UNIVERSITY OF OULU

FACULTY OF TECHNOLOGY

TOWARDS BETTER ORGANIZATIONAL ANALYTICS CAPABILITY – A MATURITY MODEL

Tuomas Antti Lassila

Industrial Engineering and Management

Master's thesis

October 2020

FACULTY OF TECHNOLOGY

TOWARDS BETTER ORGANIZATIONAL ANALYTICS CAPABILITY – A MATURITY MODEL

Tuomas Antti Lassila

Industrial Engineering and Management

Master's thesis

October 2020

ABSTRACT

Towards better organizational analytics capability – a maturity model

Tuomas Antti Lassila

University of Oulu, Degree Programme of Industrial Engineering and Management

Master's thesis 2020, 87 pp. + 2 Appendixes

Supervisors at the university: Kauppila O. & Lampela H.

Data and analytics are changing the markets. Significant improvements in competitiveness can be achieved through utilizing data and analytics. Data and analytics can be used to support in all levels of decision making from operational to strategic levels. However, studies suggest that organizations are failing to realize these benefits. Many of the analytics initiatives fail and only a small partition of organizations' data is used in decision making.

This happens mostly because utilizing data and analytics in larger scale is a difficult and complex matter. Companies need to harness multiple resources and capabilities in a business context and use them synergistically to deliver value. Capabilities must be developed step by step and cannot be bought. Bottlenecks like siloed data, lack of commitment and lack of understanding slow down the development.

The focus of this thesis is to gain insight on how these resources and capabilities can be managed and understood better to pursue a position where modern applications of data and analytics could be utilized even better. The study is conducted in two parts. In the first part, the terminology, disciplines, analytics capabilities, and success factors of data and analytics development are examined through the literature. Then a comprehensive tool for identifying and reviewing these analytics capabilities is built through analyzing and combining existing tools and earlier insights. This tool, organizational analytics maturity model, and other findings are then reviewed and complemented with empirical interviews.

The main findings of this thesis were mapped analytics capabilities, success factors of analytics, and the organizational analytics maturity model. These results help practitioners and researchers to better understand the complexity of the subject and what dimensions must be taken into account when pursuing success with data and analytics.

Keywords: analytics capability, analytics maturity, organizational analytics maturity model, analytics development

TIIVISTELMÄ

Kohti parempaa organisaation analytiikkakyvykkyyttä - maturiteettimalli

Towards better organizational analytics capability – a maturity model

Tuomas Antti Lassila

Oulun yliopisto, tuotantotalouden tutkinto-ohjelma

Diplomityö 2020, 87 s. + 2 liitettä

Työn ohjaajat yliopistolla: Kauppila O. & Lampela H.

Datan ja analytiikka muuttaa eri organisaatioiden välistä kilpailua. Huomattavia parannuksia kilpailukyvyssä voidaan saada aikaan oikeanlaisella datan ja analytiikan hyödyntämisellä. Data ja analytiikkaa voidaan käyttää kaikilla päätöksen teon asteilla operatiivisista päätöksistä strategiselle tasolle asti. Tästä huolimatta tutkimukset osoittavat, että organisaatiot eivät ole onnistuneet saavuttamaan näitä hyötyjä. Monet analytiikka-aloitteet epäonnistuvat ja vain pientä osaa yritysten keräämästä datasta hyödynnetään päätöksenteossa.

Tämä johtuu pääosin siitä, että datan ja analytiikan hyödyntäminen isossa kontekstissa on vaikeaa ja monimutkaista. Organisaatioiden täytyy valjastaa useita resursseja ja kyvykkyyksiä liiketoimintakontekstissa ja käyttää näitä synergisesti tuottaakseen arvoa. Näitä kyvykkyyksiä ei voida ostaa suoraan, vaan ne joudutaan asteittain kehittämään osaksi organisaatiota. Kehitykseen liittyy myös paljon ongelmakohtia, jotka hidastavat kokonaiskehitystä. Siiloutunut data ja sitoutumisen ja ymmärryksen puute ovat esimerkkejä kehityksen kompastuskivistä.

Tämän opinnäytteen tarkoitus on syventää ymmärrystä siitä, miten näitä resursseja ja kyvykkyyksiä hallitaan ja ymmärretään paremmin. Miten organisaatio pääsee tilaan, jossa se voi hyödyntää moderneja datan ja analytiikan mahdollisuuksia? Tutkimus muodostuu kahdesta osasta. Ensimmäisessä osassa käsitellään terminologia, analytiikkakyvykkyydet ja niiden menestystekijät. Sen jälkeen luodaan kokonaisvaltainen työkalu, organisaation analytiikkamaturiteettimalli, kyvykkyyksien tunnistamiseksi ja kehittämiseksi. Tämä malli rakennetaan ensimmäisten löydösten pohjalta. Tutkimuksen toisessa osassa aiemmat löydökset ja rakennettu malli validoidaan ja täydennetään empiirisillä haastatteluilla.

Tämän työn päälöydökset ovat kartoitetut analytiikkakyvykkyydet, niiden menestystekijät ja organisaation analytiikkamaturiteettimalli. Nämä löydökset auttavat ammattilaisia ja tutkijoita ymmärtämään paremmin aiheen monimutkaisuuden ja mitä dimensioita tulee ottaa huomioon, kun pyritään menestykseen datan ja analytiikan avulla.

