include, for example, values, expressions, variables, and identifiers expressed in case sensitive and lexical structure. Functions and special values provide the foundation of the standardized library of M. (Microsoft 4

2020) The M code of the chosen query can be accessed behind the Advanced Editor button (1), visible in figure 4.

Microsoft Power BI Desktop allows users to get data from multiple sources. The sources can be, for example, local CSV files, Rest API connections, SQL servers and many others. Different data source options are also supported, including DirectQuery or on-premises data gateway options (Microsoft 5 2020). The common data sources of Power BI Desktop are presented in the following figure.

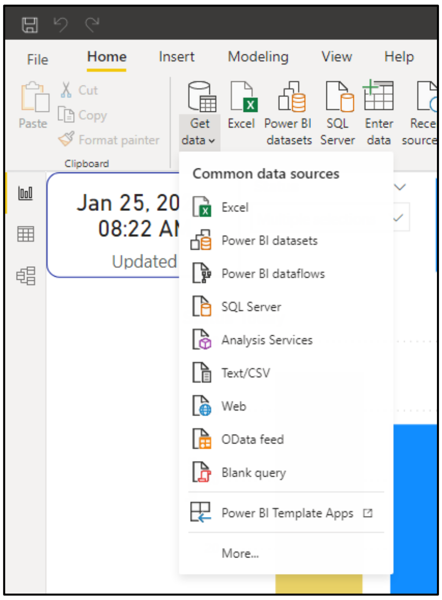

Figure 5. Power BI Desktop Common Data Sources

Figure 5 presents the common data sources used in Power BI Desktop. Power BI Dataflows is an enterprise-focused cloud solution for data preparation and access for data consumers as well as for further processing (Microsoft 6 2020). Dataflows can be used with multiple Microsoft products such as Excel and Power BI. Dataflow uses Microsoft Power Query to conduct transforms to data, the same way as it is done locally in Power BI Desktop. The main benefits of using Power BI Dataflows as an organizational data source are centralized data transformation code and centralized query maintenance (Microsoft 7 2020).

Even though data visualization tools, such as Microsoft Power BI, are cutting edge, the visualizations are only as good as the data used to model those (Stobierski 2 2021).

Therefore, it is importance to gain knowledge of the business and data before the start of data visualization or transform using tools such as Power BI Desktop.

4.2.3 Microsoft Excel

Microsoft Excel is a widely used spreadsheet software, part of the Microsoft Office 365 product family. Excel stands on the highest end of tools for multipurpose calculations, formulations and data organizing on spreadsheets. However, Excel is not a data visualization tool, even though Excel has multiple types of common charts and histograms to visualize data. (Stobierski 2 2021) Excel allows its users to share spreadsheets and access documents easily through Microsoft tools online and locally. The biggest difference from the user point of view, between Excel and Power BI Desktop, is that the big data processing power in Power BI and formulas in Power BI can only be applied to the column. In Excel you can insert different formulas on each shell. This makes a significant difference between these tools when visualizing and transforming large amounts of data.

4.3 Conceptual Framework

Conceptual framework of this study combines the most significant literature findings connected to the findings of the current state analysis of this study. The idea is to gain knowledge, best practices and understanding from literature, to tackle the weaknesses identified in the current state analysis (CSA), and for carrying out this project out successfully. The key findings of the current state analysis and the literature are presented in the following table as a visualization of the conceptual framework of this study.

Table 6. Conceptual Framework

Conceptual Framework	
Key Strengths and Weakness of the Current Resource Planning Data	Literature
Data is collected manually from multiple tools.	• **Data Mining theory** (Aggarwal 2015) • **Utilizing different Power BI data sources to collect data from multiple tools** (Microsoft 5 2020)
Lack of mappings between tools.	• **CRISP-DM, Standardized framework to understand, prepare and model the data** (North 2012), (Chapman et al., 1999), (IBM 2012), (Saltz 2020) • **Modifying, constructing and normalizing data using Power BI Desktop** (Microsoft 3 2019) • **Utilizing Power Query M Language to combine data** (Microsoft 4 2020)
Lack of required product attributes and planned growth per location in the MF PPM data	
Lack of supporting visuals in personnel and operating expenses planning process.	• **Benefits of data-driven decision-making in an organization** (Stobierski 2019), (Stobierski 1 2021) • **Components of the successful data management** (Redman 2008) • **Co-creating visualizations using Power BI Desktop** (Microsoft 2 2021)
Resource planning management is competent in using Power BI reports.	• **Top data visualization tools for business professionals** (Stobierski 2 2021) • **What is Power BI Desktop** (Microsoft 1 2021), (Microsoft 2 2021)
Power BI dataflows data sources available	• **Utilizing Power BI Dataflows automate collecting required data** (Microsoft 5 2020), (Microsoft 6 2020), (Microsoft 7 2020)

The conceptual framework of this study is presented in table 6. Literature and best practices studied are presented on the right side, according to the current state analysis key strength and weaknesses that are presented on the left. In next chapter, the conceptual framework is utilized for developing the prototype tool as a form of data set preparation and data modelling, and thus for solving the business problem in the case organization.

5 Tool Development

This chapter utilizes the current state analysis findings, the studied literature, and best practices learned in previous chapters, to build the initial proposal of the prototype tool. The data mining goal was set according to the CRISP-DM framework, to be based on the gained business and data understanding for this thesis. The data mining goal is to construct and integrate the data set, for being able to build models to support the personnel resources and operating expenses planning. The objective of this study was stated in chapter 1.

The tool development chapter contains four subchapters. The first subchapter is an overview of the tool development. The second and the third subchapters are the main chapters of the tool development, divided into two categories, data set preparation and data modelling, which lead to the fourth subchapter of the initial proposed prototype tool.

5.1 Overview of the Tool Development

The initial prototype tool was developed in close collaboration with the Resource Planning Management of the case organization. Weekly meetings were held through the tool development. Multiple small discussions and workshops were also held. The purpose of the weekly meetings was to review the progress, gain feedback and steer the project when necessary. Other clarifying discussions were held to validate the decisions and solve problems mostly with the Planning Specialist in smaller workshops.

In the beginning of the tool development, the data sources were reviewed, discussed, and brainstormed in the weekly meeting. Ideas were gathered from the participants' and the author's discovered ideas were presented. Different possible organization data sets were tested, leading to the decision. After the decision of the data sets for the prototype tool, the transformation of the data sets began. Confidential resource planning data sets were integrated and normalized to the open internal organization data sets, in order to transform the internal data set for supporting the personnel resources and operating expenses planning process.

After the data set preparation was done, the prototype tool development continued with data modelling part. The Power BI reports were co-created based on the provided research data 2 by the Resource Planning Managers and by using the transformed and integrated data set in chapter 5.2. Feedback was received throughout the tool development process in the weekly meetings. The final initial prototype tool to support the personnel resources and operating expenses planning was co-created using all the field notes written down during the tool development of this thesis as Power BI reports. The initial prototype tool is presented in subchapter 5.3.2.

The following terms are used multiple times during the tool development of this thesis. Resource pool describes a group of employee resources that belongs to a certain project, including specific technical knowledge and skills. The pool can be a team or part of a bigger organization. The lifespan is usually the lifespan of a project which represents a certain program. Competence area (CA) is used to describe a technical subarea requiring specific knowledge and skills. CA may span over multiple products and teams. The lifespan is permanent, not just during a specific project. Full-Time Equivalent (FTE) is used to indicate the working capacity of an employee. The FTE number is used because the actual number of working hours in a global organization varies between the countries.

5.2 Data Set Preparation

Data preparation phases are always unique, according to the studied literature in chapter 4.1. In this study, the data preparation is called data set preparation, and it includes parts such as selecting fit to purpose data sets, integrating data, cleaning data and constructing new attributes. The following five subchapters describe the data set preparation part of the tool development of this study and the research data collection 2, leading to the data modelling part of the tool development.

5.2.1 Choosing the Data Sets

Based on the studied literature in chapter 4 and the key CSA strengths and weaknesses, the data set quality has been maintained high by proposing to use only frequently maintained and validated data sets as well as removing possible causes for data errors. During the CSA, the possibility to replace the Excel export data sets by using the Power BI Dataflows data sets was discovered. All the data sets were chosen in order to tackle the CSA key weaknesses and the criteria defined in the thesis outline chapter, not to apply changes to the organization data structure of the case organization.

Three of the chosen confidential resource planning data sets for the tool development remained in Excel format because of confidentiality. The Excel data sets contained important data for this thesis quality wise, but they were small in terms of attributes. The first Excel data set was the export from the Corporate Financial System of the case organization, the same one as reviewed on the Location Plan Excel during the CSA. It was also decided to use the other two manual Excel data sets in the same format as in chapter 2 on the Location Plan Excel. The chosen Excel data sets contained growth and product data on a simple Excel matrix. The product and growth data were provided in a ratio format between 0 to 1.

In the weekly review meeting the need for including the Ramp Down data was discussed, and it was decided to be included in the project. The purpose of the ramp down data set was to bring in the data of locations where the resource headcount will be decreased. The structure of the ramp down data was different compared to the growth and product data sets, and it was reviewed with the Operations Manager at a separate meeting. The Ramp Down data was provided in Excel containing minus personnel numbers per department, including team and date attributes. Adding the Ramp Down data set to the other chosen data sets amounts to four confidential Excel data sources.

Multiple internal Power BI Dataflows data sets were explored, and knowledge was gained from the Planning Specialists. Three frequently maintained internal Power BI Dataflows data sets that use the data form MF PPM were found out and chosen because of the benefits of using validated data sources and centralized maintained Dataflows studied in chapter 4. The first Power BI data flows data set contained the Project Capacity Allocation data set, the same as the Excel export from MF PPM reviewed during the CSA. The second Dataflows data set contained team and unit names. The third chosen Power BI Dataflows data set was the mapping data set, containing a set of two unique id numbers used in the first and second data sets.

In terms of size, the Project Capacity Allocation data set was the largest and the major data set for this study, including 30 attributes and about 800 000 rows in total. All the other data sets included less than 5 attributes per data set. The next list shows the proposed seven data sets for the data set preparation step of the tool development. The data sources are marked in parenthesis.

- Cost data, (Corporate Financial System Excel export)
- Growth data, (Location Plan Excel)
- Product data, (Location Plan Excel)
- Ramp down data, (Ramp Down Excel)
- Organization unit and team data, (Dataflows)
- Competence Area / Resource Pool Mapping data, (Dataflows)
- Project Capacity Allocation per org level/personnel/location data, (Dataflows)

The list above presents the chosen data sets for the prototype tool development. All the chosen data sets were reviewed and documented in field notes for possible data errors, quality issues and chosen key attributes.