Avainsanat: analytiikkakyvykkyys, analytiikkamaturiteetti, organisaation analytiikkamaturiteettimalli, analytiikan kehittäminen

FOREWORD

The idea for this thesis formed from my personal interests. I was originally planning to conduct a commissioned work for a certain company but due unfortunate global events the agreed work was cancelled. However, I see the cancellation as a lucky incident. Doing this thesis on my own opened me a wide variety of opportunities and helped me to complement my skills on a subject that I personally find interesting.

I would like to thank my thesis supervisors Osmo Kauppila and Hannele Lampela for the guidance and support during the research. I would also want to thank all the interviewed organizations and their representatives for your efforts and interest towards my work. These interesting discussions gave me understanding about the subject from practitioner's point of view. Without it this thesis would have not been possible.

I would like to express my gratitude for my parents for everything. Lastly, special thanks belong to all my friends during the time in the University of Oulu. With you, the last five years has been one of the best periods of my life.

Oulu, 31.10.2020

Antti Lassila

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BDA	Big Data & Analytics
CDO	Chief Data Officer
CMM	Capability Maturity Model
IS	Information System
RDBMS	Relational Databases and Relational Database Management Systems
SME	Small and Medium Enterprise
SQL	Structured Query Language

1 INTRODUCTION

"*One executive we interviewed compared the complexity of managing the development of analytical capabilities to playing a fifteen-level chess game.*"

(Davenport and Harris 2017)

1.1 Background

Globalization and the global competition have made it challenging for businesses to thrive and companies must ponder more and more strategic questions such as: How to differentiate from other companies? How market share could be sustained and gained? (Chevalier-Roignant and Trigeorgis 2011) Davenport and Harris (2007) support this view of the challenging markets and offer one approach for these questions in their book *Competing on Analytics: The New Science of Winning*. In global business competition, geographical advantages or protective regulations do not largely matter anymore. Many industries offer similar products, and proprietary technologies are rapidly copied. For several companies', important way to differentiate is to execute your business with maximum efficiency and effectiveness. Making the smartest business decisions possible. Analytics can help in this. (Davenport and Harris 2007)

Good analytical decisions help companies to thrive by supporting them to be better at operational business processes. Identifying profitable customers, optimally pricing products, hiring the right people, optimizing inventories and supply chains are tasks that can be supported with analytics and these decisions improve operational efficiency of firms. On the other hand, analytics can also help to make strategic decisions. How to choose best locations for different facilities, how to decide the right acquisitions and mergers to scale businesses? Good decisions usually require data and analytics behind them. (Davenport and Harris 2007) In addition to this, analytics and data can stand behind not only decisions, but also products and services (Davenport & Harris 2017).

The amount of data and computing power used in analytics has been rapidly growing in recent years. Research made by (Hilbert and López 2011) shows that worlds capacity of general-purpose computing grew at an annual rate of 58% from 1986 to 2007. Capacity for bidirectional telecommunication grew 28% per year and globally stored information 28% per year in the same time frame. Currently this growth of data is increasing even more rapidly, and research made by (Reinsel et al. 2018) presents an estimation that global amount of data will grow from 33 Zettabytes in 2018 to 175 Zettabytes in 2025. Most of this new rapidly growing data is unstructured data (Dhar 2013). This rapid growth of data and computing power enable new ways to leverage data and analytics.

Firstly, this so-called big data caused technical problems in organizations due its volume, variety, and velocity but nowadays modern analytical tools and databases can handle these vast amounts of data and the big data can be seen as business opportunity (Russom 2011). (McAfee et al. 2012) describe in their article that this revolution of big data is far more powerful than the analytics used in the past. Companies can measure and manage business more precisely than ever before. Businesses shift from decisions made by gut and intuition to data and facts. Big data analytics enable better predictions and smarter decisions. (McAfee et al. 2012)

In addition to the big data in recent years there has been also emerging more and more hype around artificial intelligence and cognitive technologies. Technologies like statistical machine learning, neural networks and natural language processing can help businesses to do even better decisions when used correctly. (Davenport 2018)

However, all the benefits from these new technologies and enablers are not fully utilized yet. SAS's report made in UK in showed that in 2015 56% of businesses in UK use big data on some level but majority of the companies are just beginning to realize this opportunity and have only implemented between one to three big data analytics solutions. (Hogan et al. 2016). Also it has been reported that less than 1% of organizations unstructured data is analyzed or used at all and less than 50% of their structured data is used in active decision making (DalleMule and Davenport 2017). There are also many chokepoints which slow down the analytics adaptation. Issues like lack of commitment, siloed data, poor understanding how to use data, and failing to understand the value in it can cause integration issues for data and analytics. (Ramanathan et al. 2017)

Gaining success with analytics is a complex matter. Companies need to harness multiple recourses and capabilities (technologies, people, process, data, and organizational) in a business context and use these synergistically to deliver value. (Vidgen et al. 2017) Analytics capability cannot be bought. It takes time for the organization to build analytics capability through gradually developing all the different dimensions of analytics. (Davenport & Harris 2017) The focus of this thesis is to gain insight how these resources and capabilities can be managed and understood better to pursue a position where modern applications of data and analytics could be utilized even better.