The Excel data sets included attributes such as information of personnel cost per month, locations for growth and ramp down, growth ratios and product ratios. All of the attributes in Excels were marked as the key attributes for this project. One major data error in one Excel data set was discovered. The cost data was using different data type with a date attribute compared to the other data sets, because of the sensitivity level of cost per personnel data. Two major quality issues were detected with the Excel data sets. Manual maintenance of the Excel data sets create a possibility to input wrong data and manual maintenance work. The impact of this quality issue was reduced by using the same Excel formats as before and making users aware of the possible errors and maintenance. The second quality issue was the lack of existing unique attributes to map the data into Power BI Dataflows data sets.

Possible issues in the Dataflows data sets were also detected. The product names did not exist in the Dataflows data in the same way as the Resource Planning Managers had explained during the CSA. This issue was reviewed during the weekly meetings with the Resource Planning Managers. The outcome was that the currently used product mapping contains only R&D work type, but also management work type was seen as necessary to be included to the products. Field notes of the possible need for constructing a new product column were written down during the meeting. The second issue was minor; month and year attributes were in separate columns in the Dataflows data set. This was considered to be easily fixable by harmonizing the attributes to match with the chosen Excel data sets.

The key attributes of the Dataflows data sets were listed as attributes related to the Full-Time Equivalent (FTE) numbers, location, department, work type, programs and two unique identifiers columns (CA and Resource Pool IDs). Major quality issues were not discovered in the Dataflows data sets. However, two possible issues were detected due to the chosen attributes of product and department columns. Mapping the growth and products manually might cause issues when data changes occur in MF PPM. These two possible maintenance issues were marked down to the field notes.

5.2.2 Overview of the Data Sets Integration

This subchapter focuses on describing the methods of the chosen Excel and Dataflows data sets, for the integration to the Project Capacity Allocation data set. The manipulation of Project Capacity Allocation was needed because the confidential personnel resources and operating expenses planning data were not a part of it. The Excel data sets include data such as addressing the changes into the FTE numbers by increasing or decreasing and addressing the growth per location and products, based on the input of the Resource Planning Managers on the confidential Excel sources. Conducting this kind of changes into the data sets by using related tables would not be possible. The following table presents the planned data set integration order.

Table 7. Data Set Integration Order

Data Set Integration Order			
Data Set	Data Source	Data Set Description	Purpose
1	Power BI Dataflows (MF PPM)	Project Capacity Allocation data per org level/personnel/location (Numeric, Text, Date, True/False, null)	Personnel, product, project and work type data
2	Excel (Location Plan)	Growth data (Numeric, Text)	Address growth in different locations
3	Excel (Ramp Down)	Ramp Down data (Numeric, Text, Date)	Bring in Ramp down data per locations
4	Excel (Location Plan)	Product data (Numeric, Text)	Address growth in different products per location
5	Power BI Dataflows (MF PPM)	Competence Area / Resource Pool Mapping data (Numeric, Text)	Mapping data set
6	Excel (Corporate Finance System)	Cost data (Numeric, Text, Date)	Bring in average personnel cost per location
7	Power BI Dataflows (MF PPM)	Organization Unit and Team data (Numeric, Text, Date, True/False, null)	Bring in unit and team names

Table 7 presents the estimated data set integration order. The project capacity allocation data set (1) is the main data set, in which the integrations and changes will be addressed.

The growth data (2) was decided to be integrated first. Ramp down data (3) was planned to be brought in next, because there was also a need to be able to decrease the FTE numbers. The product data (4) was logical to integrate after the earlier mentioned data sets, because the purpose was to be able to connect growth and ramp down into certain product. The mapping data (5), cost data (6) and Organization unit and team data (7) were the last ones brought in, because those construct the possibility to filter and import attributes to the Project Capacity Allocation data (1).

The following three subchapters deep dive into the data set integration preparation, data set integration process and transformation of the combined data set. The subchapters are explained in a chronological order of the steps applied in Power Query Editor and the issues faced.

5.2.3 Data Set Integration Preparation

This subchapter covers the integration preparation of the chosen data sets presented in the previous subchapter. Each data set integration preparation is described, issues are defined and solved. The preparation was mostly done by using Power BI Power Query Editor. However, the Excel export from the Corporate Financial System was transformed in Excel because of the different data format compared to the rest of the data sets.

The first data set prepared for the integration was the growth per location data. The data was provided in an Excel matrix, including resource pool names, resource pool ids and locations. The growth was addressed between resource pool ids and countries using percentual ratios. The document was kept exactly in the same format as the Resource Planning Managers have used it before, to ensure smooth maintenance. The preparation steps were applied in Power Query Editor. All the columns of the spreadsheet matrix were unpivoted as an outcome of resource pool name and id columns and location and ratio columns. The purpose of the ratio column was to be able to modify FTE numbers matching with location and resource pool.

Next, the ramp down data set was brought into the integration preparation process. The data was presented on an Excel matrix with resource pool name, unit and department columns, including dates and minus numbers mapped to the columns. Data transformation in Power Query Editor was started with unpivoting the matrix. Then custom columns were added, to be able to connect the ramp down data into an existing resource pool and other columns on the data set 1. Also column to detect ramp down numbers were added. The further steps included combining the mapping data set in order to get resource pool id numbers in the table. Merging the mapping table and custom columns were done in order to be able to construct a similar data set as in the Project Capacity Allocation for the appended query studied in chapter 4. Appending was used because the new columns were needed on data set 1, as well as adding ramp down values on the existing FTE column in the data set 1.

The growth per product mapping table was prepared for integration after the second manipulation of FTE numbers on the personnel capacity data. The product growth mapping

Excel used the same logic as the growth per location mapping. The product mapping Excel included matrix of competence area (CA) name and CA id columns, product row and percentual values to address the growth between CA id and product. The Excel table required only unpivoting columns in Power Query Editor. The outcome was four columns in the same way as growth per location mapping. However, earlier growth per location mapping was done by using Resource Pool id and the current product growth mapping was done using Competence Area id. The purpose of percentual values was to add changes to the current FTE numbers including growth per location data.

The mapping data set contained only CA id and Resource Pool id columns. The mapping data was required to combine the product growth data sets into the Project Capacity Allocation data set, because the later mentioned data set did not contain CA id. The mapping data set did not require any preparation because the data was already designed for this purpose.

The Excel export, containing cost data, provided the average cost of the employee resources per location in a simple matrix by country, half year time frame and the average of 12 months. The time frame included a three-year period. The cost data was transformed by using simple dividing calculation to change the annual average personnel cost into months in Excel. The calculation was dividing the cost by 12 into each cell representing one month. The outcome format was columns, including date, location, and cost columns. This transformation was defined as useful because the cost data would be updated by importing the same export, with new numbers, into the Power BI source Excel from the financial system. The update schedule of an average cost data was defined to be multiple months. However, there was still one major issue. The data set did not contain any id number to map the cost data with existing data sets. This was solved by constructing an unique column.

```
= Table.AddColumn(#””, “DateRegion”, each Text.From([DATE] & ([Location]))
```

The formula presented above constructs the unique column using date and location columns combined together as a new unique column. It was found out that this method was an effective solution, since it was possible to construct a similar column also in the other data sets.

The organization and unit data set was designed for filtering purposes. The data set contained columns for filtering data and id numbers, to be able to create mapping or relationship between data sets. Multiple columns in the data set were not relevant for this project. Despite of the number of unnecessary columns, data transformation was needed with the data set. Some of the required units of the global case organization were on different organization levels. A new custom column was constructed to combine those units into one column, to be able to have all the units in the same column for easier filtering.

Cleaning was also carried out for each data set. Steps such as promoted headers, removed columns, replaced values to match the business language used in the case organization, and changed data types were done. Applied Power Query Editor steps were documented in the field notes as a text and copy of M query language. Comments were also added to the M language to ease further edits of the Power BI tool.

5.2.4 Data Set Integration Process

This section describes the steps and issues that are faced when integrating data sets 2-7 into the Project Capacity Allocation, data set 1 shown in the table 7. Most of the data sets required multiple steps of preparations to be able to achieve satisfying results documented in earlier subchapter. Before starting the integration, the project and capacity allocation data set was reviewed carefully for possible issues. Reviewing this data set took significantly more effort than the other tables due to its size and the importance of noticing possible issues for interrupting the data set integration by methods carried out in earlier subchapter.

First, the earlier detected data format issue on the project and capacity allocation data set was fixed. Date attributes were harmonized by combining columns by delimiter as a new column. Delimiter was chosen to match the date columns in the Excel data sets. Another issue was discovered while processing the project capacity and allocation data set. The size of the data set was enormous. Processing the data set has become slow due to the huge number of rows. Merging more data would make the data set even heavier. Based on the literature studied about Power BI relationships, one-way relation between data set 1 and data set 7 was created to solve this issue, in order to decrease the size by filtering only the case organization data. The next figure shows the relation between the data sets for filtering.

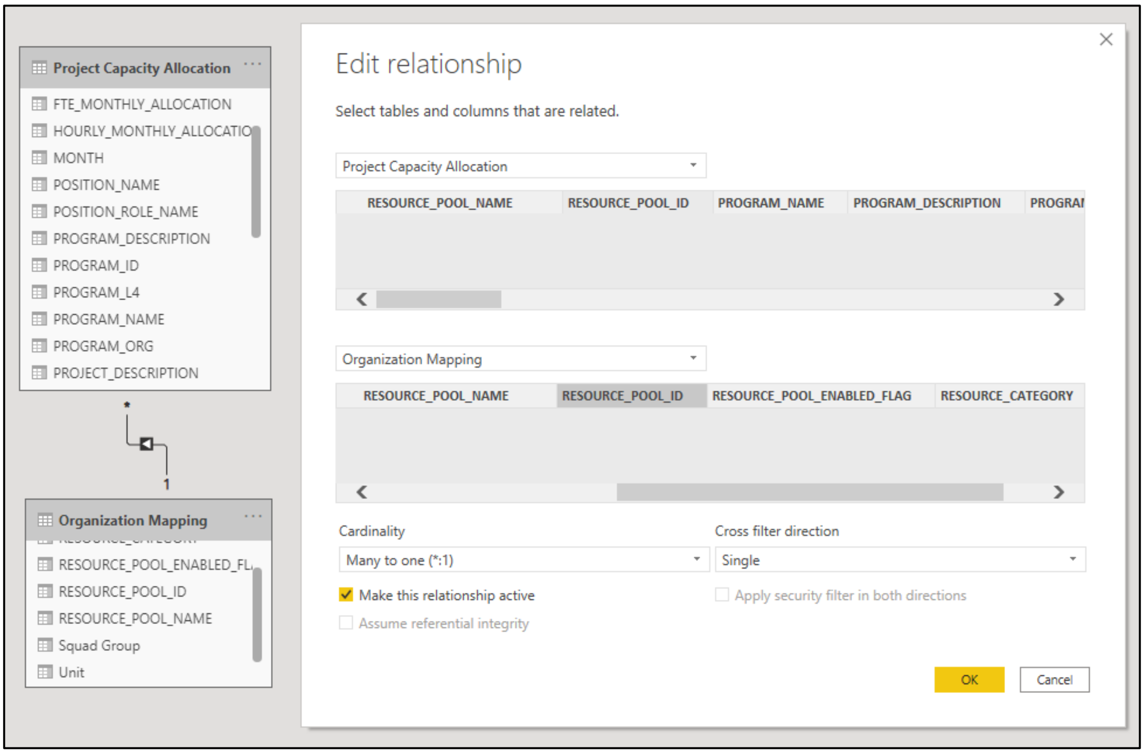

Figure 6. Relationship for Filtering Units

The figure above presents the relationship for filtering in Power BI Desktop. The relationship was many to one relationship due to the structure of data sets. The relationship was perfect for filtering, because each Resource Pool id on the organization mapping table is unique and corresponds to multiple id numbers on the project capacity allocation table. Cross filter direction was chosen to be single for filtering only the case organization data on the data set 1. This reduced the size of the data set significantly. Because of this issue the earlier data set integration order plan was reviewed again, but no other conflicts were found. The plan remained the same, except two related tables were planned to exist now in the final prototype tool proposal.

The growth data was the first one to be merged to the data set 1. The merge was conducted by using Resource Pool id as a matching column and left outer join which included all the rows from the data set 1, and all the matching ones from the growth data set. The location corresponding with the growth column and the ratio for manipulating the FTE numbers were brought in. Then new columns for Location and FTE were created to be able to manipulate the original FTE and location columns. The M language for both customs is shown next, first is the new FTE column and then the new location column including growth.

```
= Table.AddColumn(#"","FTE_Scenario", Each if Text.Contains([RE-
SOURCE_POOL_NAME],"RampUp") then [FTE_MONTHLY_ALLOCATION]*[Ratio] else
[FTE_MONTHLY_ALLOCATION])

= Table.AddColumn(#"","Location_Scenario", each if Text.Contains([RE-
SOURCE_POOL_NAME], "RampUp") then [Region Scenario] else if [REGION_NAME] = ""
then [Region Department] else [REGION_NAME])
```

Using the AddColumn formulas presented above the growth has been included into the original FTE (FTE_MONTHLY_ALLOCATION) and location (REGION_NAME) columns. At this point, the data set was tested by visualizing the FTE_Scenario and Location_Scenario attributes in the following figure.

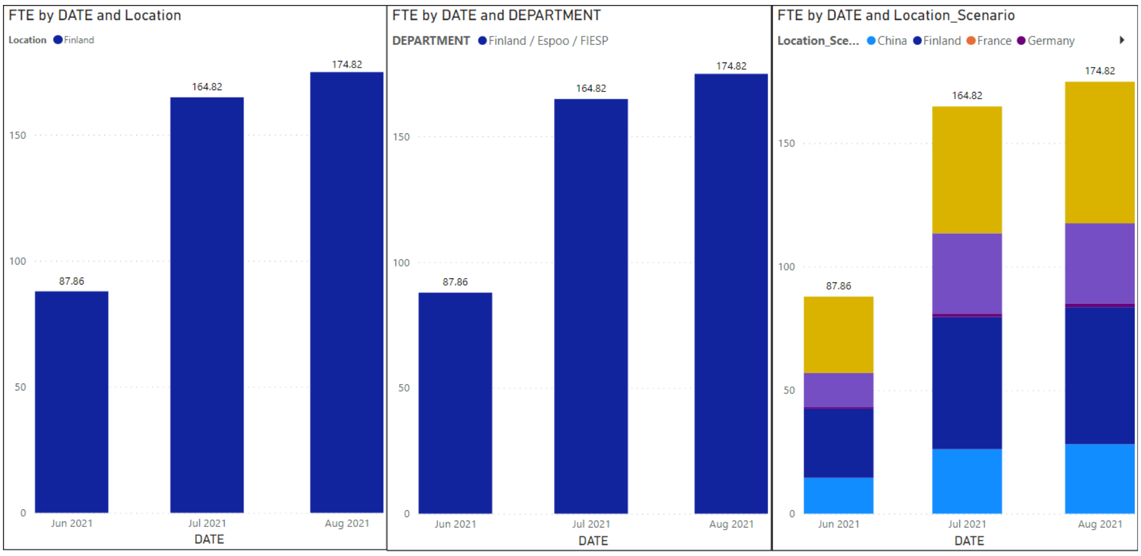

Figure 7. Data Set Test View

In figure 7 it can be seen that the previously estimated known issue became an accurate issue, because of the integrated growth data set that contains only countries. With the growth addressed to the accurate locations, the department column visualized in figure 7 became inaccurate, because changes were only applied to the FTE and location attributes. The same inaccuracy also applies to the team column, since according to the studied literature about the CRISP-DM process, now could be the correct time to return back to adjust the chosen data sets. However, this issue was reviewed at the weekly meeting with the Resource Planning Management, and it was decided to leave the issue as a known issue because of the thesis outline, and the issue did not affect seeing the overall performance of the prototype tool. It was clearly seen that the growth data set or organization data structure was causing the known issue.

Next, the ramp down data set was integrated to the data set 1. The integration was conducted by using append queries between two tables, because the purpose was to add rows on already existing columns and create new columns for integrating ramp down data on already existing columns on data set 1, and to create new columns for filtering. The appended ramp down data set included resource pools, team, location, department and ramp down numbers. The outcome was excellent, and the step was tested by creating test visuals for visualizing minus numbers per location. The appended query added the ramp down FTE numbers on the existing FTE column without creating additional custom column.

The third integration was the mapping data set. Again, the integration was conducted by using merge. The purpose of this integration was to bring the CA id column into the data set 1, in which only Resource Pool id existed. The merge was conducted by using Resource Pool id column, to combine all matching CA ids from the mapping data set to the data set 1.

The fourth integration was the product data. The merge was done by using CA id and left outer join. CA name, product name and product ratio columns were brought on the data set 1. This allowed manipulation of the already existing custom FTE column with new product mapping data, using ratio numbers. The next line of code shows how the product growth mapping has been constructed into the data set 1, in which growth was already addressed to the location.

```
= Table.AddColumn(#"", "FTE_Scenario_Product", each if [Product] = null then
[FTE_Scenario] else [FTE_Scenario] * [Ratio_Product])
```

The M code above describes how the FTE column, including location and product data has been constructed. Now the column includes now the original FTE numbers, addressed growth data per location from the data set 2 and addressed growth per product data from the data set 3.

The fifth data set integration was done for the cost data. The data set contained a lack of a unique identifier, which was constructed in the preparation subchapter. A similar column was constructed to the project capacity allocation table by using the same formula for being able to conduct the integration. The cost data was merged by using the left outer join, containing all the rows from the project capacity allocation table and the matching rows from the cost data. The following figure presents the relationship view of Power BI Desktop, in which the current integrated data set (Project Capacity Allocation) can be seen.

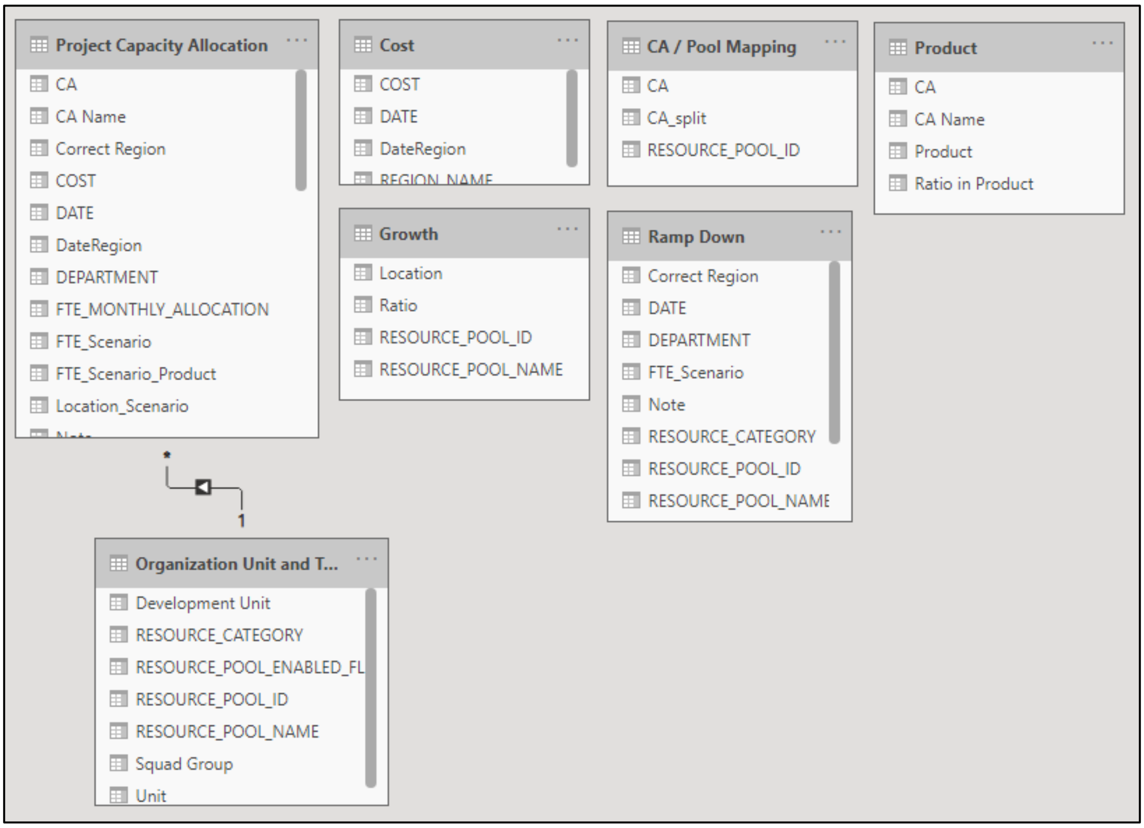

Figure 8. Current Power BI Relationship View

Figure 8 presents the integrated Project Capacity Allocation data set, all the other data sets used for the integration process, and the relation between the Project Capacity Allocation data set and Organization Unit and Team data set.

5.2.5 Transformation of the Integrated Data Set

In the earlier subchapter the data sets were integrated into the Project Capacity Allocation. The expected outcome of this subchapter is the transformed and cleaned Project Capacity Allocation data set that contains the key attributes, chosen earlier, to support the personnel resources and operating expenses planning process. The outcome data set will be used for the data modelling part of the tool development of this study. The purpose is to achieve all the requirements defined in chapter 1. The transformations were conducted by visualizing the data using Power BI, in order to capture the missing attributes and to figure out columns containing wrong arguments and renaming columns to match the business language used in the case organization.