1.2 Research objectives and scope

In the literature, data and analytics capabilities are seen of as separate entities. Even though data without tools and applications, as well as the processing applications without data gives no benefits for the company (Aydiner et al. 2019) The goal of the research is to understand different stages of data and analytics maturity and to synthesize a holistic framework based on existing literature for evaluating organizations data and analytics capabilities. What factors should be measured when analyzing these capabilities and how to identify them. This framework can be then used to determine how mature an organization is in different fields of data and analytics and help the organization to clarify their focus where and how to build these capabilities for further analytics adaptation. Also, the insights from empirical interviews will be used to complement the model. The following research questions form the basis of the thesis.

RQ1: What are organizational data and analytics capabilities?

RQ2: What factors accelerate the development of organizational data and analytics capabilities?

RQ3: How can organizations identify and review their current data and analytics capabilities and how to gain better insight about them?

The thesis will only partially study how the data and analytics capabilities should be developed. The main focus will be on identification and examination of these capabilities.

2 ANALYTICS, ANALYTICS CAPABILITIES, AND ANALYTICS MATURITY

This chapter firstly reviews prior research and literature about data, analytics, and different disciplines of working with data. Also, the definitions of capabilities, maturity and maturity models are handled. Aim is to create a holistic understanding about these themes, but also clarify the used terminology in this thesis. Based on this theorical foundation of basic concepts, research on data and analytics capabilities and maturities is conducted to build a frame for answering research question 1. Success factors for data and analytics adaptation are inspected to help to understand research question 2. Finally, the assessed literature is synthesized into a theorical framework to build basis for research question 3.

2.1 Data, big data, databases, and data governance

Data and big data

"Data is a representation of facts, concepts, or instructions in a formalized manner, suitable for communication, interpretation, or processing by humans or by automatic means" (Organisation for Economic Co-operation and Development 2003). Data by itself has little relevance or purpose. It only describes facts about past events. But data is important for organizations because it is essential raw material for creating information. Once data becomes information it has purpose and has value. (Davenport and Prusak 2000). This process of transforming raw data into usable information is called data analysis (Organisation for Economic Co-operation and Development 2003).

When the organized collection of data, the data set, is so large, quickly changing or coming from multiple sources, so that you have to change your mind-set how to analyze it or use it in a different way compared to a normal data set, it is called big data (Tonidandel et al. 2016). Even though the definition of big data is vague, it is usually defined by its three main characteristics. These three characteristics are the usually called the three Vs of big data. They are volume, variety, and velocity of data. The volume aspect of the big data refers the sheer size of the data in term of the number of data points and how much disk space it uses. The variety of big data means that analyzed data might have multiple sources or multiple forms. The varied data may also have unstructured or semi structured forms, or the data can come from audio, video, and other devices, which makes it even more difficult to analyze with traditional means. The velocity aspect expresses frequency of new data generation or frequency in data delivery. For example, big data can be collected in real-time. Sensors detecting the surrounding environment, automated measures of manufacturing processes, or web sites collecting the actions of the visitors are examples of real-time gathered big data. (Russom 2011)

Sometimes two new V's are added to the definition of big data. These are viscosity, meaning the latency or the data's delay to changes, and veracity which means the accuracy of the data. However (Tonidandel et al. 2016) argue that the veracity should not be considered as one of the defining characteristics of big data when differentiating from more traditional data sets since veracity and the accuracy of data is essential for all the data.

A survey made (Ward and Barker 2013) also adds that many times when speaking of big data also the technologies and infrastructure of big data are included in the definition of big data. "Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to NoSQL, MapReduce and machine learning."(Ward and Barker 2013)

Since big data does not have single commonly agreed definition this thesis will understand the big data only as an extension of traditional data. As a raw material for analytics like traditional data but having the three differentiating characteristics earlier described by Tonidandel et al. (2016) and Russom (2011).

Databases, storing data, and collecting data

To organize, represent, and keep the data consistent relational model of data was created in 1970 (Codd 2002). Based on this model relational databases and relational database management systems (RDBMSs) where data could be stored digitally were created. User could interact with an RDBMS with writing queries in Structured Query Language (SQL). Most importantly this interacting meant taking the data tables, joining, and morphing them into new, more complex tables. These so-called SQL databases were seen as de facto option for any instance until non-relational database paradigm or NoSQL databases emerged. (Perkins et al. 2018) The main difference between SQL and NoSQL databases is the data model that the database uses. In SQL database it is mainly relational but in NoSQL it can be something else. For example, the data model can be key-valued or column-oriented. (Han, Haihong, et al. 2011)

These NoSQL databases were able to solve some limitations of more traditional SQL databases such as problems with concurrent reading and writing causing high latency, big data storage and access needs, scalability and availability problems, slow data manipulation speeds when database contains large amounts of data, and high maintenance costs. (Han, Haihong, et al. 2011) However, NoSQL has its own limitations and has not replaced SQL databases. Both types have their use cases and the options for database should be considered based on the needs of the data, data model, and the database. (Perkins et al. 2018)

Collection of integrated databases designed to serve only informational or analytical needs is called a data warehouse. This data warehouse house is usually separated from operational databases because the data serving operational needs is physically different from the data serving analytical purposes and because the supporting technology for operational processing is essentially different from the technology used to assists informational or analytical needs. The main reasons to use data warehouse are the following:

- There is only single source of the truth.
- The data can be reconciled if necessary.
- Data is immediately available for new and unknown uses.

(Inmon 2005)

To support needs of big data and large and quickly arriving volumes of unstructured data, data lakes were introduced (Miloslavskaya and Tolstoy 2016). Data lakes are storages and processing systems that ingest data without compromising the structure of the data in contrast to data warehouses highly structured data. Data lake holds a vast amount of raw data in its native format until it is needed. Comparing data lakes to data warehouses there is also other benefits. Data lakes are more cost efficient and less constrained by performance and storage capacities. Also, the lack of predefined schema gives possibility to analyze the data in its raw unstructured form. (Laskowski 2016) Fang (2015) supports this view in his research and notes also that data lakes are tightly tied to Apache Hadoop and its ecosystem because it technologically feasible and cost effective way to fill the needs of data lakes and big data (Fang 2015)

In recent years alternative ways to store data has emerged. Instead storing the data locally organizations can use off-site storage maintained by third party. These remote access databases are called cloud storages. Connection between the user of the data and the database is provided through Internet. Cloud storages are viable choice when data must be accessed from any location or by multiple users conveniently. Cloud computing and storages also offers huge scalability, reliability, high performance, and specifiable configurability. (Wu et al. 2010)

Organizations can gather data internally by observing, measuring, or collecting it by means of questioning as in surveys. However, most of the data nowadays is captured via automatic means and for example by measuring processes. (Organisation for Economic Co-operation and Development 2003) s.155 In addition to these internal data collection methods, organizations can also obtain data from external sources, partnering or collaborating with other organizations or from buying the needed data from commercial data providers. For example, Fey & Birkinshaw (2005) argue in their research paper that using these external data sources might have positive impact on performance of R&D activities. Though it was noted that it may also have negative impact depending on the source. Particularly in R&D activities partnering with universities for gathering data had a positive impact on the performance. (Fey and Birkinshaw 2005)

Data quality and data governance

When processing and analyzing the data, a poor-quality data might have concerning social and economic impacts (Wang and Strong 1996). Problems with data quality costs US businesses more than 611 billion dollars in 2009 (Khatri and Brown 2010). To gain valuable information out of data, it is essential that the used data is accurate and high quality. Data quality can mean different thing for different data users and there has been recorded over 118 attributes linked to the data quality (Wang and Strong 1996). According to Wang and Strong (1996) these can be grouped to four main categories:

- Intrinsic data qualities. Which means accuracy, objectivity, believability, and reputation of the data.
- Contextual data qualities. These qualities consist of value-added, relevancy, timeliness, completeness, and appropriate amount of data.
- Representational data qualities. Which denote interpretability, ease of understanding, representational consistency, and concise presentation.
- Accessibility data qualities. These qualities are accessibility and access secure of the data.

Newer research has been conducted to expand the definition of data quality and fulfill the gaps in the body of knowledge created by modern characteristics of data. Analysis conducted by Jayawardene et. al (2015) splits the data quality dimension into eight main categories. These are completeness, availability & accessibility, currency, accuracy, validity, usability & interpretability, reliability and credibility, and consistency. (Jayawardene et al. 2015)

To ensure this high quality of data, governance is needed. Data governance sets the requirements of intended use of data and the standards for data quality in the organization. In addition to data quality, data governance also covers domains of data principles, metadata, data access and data lifecycle. Purpose of data principles is to clarify the role of data as an asset and establish the direction for all other decisions regarding the data. Data principles answers questions like what are uses of data for the business, who is the owner of data assets, how are opportunities for sharing and reuse of data identified, and how should businesses communicate about the data. Domain of metadata includes the basis for how data is interpreted and explains the content of the data for its users. Data access governance is specifying access requirements of data and what are the standards and procedures for data access. Data lifecycle domain determines the production, retention, and retirement of data. (Khatri and Brown 2010) Ladley supports this view and summarizes in his book *Data governance: How to design, deploy, and sustain an effective data governance program* (2019) the definition of data governance into the following

"Data governance is the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets." (Ladley 2019)

Ladley (2019) underlines that data governance is not a function performed by those who manage the information. Governance should be left for top management. Governance only provides the rules and policies how data management should happen. Only practical data management is done under information management function. (Ladley 2019)

2.2 Data analysis, statistical analysis, business intelligence, data science, and other data related disciplines

"Data analysis is the process of transforming raw data into usable information" (Organisation for Economic Co-operation and Development 2003). According to Ramsay (2004) goals of data analysis include at least the following:

- Represent the data in ways that aid further analysis.
- Display the data as to highlight various characteristics of the data.
- Study important sources of pattern and variation among the data.
- Explain variation in an outcome or dependent variable by using input or independent variable information.

(Ramsay 2004)

In addition, Hair et al. (2019) argue that prediction of the outcome is one goal for data analysis.