First, the required changes, according to the field notes gathered during the data set integration process were carried out. During the integration preparation of the Ramp Down data set, the custom columns were created to be able to separate the Ramp Down data using filter and to combine the already existing columns together with ramp down data. This was carried out by using if function in Power Query Editor.

```
= Table.AddColumn(#””, “Student_RD”, each if [Student_FLAG_RD] <> null then
[Student_FLAG_RD] else [STUDENT_FLAG])

= Table.AddColumn(#””, “Team_RD”, each if [Team_Ramp Down] <> null then
[TEAM_Ramp Down] else [TEAM])
```

Both of the presented formulas constructed new columns, in which the earlier columns are combined into one, for including or excluding students and filtering the team attribute.

It was noticed that the earlier prepared and integrated product data set only contained the product mapped to the CA on the data source. However, the Project Capacity Allocation data set contained much more CA ids, which caused a large blank product. This was reviewed at a weekly meeting, and the blank product was identified to contain work types were, for example, management and other certain product-related work. At one of the weekly meetings it was decided to reduce the blank product by combining work types, such as the previously mentioned management, as a part of the product column. The new custom column was made by using text.contains and if function to reduce the size of blank product. The maintenance of the product column was also discussed at the weekly meeting. Possible changes in the Power BI Desktop file and product data set were acknowledged when Cas change in MF PPM. The lifespan of product combinations is multiple months, which made this logic seem fine. The following formula was used to construct the custom column.

```
= Table.AddColumn(#"", "Product+", each if Text.Contains([Note], "RampDown")
then [Product] else if Text.Contains([WORK_TYPE], "Management") then "Manage-
ment" else if Text.Contains([WORK_TYPE], "Management1") then "Product1"…
```

The code presents the method of collecting the required product names from the work type column, to reduce the blank product. In the first clause of the constructed conditional column, the ramp down rows were separated by using a note, which identifies the rows that contain ramp down data. The second clause on the code above presents how the required product name was captured from the work_type column.

The total cost column was created in order to tackle down the CSA finding, which is that the Resource Planning Managers do not have a tool to visualize operating expenses in a big picture. The total cost column was created by adding a calculated column. The column was constructed by calculating cost times FTE_Scenario_Product. As a final transformation to the integrated data set, the size of the Project Capacity Allocation data set was decreased. The date column was decreased to only contain one previous year, and the columns that were not necessary for data modelling phase, according to the weekly review. The load of existing data sets without relations were unenabled to reduce the size of the report.

As an outcome of the data set preparation part of the tool development of this thesis, the chosen data sets have been prepared and integrated to the transformed Project Capacity Allocation data set. The integration and transformation of the Project Capacity Allocation data set have been done in order to build a data set that supports personnel resources and operating expenses planning process and contains all the required attributes defined in chapter 1. The following figure presents the final outcome of the data set preparation.

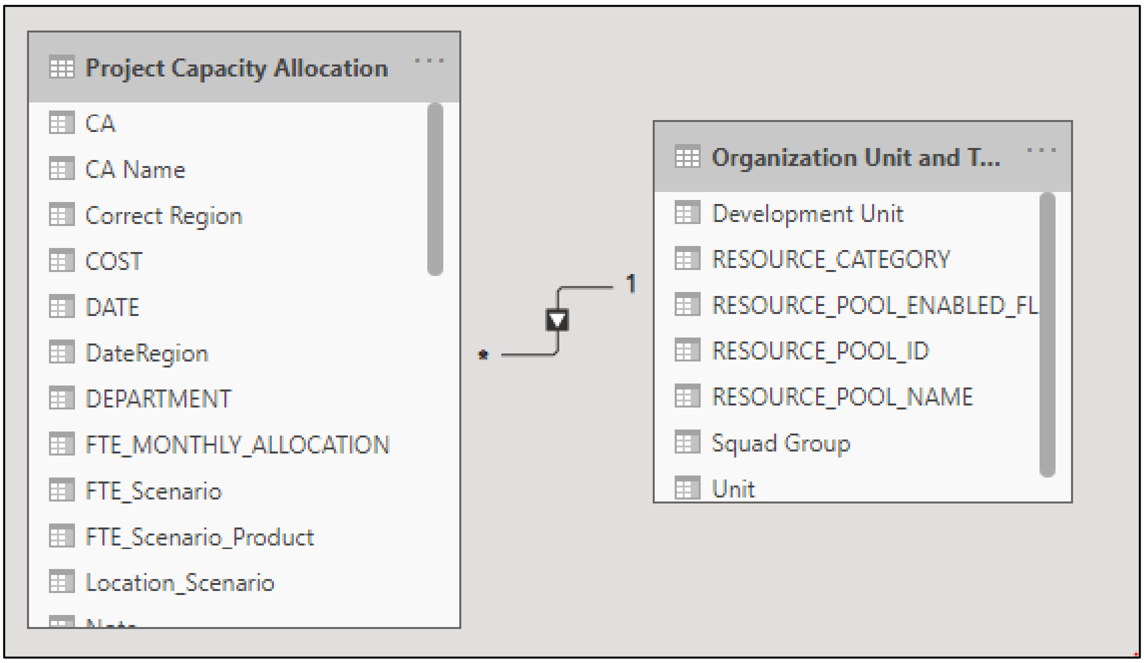

Figure 9. The Outcome of the Data Set Preparation

Figure 9 presents the integrated Project Capacity Allocation data set and the related Organization Unit and Team data set as the outcome of the data set preparation part of the tool development. The relationship direction in figure 9 presents that the Organization Unit and Team data set filters the Project Capacity Allocation data set, but not the other way around.

5.3 Data Modelling

This subchapter focuses on data modelling, leading to the initial proposal for the Power BI prototype tool. The data set constructed in subchapter 5.2 is used during the data modelling. Power BI Desktop is used as the data modelling tool because of the comprehensive data visualization and viewing options studied in chapter 4 and the key strength discovered in chapter 3. Maintaining a high quality at this part of the study is ensured by comparing the resulting FTE numbers with the validated internal data source and by gathering feedback from the Resource Planning Management team.

The models have been co-created with two key decision makers of the case organization. Individual meetings have been organized, in order to document the accurate use cases to support the needs of the key user and to gain feedback. Weekly meetings were also held to support the data modelling part of the tool development of this thesis, and reviewing the outcomes. The Line Manager and the Operations Manager provided use cases which are presented in the list below.

1. The Operations Manager specifies as follows: “I have a need to monitor how Competence Areas per headcount are correlating”.
2. The Operations Manager mentioned the following: “It is important to see how the operating expenses are correlating with the personnel growth”.
3. The Line Manager pointed out the following: “I need to be able to evaluate and validate the materialized budget per product, in order to decide on the need for re-planning”.
4. The Line Manager described as follows: “We need to be able to allocate resources based on received budget per product, and a visualized big picture of OPEX would definitely help”.

The list above describes the use cases provided by the key decision makers of the case organization. In next subchapter, the data models are first co-created based on the use cases introduced earlier. Secondly, the modelling outcomes are presented as the initial proposed prototype Power BI tool.

5.3.1 Building Models

The models were co-created with the Operations Manager and the Line Manager of the case organization, based on the use cases presented in the former list. In the beginning of the individual data modelling meetings, the main attributes were reviewed, and the transformed and integrated data set structure was described, and the known issue related to it was demonstrated. The main attributes included the modified FTE, CA names or products, location containing the planned growth, departments, teams and date. Parameter settings were set to only include the case organization data. The other parameters were planned to be conducted by using slicers, to allow the end-user control.

The first Power BI report for visualizing the correlation of CA names and FTE numbers was co-created based on the use case (1) by the Operations Manager. Two visualizations were constructed by using stacked column charts. Percentual and numeric options were used, as well as slicers, for all the main attributes. A multi row card was used to present the numbers, as requested in chapter 1. In the following figure, the co-created report of correlation of CA names and FTE numbers is presented.

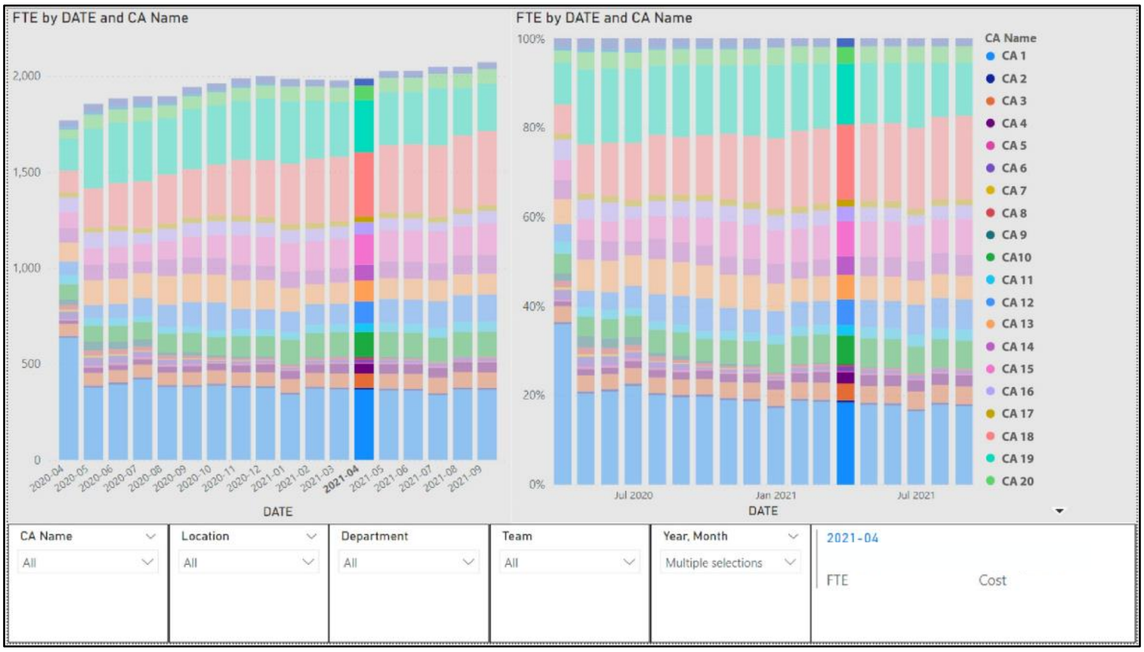

Figure 10. FTE CA Correlation Report

The figure presents the report that was co-created based on the use case (1) provided by the Operations Manager of the case organization. In figure 10, April 2021 has been chosen to demonstrate the multirow card functionality. A similar report with cost on the value axis was also constructed for providing another option, in which FTE, presented in figure 10 was replaced with Cost. In figure 10 dummy data have been used due to confidentiality of the internal resource planning data.

The second report was co-created based on the second use case (2) provided by the Operations Manager. The purpose of this report was to visualize the correlation between the operating expenses (OPEX) and FTE numbers, including growth per competence areas. This time the multirow card visualization was changed to the table for better visibility of numbers. A lack of possibility to include or exclude students and outsourced resources were also added to the report as a slicer, based on the feedback from the Operations Manager. The next figure presents the operating expenses and personnel growth correlation report.

Figure 11. OPEX and FTE per CA Report

Figure 11 visualizes the Power BI report, which was co-created based on the second use case (2) provided by the Operations Manager. The chart in figure 11 uses line and stacked column chart. The FTE values are on the left Y-axis and costs are on the right Y-axis. The table is meant to be interactive in relation to the user input. April 2021 is chosen from the chart, in the same way as in figure 10, to demonstrate the interactive table in the report. In figure 11 dummy data have been used due to confidentiality of the internal resource planning data.

The Power BI report for validating materialized budget per product was co-created by using the same charts as in the earlier report, in order to provide a clear structure of the prototype tool in different Power BI report pages. This report was co-created with the

Line Manager. The product column, constructed previously in subchapter 5.2.5, was used instead of CA attribute as the legend of the line and stacked column chart. The next figure presents the Power BI report.

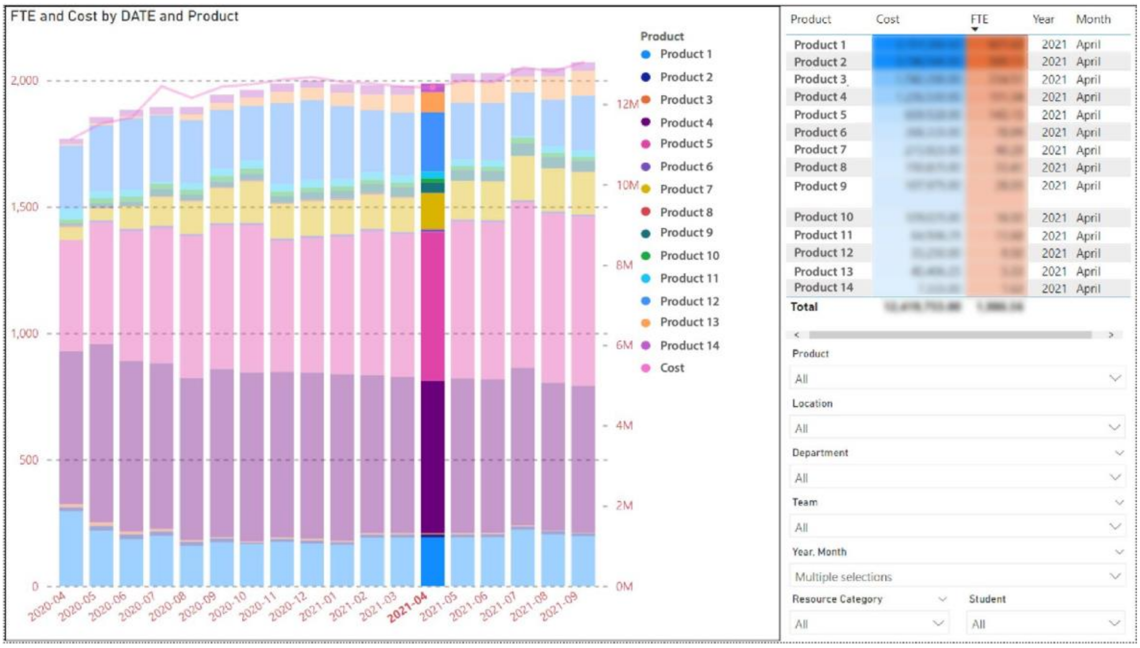

Figure 12. OPEX and FTE per Product Report

The figure above presents the report that was co-created based on the use case (4) provided by the Line Manager. Other visualizations using 100% stacked column chart with FTE and cost by product and date was also constructed for supporting both use cases (3) and (4) provided by the Line Manager. Dummy data have been used due to confidentiality of the internal resource planning data, presented in figure 12.

5.3.2 Initial Prototype Power BI Tool

The prototype Power BI tool development of this thesis was divided into data set preparation and data modelling parts. The outcome of the data set preparation part is the integrated and transformed Project Capacity Allocation data set, presented in figure 9 in subchapter 5.2.5. The outcome of the tool development part is the initial Power BI tool which was co-created in the earlier subchapter based on the use cases provided by the Operations Manager and the Line Manager. The Power BI tool uses the integrated Project Capacity Allocation data set to report the personnel resources and operating expenses data visually and in numbers. The initial proposed prototype Power BI tool as the Power BI reports is presented in figures 10-12.

The initial prototype Power BI reports, co-created to support the personnel resources and operating expenses planning, were reviewed by the Resource Planning Managers of the case organization. The Operations Manager and the Line Manager reviewed the co-created reports, the Power BI tool, in the individual meetings. Both of the key users were satisfied with the tool development outcomes presented in earlier subchapter in figures 10-12. The Operations Manager summed up the biggest benefits of the co-created Power BI reports as follows:

> Seeing the visual of the big picture makes challenging of the needs for new recruitments easier, when we can see the big picture of headcount per location, including the planned growth. (Operations Manager)

Other benefit that was mentioned during the individual review of the co-created report was:

> This tool presents the materialized and planned OPEX so much better than the old tool, which helps an adjusting the budget. (Line Manager)

The set data mining goal was also seen as achieved because the chosen data sets were transformed into being able to build models to support the personnel resources and operating expenses planning process.

At the meeting with the Resource Planning Managers, the results were seen as promising, regardless of the known error that was detected earlier. The data provided by the Power BI reports was compared with the internal data source, and the FTE numbers were accurate per country. Growth mapping and products were looking accurate, when compared with the simultaneously updated Location Plan document on a brief review. At the individual meeting with the Operations Manager the most important notice was that:

> The team, resource category and student slicers must be on each report. (Operations Manager)

Resource category and student slicers had an important role, because the operating expenses for external employees and students are calculated differently. The Line Manager had also brought up the previously mentioned slicers and a need for department slicer, which was the known issue due to because of the structure of the growth data set described earlier in subchapter 5.2.3.

Based on the feedback gained from the Operations Manager and the Line Manager the table in the report was seen better than the multirow card for displaying the numbers. The percentual ratio of cost was also requested to be added in the table. The key decision makers thought the line and stacked column chart an excellent choice and fitting the purpose. No pre-defined slicers were identified to be good and supporting multiple different resource planning use cases, based on the feedback. The known issue with the department and team attributes was seen as an excellent example of creating a need for further development of the initial prototype tool. Regardless of the known issue, the Power BI reports were seen as a working prototype tool, and it was thought that the integrated data set will already be able to support the personnel and operating expenses planning process. The validation of the initial prototype Power BI tool and the data set was seen as necessary in the review.

6 Validation

This chapter covers the validation of the initial Power BI prototype tool. Validation is a part of a development process and of the research design of this study. This phase has been carried out in order to test and validate the developed initial outcome of this study. This chapter combines the final prototype tool proposal as the outcome of this thesis. The validation chapter of this thesis consists of three subchapters, including the data source validation, output data validation and the final proposed prototype Power BI tool.

Validation meetings were held with the Resource Planning Managers. Also, there were detailed reviews and discussion of the final prototype tool and use case tests with the Operations Manager and the Line Manager as the key users of the prototype tool. The prototype Power BI tool was also reviewed by the Planning Specialist of the case organization. Data collection 3 for validating the prototype tool is presented in chapter 2. In the subchapters of this chapter, it is first described how the chosen data sets for the Power BI prototype tool were validated, and secondly, how the output data from the Power BI prototype tool was validated and planned to be maintained. This is followed by a description of the final proposed prototype Power BI tool as the outcome of this thesis.

6.1 Data Source Validation

This subchapter covers the validation of chosen data sources that were used in the tool development of this study. Seven different data sets were used as data sources to construct the Power BI prototype tool. However, three of them were from a centrally maintained and validated internal Power BI Dataflows source which did not require additional validation. The Dataflows data sets were the Project Capacity Allocation, Unit and Team and mapping data sets. The maintenance cycle of the Dataflows data sets was also significantly fast. The list of the data sets can be seen in table 7 in subchapter 5.2.2.

The four manually maintained resource planning Excel data sets were the cost data, growth data, ramp down data and product data. It was decided to place the manually maintained Excel sources in the internal Microsoft Teams folder, in which the access was already limited only to the Resource Planning Management team due to confidential data. In case of an update, the Resource Planning Managers could easily access the documents easily in the Team folder. Placing the Excel data sets into the Teams folder also made the refreshing of the Power BI report easier and enabled a fully automated Power BI report refresh for both the Excel and Dataflows data sets, for later deployment of the tool.

The cost data comes from the Corporate Financial System, which is an already validated source that does not require anything else than a review for the updates. The Financial System Provides the data in standard format, which was used in the cost Excel data set. The update of the cost data can be done by copying and pasting the new numbers into the document, by replacing the old numbers. The growth Excel data set is in the same format as the Resource Planning Managers were using, which made it easy to be updated according to the higher organization plans. The already existing frequent screening validation method is planned to be used for this data source.

The Ramp down and product data are also coming from sources that are were in the same format as the Resource Planning Managers were using before, and the data is based on the higher organization plan. Same screening, validation and maintenance methods are planned to be used, as previously for the Excel data sets. The ramp down data is updated less frequently. The updates of the ramp down values are planned to be added to the Excel data source.

The maintenance cycle for the Growth data and Product data sets was set to be according to the maintenance cycle mentioned during the CSA. The product and growth data updates are planned to be conducted by replacing the values into the Excel data sets. When changing only the growth or product ratios, no further edits are required to the Excel data sets or Power BI Desktop document, only a refresh. In case new products, locations, CAs or Resource Pools are added to the product or growth data sets, changes are required in the Power BI desktop file. The assigned ownership of the proposed final Power BI desktop document was considered as needed in the Resource Planning Management team.

6.2 Output Data Validation

The output data of the Power BI report was validated by using multiple methods. As the prototype Power BI tool uses the integrated and transformed Project Capacity Allocation and Excel data sets, careful validation of the output data was done. In order to succeed in the validation, all the Power BI data sets, the Location Plan document, Ramp Down document, and the validated internal tool for data validation were ensured to be updated at the same time. Slicers to choose between original Project Capacity Allocation data from MF PPM, growth and ramp down, were added on the Power BI report.

The full-time equivalent (FTE) numbers and personnel costs data, provided by the prototype Power BI tool, were validated by comparing the values with the validated internal tool and the prototype tool data sources. The most recent resource planning material, per location from the Location Plan document, was also used during the validation. The product and growth data provided by the prototype tool were validated by using the planning material from the Location Plan document for comparing. The ramp down numbers were validated with the internal source and the Ramp Down data source Excel by comparing the data with the data provided by the prototype tool.