The process of analysis starts with defining problem, determining what data is needed, collecting the data, then using different methods to summarize and analyse the data, and making decisions based on the data. (Newbold et al. 2013) This method of answering the question based on the data is called confirmatory data analysis. When the analysis is conducted without any pre-conceived ideas to discover what the data can tell, the research is called exploratory data analysis. (Tukey 1977)

The overall process of discovering useful knowledge from data is called *Knowledge Discovery in Databases* process or *Knowledge Discovery* Process. This includes the process of analysis of the data but also additional steps. These steps are the following: data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, data analysis or data mining and proper interpretation of the results of the analysis. (Fayyad et al. 1996) The figure 1. clarifies the process steps and their outputs.

Figure 1. Knowledge discovery process (modified from Fayyad et al. 1996)

Descriptive, predictive, prescriptive, and other types of analytics

Usually analytics techniques are categorized into three main types: descriptive, predictive, and prescriptive (Souza 2014, Davenport and Harris 2017). Descriptive analytics aims to answer question what is happening based on data. It provides ability to report, explore and alert. (Davenport and Harris 2017)

Predictive analytics can be seen more advanced and provides abilities to understand why the phenomenon is happening, what happens if the trend continues, and what will happen next. Predictive modelling uses quantitative methods and technologies to predict future based on past data. (Souza 2014, Davenport and Harris 2017)

Prescriptive analytics gives decision recommendations based on variety of predictive and descriptive analytics models. It answers questions of what will happen if we try this, and what is the best that can happen. Goal is to specify the optimal behaviour and actions. (Souza 2014, Davenport and Harris 2017)

Davenport & Harris (2017) also adds one category more in addition to these previous three. This is called autonomous analytics and it employs techniques like artificial intelligence and cognitive technologies. Goal is to create and improve models and learn from data without human hypotheses. Answers the question “What can we learn from the data?” (Davenport and Harris 2017)

Predictive, prescriptive, and autonomous analytics are sometimes referred as advanced analytics. Descriptive analytics answering questions about what happened and why things happen is considered to be traditional analytics. (Intel 2017)

Two approaches for data analysis

Development of technologies and era of big data has also created new demands for the analytical techniques to deal with new and varied sources of data. This has resulted to acknowledgement of two different and distinct “cultures” or schools of data analysis. These cultures are statistical/data models and algorithmic models/data mining models. (Breiman 2001) & (Hair et al. 2019) They both work under the same conditions. Both have data or variables, research problem, and goal to predict the outcome based on the inputted variables or to gain information how these variables affect the outcome. However, the difference between these two disciplines is how the different models approach the problem. The statistical or data model way approaches the problem by basing the model upon theory and then executing a research design to test that model and underlying theory. The algorithmic or data mining models handle the problem differently. Instead of describing the process they focus on the best algorithms that can reproduce the process and perform on highest predictive accuracy. (Hair et al. 2019) Table 1. further clarifies the differences between these two approaches.

Table 1. Comparison between statistical/data models and data mining/algorithmic models (modified from Hair et al. 2019)

Characteristic	Statistical/Data Models	Data Mining/Algorithmic Models
Research Objective	Primarily Explanation	Prediction
Research Paradigm	Theory-based (deductive)	Heuristic-based (inductive)
Nature of Problem	Structured	Unstructured
Nature of Model Development	Confirmatory	Exploratory
Type of Data Analyzed	Well defined, collected for purpose of the research	Undefined, generally analysis used data available

Scope of the Analysis	Small to large datasets (number of variables and/or observations)	Very large datasets (number of variables and/or observations)

These two ways of conducting data analysis have their own purposes, strengths, and weaknesses. The differences between these cultures do not make one method better than the other. Choosing the right model depends on the situation. For example, the situations where the analyzed process is so complicated (e.g. autonomous cars) that it is almost impossible to model with statistical means, algorithmic models like machine learning seems more appropriate. And when more insightful analysis about variables affecting the outcome is needed, theory-based data models might be more useful. (Hair et al. 2019) The purpose of distinguishing of these two models is not to replace one with another but give scientist and analysts wider variety of tools to conduct analyses (Breiman 2001).

Statistical analysis and data models

Typically, statistical data analysis is encountered in physical sciences but also when analyzing business opportunities and making better decisions in uncertain environment (Newbold et al. 2013). Statistical data analysis has been conducted traditionally via mechanical calculators and by hand until the 90's when growing computing power allowed more complex and efficient analyses with computers. The basis of statistical data analysis is still in the mathematics and statistics and applying statistical methods such as hypothesis testing, linear regression, analysis of variance and maximum likelihood estimation on the data. (Efron and Tibshirani 1991)

Statistical analysis use data models which make the basis of any statistical data analysis. These stochastic data models aim to represent and simulate the process, that is examined, as well as possible. The model is formed by the researcher and is then estimated using the data available to assess the model fit and its usability. Data models help to analyze the process and its outcome but there are also some risks regarding the models. Flawed data model might cause incorrect interpretation of the process. (Hair et al. 2019)