The slicers for department and team attributes were incorrect because only planned growth data per location existed in the Growth data set. This is an issue and requires further development, because the department and team attributes were the key attributes for the personnel resources and operating expenses planning, and the validated initial prototype Power BI tool certainly requires them. As an outcome of the output data validation, the data provided by the Power BI report were seen as accurate, except for the known issue with the department and the team attributes.

6.3 Final Prototype Power BI Tool

This chapter describes the validated final proposed prototype Power BI tool developed to support the personnel resources and operating expenses planning. The tool development of this thesis was divided into data set preparation and data modelling parts described in chapter 5. The integrated and transformed Project Capacity Allocation data set remained the same as presented in figure 9 in chapter 5.2.5. Minor changes were added into the Power BI tool since the initial prototype tool that was presented in sub-chapter 5.3.2.

The co-created reports, chosen by the Line Manager and the Operations Manager of the case organization, were combined into their own pages in the final Power BI tool. The following figure presents the CA page.

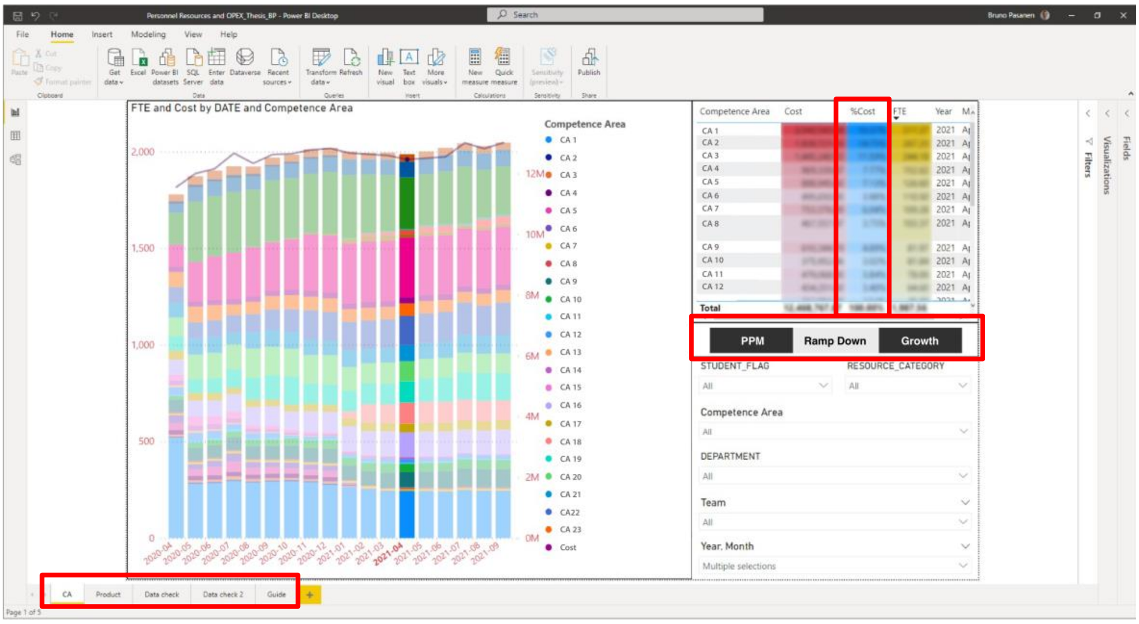

Figure 13. CA Page of the Final Proposed Power BI Prototype Tool

As seen in figure 13, minor changes have been made since the initial proposed prototype Power BI tool. The first two pages are based on the initial co-created reports, the second and third pages are for checking the data, and the last page is a guide. The percentual numbers of costs were added on the table based on the feedback. Individual reports have also been combined together into one report, containing five pages in total. Slicer to choose between original data coming from MF PPM, ramp down and growth data sources were also added, as this was enabled during the data set integration process. The next figure presents the Product page of the final proposed Power BI prototype tool.

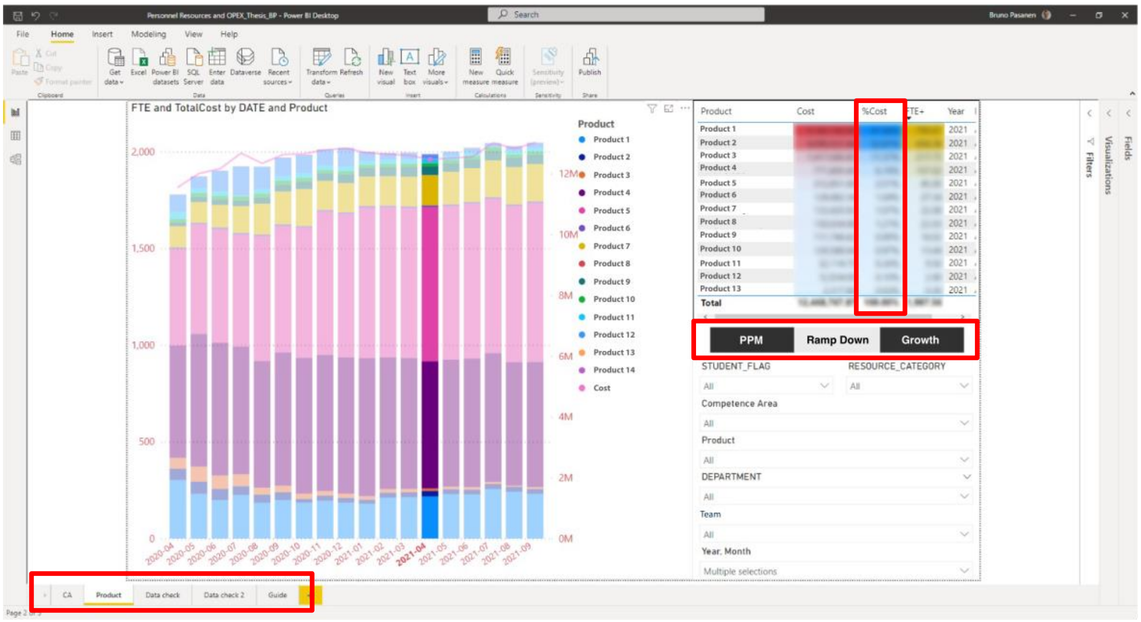

Figure 14. Product Page of the Final Proposed Power BI Prototype Tool

Figure 14 presents the product page, including the same changes as the previously explained CA page. Dummy data have been used, due to confidentiality of the internal resource planning data, in figures 13 and 14.

The Data Check pages and the Guide page were also added to the final proposed Power BI tool, in order to present the data behind the visualizations and to provide instructions for the end-users. The next figure presents the data check page.

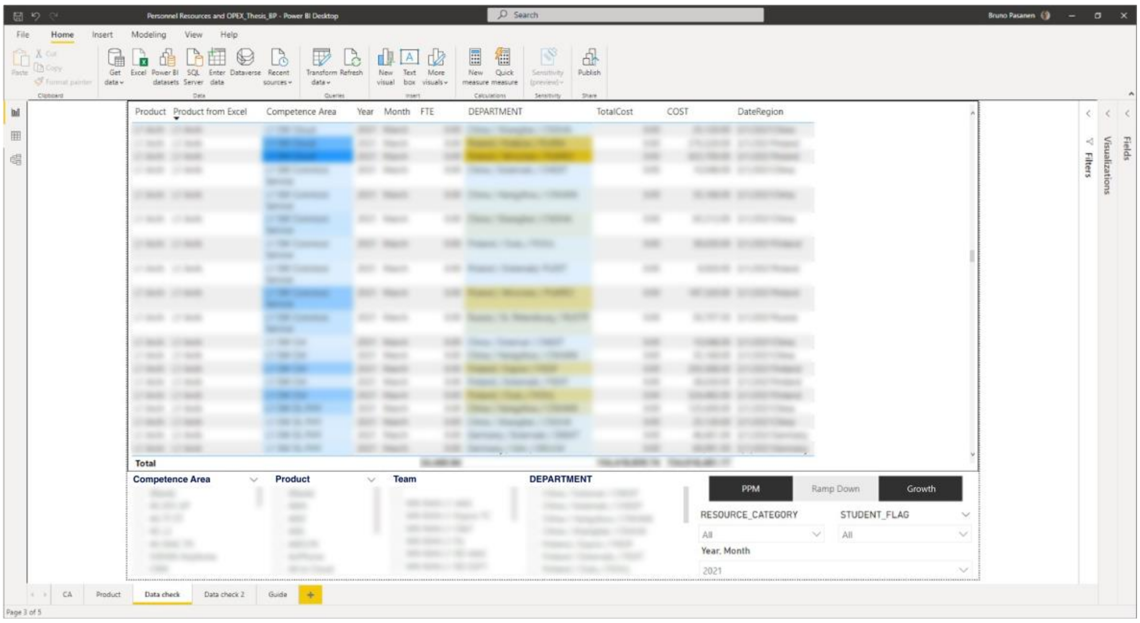

Figure 15. Data Check Page of the Final Proposed Power BI Prototype Tool

As seen, the data check page of the final proposed Power BI prototype tool contains a large table to view the data with slicers to filter it. Data Check 2 page contains different columns and slicers, but the layout of the page is the same as Data Check page's. Guide page is a text page that provides information about the Power BI tool for personnel resources and operating expenses planning. Dummy data have been used due to confidentiality of the internal resource planning data in figure 15.

The final proposed Power BI prototype tool contains seven initial data sets, including Power BI Dataflows data sets and manual Excel data sets. The data sets are visualized in figure 8 in chapter 5.2.4. The purpose of the Excel data sets was to integrate the resource planning data into the organization data as an outcome of transformed and integrated Project Capacity Allocation data set. In the final proposed Power BI prototype tool, the end user can interact with the chart and table presented in figures 13 and 14 by using the slicers and clicking the data in the chart or in the table. The end-user can also review the data by using data check pages and read instructions for using the tool from the guide page. In figures 13 and 14, April 2021 has been chosen to demonstrate the interaction of the chart and the table. In figures 13 and 14, only date attribute has been filtered by using the slicer. Other slicers visible in figures 13 and 14 have been set to all. The case organization data is set as a pre-filtered parameter in the final prototype Power BI tool on each page.

The following figure shows the summary of the prototype Power BI tool development of this study.

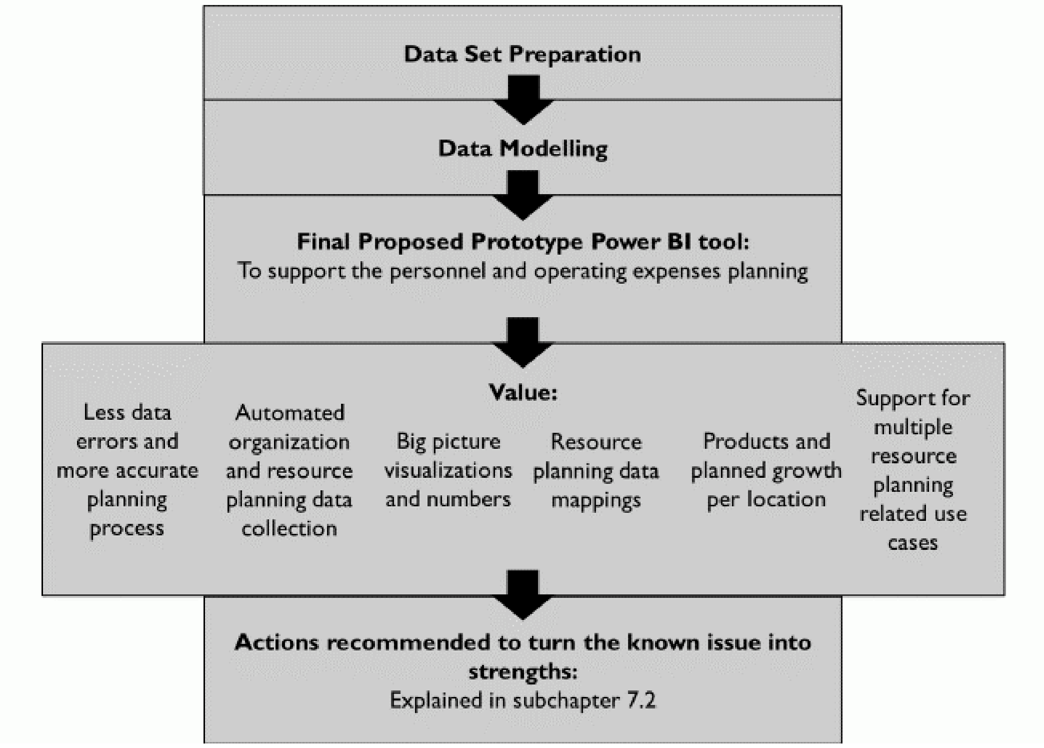

Figure 16. Summary of the Proposed Prototype Power BI Tool Development

Figure 16 presents the summary of the proposed prototype Power BI tool development structure, leading to the outcome, the final proposed prototype Power BI tool and the value gained from the outcome of this study. The summary figure also presents the path to the actions that are recommended for turning the known issue into strengths.

The Resource Planning Managers of the case organization saw the final prototype Power BI tool supporting the personnel resources and operating expenses planning process. The tool collects the data from multiple different resource planning and organization data sources, which was a criterion of this thesis. The final proposed prototype Power BI tool is bringing value in many aspects, and it suits the use cases defined in this study. The tool was capable of building a big picture visualization of personnel resources and operating expenses, and it could provide the data in numbers and including other requirements stated in chapter 1, such as containing product data. The used data sources were validated to be trusted and frequently maintained, and the tool was possible to be updated on a frequent basis, and updates of Power BI report were also possible to automated. The constructed data set included data mappings between the organization and resource planning data, which significantly decreased the manual work and possible data errors.

The known issue with department and team attributes was discovered in subchapter 5.2.3 and it was validated to require further development in subchapter 6.1. The recommendations to solve the known issue are presented in subchapter 7.2 as a next step for this thesis. The recommendations aim to turn the known issue into strengths and to provide more accurate resource planning data for the case organization and an easily maintained Power BI tool to support the personnel resources and operating expenses planning process, as well as for other resource planning use cases.

7 Conclusion

This is the final chapter of this study and its purpose is to conclude this thesis. This chapter contains an executive summary, recommendations for a next step, thesis self-evaluation and final closing words.

7.1 Executive Summary

The objective of this study was to propose a prototype tool that combines personnel resources and operating expenses data, including product and growth data for monitoring and planning purposes. The expected outcome of this study was a proposal for a prototype tool that allows the Resource Planning Managers to have one supporting tool for personnel resources and operating expenses planning which collects the data from multiple reliable sources. Currently, the Resource Planning Managers have made use of multiple different tools to conduct the planning. This topic was chosen due to the importance of collecting evidence-based data, the need to reduce manual work and the aim for a more accurate and simplified business decision-making process with continuously maintained and validated data.

Action research was chosen as the research approach for this study. Qualitative data modelling and gathering method were also utilized. The research design of this study included four phases. The first phase was the current state analysis including comprehensive research of the current resource planning team, data and tools, leading to the key strengths and weaknesses of the current resource planning data. The second phase was the studying of the literature and best practices, based on the key current state analysis findings which led to the conceptual framework of this study. The third phase was the tool development, which was divided into data set preparation and data modelling parts, leading to the initial proposed prototype Power BI tool. The tool development phase was conducted in close collaboration with the Resource Planning Managers of the case organization. The fourth phase was the validation of the initial proposed prototype Power BI tool, leading to the outcome of this thesis as the final proposed prototype Power BI tool. The recommendations for the next step are presented in the chapter 7.2.

The current state analysis of this thesis was carried out with the help of multiple meetings arranged with the Resource Planning Managers of the case organization, starting from the kick-off. Business and data understanding of the case organization were also studied comprehensively by working as a part of the case organization, participating in resource planning related meetings and observing multiple internal documents related to the resource planning of the case organization. The current state analysis included analysis of the Resource Planning Management team, tools used in resource planning and current resource planning data. The key findings were collected after combining current state analysis (CSA) findings as strengths and weaknesses.

Literature and best practices were studied based on the key findings of the CSA, in order to find the best possible ways to develop the requested prototype tool. The Cross Industry Standardized Data Mining Process (CRISP-DM), other data mining and analytics related sources, Microsoft Power BI and Excel were studied which led to the conceptual framework of this study. Conceptual framework included the studied literature and best practices for solving the key findings of the current state analysis presented in chapter 4.3.

The tool development of this thesis was conducted in close cooperation with the Resource Planning Managers of the case organization. Weekly meetings were held during the project and additional meetings were scheduled for solving problems and gaining deeper understanding, as well as feedback. The tool development included data set preparation and data modelling parts. The data set preparation was the largest part of this study, as it included choosing, transforming, integrating and normalizing the chosen data sets in order to develop the data set for supporting the personnel resources and operating expenses planning. In the modelling part, the data set was used to co-create Power BI reports based on the use cases provided by the Line Manager and the Operations Manager of the case organization leading to the initial proposed prototype Power BI tool in chapter 5.3.2.

The validation phase of this study included the validation of the chosen data sets that were used to construct the integrated and transformed Project Capacity Allocation data set and the validation of the output data of the initial proposed Power BI tool. The initial Power BI prototype tool was reviewed, and feedback was gained from the Resource

Planning Managers of the case organization in multiple meetings. The Power BI prototype tool was seen as an excellent tool for supporting the personnel resources and operating expenses planning. The prototype Power BI tool, the known issue discovered in subchapter 5.2.3 and this thesis were also clearly presenting the need for improving the internal resource planning data structure to support the resource planning process better.

The final proposed Power BI tool included the co-created Power BI report with five pages that all used the integrated and transformed Project Capacity Allocation data set. The data set preparation is described in subchapter 5.2. The outcome of the data set preparation part is presented in figure 9 in subchapter 5.2. The initial proposed Power BI tool was co-created in subchapter 5.3, and the final proposed Power BI tool is presented in subchapter 6.3 in figures 13-15 and summarized in figure 16. The final proposed Power BI tool contains minor changes that were done in relation to the feedback from the Resource Planning Managers during the data modelling and validation part of the tool development. The final proposed prototype Power BI tool achieves the objective of this study and is bringing value in many aspects as well as suiting the use cases defined in this study.

7.2 Recommendations for Next Steps

Before the deployment of the final proposed prototype Power BI tool to the case organization, it is strongly recommended to carry out at least one of the alternative recommendations. This subchapter presents two alternative recommendations for tackling the known error discovered in this study and a recommendation for adding product attribute in MF PPM data. All the recommendations increase the resource planning data accuracy and ease the maintenance of the proposed Power BI tool. First two alternative recommendations are presented. The third one is a recommendation to add the product attribute. The third recommendation increases the resource planning data quality even without using this Power BI tool.

In relation to the gained knowledge and received feedback during this thesis, the final proposed prototype Power BI tool is already able to achieve the objective of this thesis. However, resolving the discovered known issue with department and team attributes is certainly necessary to provide the best possible value from the Power BI tool, and consequently, received value from the resource planning data will increase significantly. The team and the department attributes are important for planning the personnel resources and operating expenses, defined with the Line Manager and the Operations Manager during this study. The known error of the inaccurate department column was identified in chapter 5.2.3 because of the growth mapping data set or the lack of planned growth data per department and team in MF PPM tool.

7.2.1 Recommendation for Changing Growth Data Maintenance Logic

It is recommended to add the planned growth data per location, department and team, to the Project Capacity Allocation data set in MF PPM tool. The outcome of this recommendation is estimated to be that the correctly addressed growth data would be a part of the Project Capacity Allocation data set in MF PPM tool. This would reduce the maintenance of the proposed Power BI tool and increase its data accuracy significantly. The growth data set would become redundant without affecting the other data sets of the Power BI tool. However, maintenance of the Power BI tool would be required. This recommendation would also have an increasing effect on the resource planning data accuracy throughout the case organization.

During the CSA, the maintenance of the growth data was mentioned to be done in the Resource Planning Management team. The logic for growth data maintenance is most likely better to be spread into each team in the case organization. This would reduce the overload of work for the Resource Planning Management team when it is required to maintain the growth data in individual teams of the case organization. The individual teams most likely have the best view of their growth and recruitment statuses. Also, the maintenance cycle of the planned growth data must be discussed. It is also extremely important to harmonize the maintenance logic between each team if this recommendation is taken into use.

However, it must be considered that pushing any kind of resource planning data into a tool such as MF PPM cannot lead towards wrong decision making in the case organization. Piloting this alternative recommendation is suggested to see its validity.

7.2.2 Recommendation for Combining Growth and Ramp Down Data Sets

The second alternative recommendation is to fix the known issue by utilizing the team and department attributes in the same way as with Ramp Down data set. During this study, the Ramp Down data set was validated to be working and containing required attributes. It is recommended to combine the growth data and Ramp Down data into the same data set by using a similar format as the Ramp Down data set that was used in this study. This could be done by adding growth numbers into the same document and in the same time frame as the ramp down numbers which are negative. The next figure presents how a recommended growth and ramp down data set would look.