Data mining

Growing amount of data emerged more disciplines, such as data mining, to support statistical data analysis. Data mining is act of discovering interesting, unexpected, or valuable structures or patterns in large data sets. (Hand 2007) & (Hong et al. 1999) Datamining is also referred as *knowledge discovery from data* (Han, Pei, et al. 2011) As such, it has two rather different aspects. One of these concerns large-scale, 'global' structures, and the aim is to model the shapes, features of the shapes, or distributions. The other aspect concerns small-scale, 'local' structures, and the aim is to detect these anomalies and decide if they are real or chance occurrences." (Hand 2007)

Data mining is one step of the knowledge discovery process (the step where intelligent methods are applied to extract information from cleaned data). However often the term data mining is used to refer to the entire knowledge discovery process. (Han, Pei, et al. 2011)

The most common outputs and the goals of the datamining can be categorized to the following types:

- Class/concept description. Data entries can be associated to a class or concept. The description of these classes can be acquired using data characterization, data discrimination or using both simultaneously. Data characterization means summarization of general features of target class of data. Data discrimination means contrasting classes or act of discovering differentiating features for two different classes.
- Discovering frequent patterns. Process of detecting interesting associations and correlations within data. For example, a frequent item set that is bought together or sequential pattern of items bought in sequence.
- Classification and regression. Classification is the act of distinguishing data classes or concepts using a model based on training data with class labels. The model is then used to predict the class label of new data which has unknown class label. For example, banks classifying the risk level of new loans based on history data about previous loans and their classification. Regression predicts continuous values instead of class labels. Regression analysis is also used for detecting trends.
- Cluster analysis. This means grouping data objects with goal of maximizing the intraclass similarity and minimizing the interclass similarity. Process of creating

classes without pre-defined class label. For example, identifying homogenous subpopulations among all the customers.

- Outlier analysis. Analysing a data object which behaviour does not comply with general behaviour of the data. Goal is to find out the reasons and the features of this outlier. This is used for example in fraud detection.

(Han, Pei, et al. 2011)

Machine learning, Neural Networks, & Artificial Intelligence

To reach the goals of data mining efficient algorithms are required. These algorithms learn from available data and then produce most appropriate output to estimate unknown data. These kinds of algorithms are called machine learning. (Marsland 2015) Machine learning is not able to replicate the examined process completely, but a good and useful approximation of the process, a model, can be built on data. This model has the same base and theory as statistical and mathematical models but in machine learning the model is modified and optimized by the algorithm itself by studying the earlier data and its features from this examined process. (Marsland 2015) & (Alpaydin 2020)

Different problems require different kind of machine learning algorithms. Depending on different sources there are three to five main types of machine learning algorithms. The three main algorithms are the following:

- Supervised learning. Training data for the model includes correct responses or labels for the data. The machine learning algorithm then generalises to respond correctly to all possible inputs, based on this training data.
- Unsupervised learning. Correct responses or labels are not provided, and the algorithm is trying to identify similarities between inputs. Goal is to categorise the inputs that have something in common.
- Reinforced learning. The algorithm is only told when it is wrong, and it must explore and try out different possibilities until it gets the right answer.'

(Marsland 2015, Davenport 2018)

Marsland (2015) argues also that evolutionary learning is also one distinct type of machine learning algorithms. Simulation of biological evolution as a learning process. The model is scored for how good the current solution is and then next generation of models are generated based on this. (Marsland 2015)

Ayodele (2010) adds that there is also semi-supervised learning where the training data combines labelled and unlabelled data to train the model, transduction learning where the model learns from the training data but also from the outputs and from the new inputs. In addition, there is also learning to learn type of algorithms which learns its own inductive bias based on earlier experience. (Ayodele 2010)

More sophisticated form of machine learning is the neural network. The basis is the same as for machine learning, there are inputs with different features which affect the outputs. However, the logic behind the algorithm is more complex and difficult to interpret. The neural network algorithm combines input variables into perceptrons which are used to estimate the output but typically have little meaning to humans. Especially when the neural network has multiple layers, when the number of different features and perceptrons affecting the output can be in thousands. The multi-layered neural networks are called deep learning. (Davenport 2018) & (Alpaydin 2020)

Technologies that have some kind of cognitive capabilities, such as machine learning and its ability to learn, are called as artificial intelligence (AI). There is considerable ambiguity in the term of artificial intelligence, but Davenport (2018) proposes that, in addition to machine learning, neural networks and deep learning, at least the following technologies should be considered as AI:

- Natural language processing. Process of analysing and understanding human speech and text.
- Rule-based expert systems. Systems which have set of logical rules derived from human experts.
- Physical robots which automate physical activities.
- Robotic process automation. This means automation of digital tasks and processes.

Many aspects of AI are out of the scope of this thesis. Therefore, this thesis will discuss only about characteristics of AI that are related to data analysis such as machine learning and neural networks which can be seen as advanced techniques for data mining or data analysis.

Data science and Business intelligence

Data science is “A term intended to unify statistics, data analysis and related methods. Consists of three phases, design for data, collection of data and analysis of data.” (Everitt and Skrondal 2010) Data science can be seen as a combination and extension of statistics and data mining, but it differentiates from statistics and other related disciplines in several ways. Whereas more traditional statistical data analysis uses relatively small and structured data, data science can use heterogenous and unstructured data such as text, images, and videos. To analyze these types of data, tools from computer science, linguistics, sociology, econometrics, and other studies are needed. (Dhar 2013) Provost & Fawcett (2013) support this view in their article and note that core of data science is use of techniques for mining data, but it also covers more than that. Good data scientist should understand the data, source of the data, database, and the problem and its context (Provost and Fawcett 2013).

Analytics adaptation to business context emerged in the 90’s and 2000’s. This new disciple was business intelligence. (Chen et al. 2012) Goal of business intelligence is to present complex and competitive information and knowledge to planners and decision makers combining operational data and analytical tools (Negash and Gray 2008). In addition to analysing data, business intelligence is considered to be an umbrella term which includes also data mining, data warehousing, data gathering and knowledge management in business environment. (Negash and Gray 2008) & (Xia and Gong 2014) Newer research argue that the data science has been included under the umbrella term of business intelligence (Larson and Chang 2016).

Business intelligence as a term is heavily linked to business intelligence systems. These are the IT systems or software applications of business intelligence which conduct the practical analyses and deliver the information to decisions makers. Usually these business intelligence systems include technological components such as interface for decision makers to visualize and work with the data by themselves and databases to store the data. (Richards et al. 2019), (Negash and Gray 2008) & (Xia and Gong 2014)

Business intelligence and data science have significant similarities and overlapping. However, research has shown that data scientist are generally much more data and technology orientated than business intelligence professionals. Data science tool kits are usually more sophisticated and more diversified comparing to business intelligence tools. Where business intelligence is heavily focused on the business context, data science is not locked to the business domain and is utilized in other fields as well. The same survey also reveals that data science is more focused on working on big data than on normal data. (Cao 2017)

The meaning of business intelligence has been changing during the years and it does not have a single agreed definition. There has been also discussion that business intelligence means only the business intelligence systems. This thesis will understand business intelligence as a discipline which goal is to present complex and competitive information and knowledge to planners and decision makers by combining operational data and analytical tools as Negash and Gray (2008) described it but heavily focusing on business side and leaving the most sophisticated analysis methods to data science as Cao (2017) showed in his article.

2.2.1 Summary of the disciplines and terms

People from different backgrounds use different terminology to describe the same actions and phenomena. Also, the popularity of different terms has changed a lot during recent years. Figure 3. shows how much different terms have been searched relatively to each other in Google search engine over time. Search interest in data science and data analytics has been growing in recent years where interest in business intelligence, statistical analysis, and data mining has been gradually decreased over time. (Google Trends 2020)

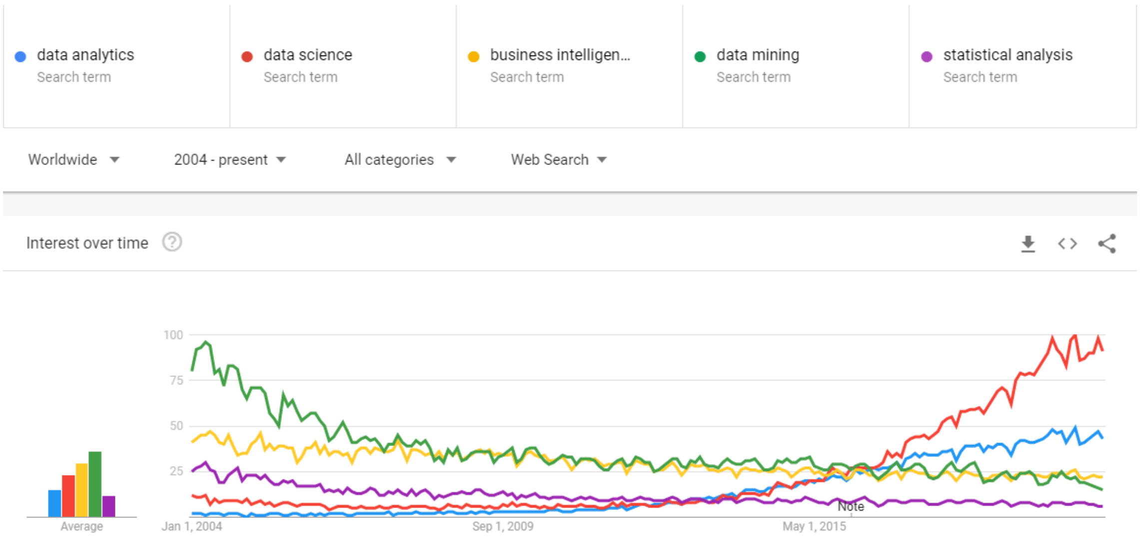

Figure 2. Search interest over time (Google Trends 2020)

Since the disciplines of data and analysis are wide and rather incoherent, this thesis will use the term "analytics" refers to the use of data and related insights created with applied analytical disciplines discussed earlier to drive fact-based decision-making, planning, management, execution, and learning. Davenport and Harris (2017) also support this view and define analytics as "The extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions". This means that analytics is not synonym with technology. Like mentioned before analytical capabilities include three main areas: organization, human, and technology (Davenport & Harris 2017)

2.3 Data and analytics capabilities

IT capabilities have significant impact on firm's performance and while resources can easily be copied, an unique set of capabilities assembled by a firm is not easy to imitate and this will generate sustained competitive advantages (Santhanam and Hartono 2003). These IT capabilities are defined as "firm's ability to mobilize and deploy IT-based resources in combination or co-present with other resources and capabilities" (Bharadwaj 2000). Davenport (2007) & McAfee et al. (2012) support this view and add that especially developing analytics and data capabilities can lead to competitive advantage.