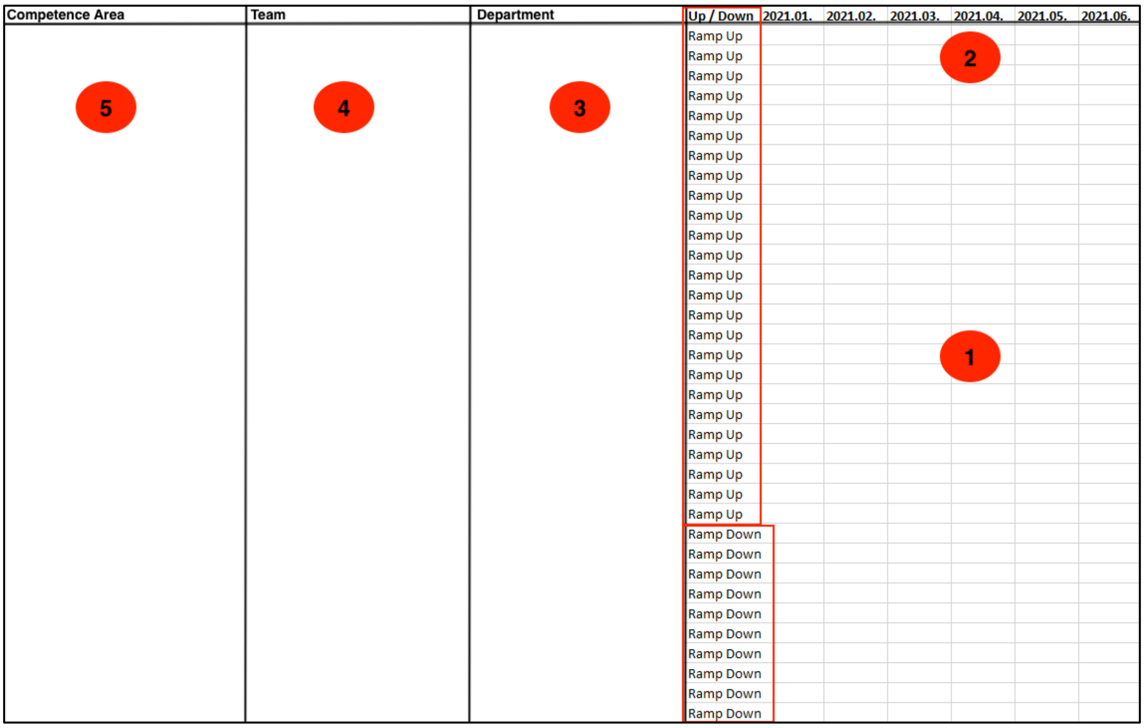

Figure 17. Recommended Combined Data Set Example

Figure 17 presents the recommended combined growth and ramp down data sets. The column circled with red identifies the growth as ramp up and ramp down values. The positive or negative FTE values would be placed in table (1) and a date row (2) on the top, making it possible to unpivot the table in Power Query Editor easily. Department (3), Team (4) and Competence Area (5) columns are recommended to be created in the same format as in the Project Capacity Allocation data set in MF PPM. The same format significantly eases the integration, which is recommended to be done using append, because then the new rows will be added on the already existing ones. However, it is good to remember the column names are case-sensitive. This recommendation should make it possible to use team and department attributes without any issues. The overall benefit from implementing this recommended change is that the sensitive resource planning data will be less complex to maintain in a single document and to use for monitoring and planning purposes.

7.2.3 Recommendation for Adding the Product Attribute

The third recommendation is to add the product attribute to the Project Capacity Allocation data set in MF PPM tool. The estimated outcome of this recommendation is that the product attribute is available in the Project Capacity Allocation data set. The attribute could be created by mapping the certain Competence Areas, Resource Pools or work types to a certain product name, as was done during the data set preparation of this thesis by using Power Query editor. The product attribute in MF PPM tool would decrease the maintenance effort of the final proposed prototype Power BI tool. By implementing this recommendation, the product data set and the constructed column would become useless when the attribute already exists in the MF PPM tool. The product names exist in multiple places through the case organization and are certainly needed in resource planning, according to this study and the work at the case organization. The product attribute in MF PPM data would harmonize the data understanding of an employee and increase the value and quality of the resource planning data in the case organization even without using this Power BI tool. Most likely this recommendation would provide significant value to the whole case organization.

Possible value gained for implementing the recommendations is clearly significant, which definitely makes continuing this project necessary. Validating the possible needs for this kind of supporting tool for Resource Planning Managers in other organizations of the case company is also suggested. Regarding the discussions with the Resource Planning Managers of the case organization, and having carried out this study, it can be said that the same issues might be bothering other organizations also.

7.3 Thesis Self-evaluation

The objective of this thesis was to propose a prototype tool which combines personnel resources and operating expenses data, including product and growth data for monitoring and planning purposes stated in subchapter 1.2. Three research data collection rounds were conducted, and feedback was gathered throughout the study. In order to ensure the logic of this study the robust, a four-phase research design, was created in the beginning. The key findings of the current state analysis are concrete strengths and weaknesses in the case organization, collected from multiple discussions, observations and meetings. The tool development of this study was the third phase, and it was divided into two parts, and it is a detailed path to the comprehensive initial outcome of this study presented in chapter 5. The tool development was co-created with experts of the case organization. Validation which leads to the final outcome of this study is presented in chapter 6. Continuous collaboration, analysis, development process, and validation of this study were excellent in terms of research, which can be seen from the concrete outcome and recommendations as a next step of this thesis.

The final outcome of this thesis is a co-created working prototype Power BI tool which supports the personnel resources and operating expenses planning process. In the tool development phase of this thesis, an issue with department and team attributes was detected, and it was decided to be left as a known issue because fixing it would have gone over the scope of this project. Consequently, the outcome of this study achieves the objective with stated criteria in subchapter 1.3.

Action research is the research approach of this study and definitely the most suitable one for it. The author of this study is working as a member of the Resource Planning Management team. Kananen (2013) states that a conductor of action research participates in the research process, which does not happen in design research (Kananen 2013:41). Relevant elements for action research are, for example, development of operations and continuous cooperation (Kananen 2013:44). Both of the earlier mentioned elements occur in this thesis clearly. I have looked at the research from multiple angles, gathered feedback, and comprehensive research data has been collected for current state analysis, tool development and validation.

Being a part of the team and understanding the internal organization data and business have given a significant advantage to being able to develop the prototype Power BI tool. Arranging thesis-related meetings and other activities with the stakeholders were a part of natural daily work. Involving the key stakeholders was mandatory for this thesis. However, I believe that having deeper understanding of MF PPM Project Capacity Allocation data set structure in the beginning and knowing how the key attributes have been formulated could have eased the project. In other words, conducting a deeper current state analysis would have made it faster to carry out this study. At each phase of this thesis, my learning has been constant, and towards the end more significant findings were covered and those will continue outside of this study. During this study, changes of the internal organization structure created additional challenges for this project. Nevertheless, the changes in the organization data structure during this thesis had a strengthening effect on the understanding of the author. Learning by doing provided a great path towards the outcome of this thesis.

7.4 Final words

Development of new tools and ways of working are necessary for staying on track with the constantly evolving world and technology. Quality data and information are creating a tremendous amount of value in an organization. This project underlined the importance of being able to use quality data for business decision making, and even more importantly, a real-life practice of developing, learning about and constructing the data for receiving real business benefits and value to the organization.

References

Aggarwal, C.C. 2015. Data mining: the textbook. Springer International Publishing.

Chapman et al. 1999. The CRISP-DM Process Model. [Online]. Available at: https://www.spss.ch/upload/1109179904_CRISP-DM%20Process%20Model.pdf [Accessed 17 01 2021].

Data Science Project Management. 2020. What is CRISP DM. [Online]. Available at: https://www.datascience-pm.com/crisp-dm-2/ [Accessed 20 01 2021].

Erro-Garcés, A. and Alfaro-Tanco, J. 2020. Action Research as a Meta-Methodology in the Management Field. International Journal of Qualitative Methods, 19. [Online]. Available at: https://journals.sagepub.com/doi/pdf/10.1177/1609406920917489 [Accessed 11 02 2021].

IBM. 2012. IBM SPSS Modeller CRISP-DM Guide. [Online]. Available at:ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerCRISPDM.pdf [Accessed 21 01 2021].

Kananen, J. 2013. A Practical Guide for Thesis Research. JAMK.

Microsoft 1. 2021. What is Power BI. [Online]. Available at: https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview [Accessed 17 01 2021].

Microsoft 2. 2021. What is Power BI Desktop. [Online]. Available at: https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-what-is-desktop [Accessed 17 01 2021].

Microsoft 3. 2019. Shape and combine data in Power BI Desktop. [Online]. Available at: https://docs.microsoft.com/en-us/power-bi/connect-data/desktop-shape-and-combine-data [Accessed 08 02 2021].

Microsoft 4. 2020. Power Query M language specification. [Online]. Available at: https://docs.microsoft.com/en-us/powerquery-m/power-query-m-language-specification [Accessed 05 02 2021].

Microsoft 5. 2020. Power BI data sources. [Online]. Available at: https://docs.microsoft.com/en-us/power-bi/connect-data/power-bi-data-sources [Accessed 01 02 2021].

Microsoft 6. 2020. What are dataflows?. [Online]. Available at: https://docs.microsoft.com/en-us/power-query/dataflows/overview-dataflows-across-power-platform-dynamics-365#benefits-of-dataflows [Accessed 06 02 2021].

Microsoft 7. 2020. Creating a dataflow. [Online]. Available at: https://docs.microsoft.com/es-us/power-bi/transform-model/dataflows/dataflows-create [Accessed 01 02 2021].

Microsoft 8. 2021. What is Power BI Service. [Online]. Available at: https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-service-overview [Accessed 17 01 2021].

Nokia 1. 2021. About Us | Nokia Corporation. [Online]. Available at: https://www.nokia.com/about-us/ [Accessed 21 03 2021].

Nokia 2. 2021. Our history | Nokia Corporation. [Online] Available at: https://www.nokia.com/about-us/company/our-history/ [Accessed 21 03 2021].

North, M. 2012. Data Mining for the Masses. Global Text.

Redman, T. 2008. Data Driven. Harvard Business Press.

Saltz, J. 2020. CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects. [Online]. Available at: https://www.datascience-pm.com/crisp-dm-still-most-popular/ [Accessed 03 02 2021].

Stobierski, T. 2019. The advantages of data-driven decision-making. Harvard Business School. [Blog] Business Insights. Available at: https://online.hbs.edu/blog/post/data-driven-decision-making [Accessed 23 01 2021].

Stobierski 1, T. 2021. What's the difference between data analytics & data science?. Harvard Business School. [Blog] Business Insights. Available at: https://online.hbs.edu/blog/post/data-analytics-vs-data-science [Accessed 24 01 2021].

Stobierski 2, T. 2021. Top data visualization tools for business professionals. Harvard Business School. [Blog] Business Insights. Accessible at: https://online.hbs.edu/blog/post/data-visualization-tools [Accessed 24 01 2021].