When organization is capable to execute its processes well and have capability to manage development and maintenance organization-wide, the organization is mature. Organizations can have different types of maturities in different functions of a business. Maturities are usually presented in stages. Immature organizations can develop their capabilities and competencies step by step to reach the highest level of maturity. (Paulk et al. 1993)

This subchapter aims to answer first part of research question 1. What are the data and analytics capabilities?

2.3.1 Different types of data and analytics capabilities

Comprehensive analytic capabilities can be split in three main areas. These main areas are organizational capabilities, personnel related capabilities and technology related capabilities. (Davenport and Harris 2007) Cosic et al. (2012) support this view in their research but uses governance capabilities and analytics culture instead of organizational capabilities. Holsapple et al. (2014) conducted an extensive research that identified 21 key maturity items for business analytics. These then could be grouped in three distinct groups: integration and management support capabilities, process level ability to benefit from analytics, and technology and technical data analytics capabilities. (Holsapple et al. 2014) There are also other studies which divide the capabilities in their own way. For example, Shuradze & Wagner (2016) with their research where they categorized the analytics capabilities into infrastructure capabilities, personnel expertise, and relationship infrastructure. However, most of the research uses the same base made by Davenport and Harris (2007) with slight modifications. Therefore, this thesis will use the same split as Davenport and Harris (2007). All the identified capabilities and maturity items are discussed under these three main categories.

Organizational capabilities

To ensure successful analytics integration and significant impact on business performance organization should be able to translate analytical insights into their performance drivers by driving costs, profitability, growth, and shareholder value. Analytical insights by themselves are not generally very useful if they are not put into action. (Davenport and Harris 2007)

Translating insights into performance drivers is intertwined with organizations ability to execute strategy and manage performance. The organization must be able to convert its strategy into business objectives and align metrics with these business objectives such deep strategic insights in higher levels of analytics maturity, or simpler market and customers insights in lower levels of analytics maturity. Also, the organization should be able to focus one or two areas in their data and analytics strategy and to build their understanding progressively over time, learning from each new experiment and analysis. (Davenport and Harris 2007) In addition to usual business strategy, companies should have also robust data strategy for organizing, governing, analyzing, and deploying organizations information assets (DalleMule and Davenport 2017)

Ability to execute strategy also includes firm's capability to modify its business processes. To fully benefit from analytics and leverage data, organizations need to fully integrate analytics systems and other data infrastructure to their business processes. (Shuradze and Wagner 2016). Davenport and Harris (2007) support this view and list the company's ability to redesign processes as one of the key aspects of organizational analytics capabilities. This ability to modify firm's processes include ability to restructure business processes, restructure IT processes, and organizations ability adopt analytics applications (Shuradze and Wagner 2016)

Analytics management capability refers to management's ability to handle routines in a structured manner rather than ad hoc to manage IT and analytics resources in line with business needs and priorities (Wamba et al. 2017). Governance is frequently used concept to refer to all the activities and decision-appropriation mechanisms related to IT resources (Mikalef et al. 2018). The lack of governance has been recognized to be one of the main reasons to failures in leveraging data efficiently (Posavec and Krajnović 2016). Governance in data related context means firms capability to create networks internally and externally. This includes structured governance (assigning responsibilities, planning, and leading), relational governance (conflict resolution, business partnerships, and idea exchanges), and procedural governance (cost control, resource allocation, and guiding behavior through value analysis). (Mikalef et al. 2018) In addition Gupta and George

(2016) add that one organizational capability is ability to correctly estimate length and cost of analytics projects. This helps the projects to achieve their goals. (Gupta and George 2016)

A data-driven culture has been found to be significantly affecting use of analytics and driving the integration of analytics (Cao and Duan 2014). This helps building firms ability to leverage competitive advantage from analytical insights by promoting data-driven decision making instead of managerial experience or intuition. Data-driven culture has been noted to increase the success rate and continuation of data projects. Data-driven culture also promotes cross-organizational collaboration in data and analytics related matters thus enabling better analytics-generated insights. There is also less siloed data in data-driven organizations. (Mikalef et al. 2018) Survey made by Kiron et al. (2014) highlights this data-driven culture to be one of the key components of overall analytics capabilities. In data-driven culture, analytics changes the way business is conducted and causes a power shift in the organization. Data is seen as a core asset and more investments are done in analytics technology, talent acquisition and training. Analytics best practices and collaborative use of data is promoted across the company lines. (Kiron et al. 2014)

To successfully utilize firms IT-resources cooperation and interaction between business and IT functions is needed (Bassellier et al. 2001). Firms ability to inter-functionally coordinate activities and employees' social capital has been noted to be one aspect of overall analytics capabilities. The social capital aspect includes relations, respect, and trust between employees from IT and business departments, and common language and understandable communications between functions. Inter-functional coordination means abilities, in analytical capabilities context, such as: joint-coordination of business and analytics functions, use of cross-functional teams, information share between departments, sharing goals and priorities, and top managements promotion of coordination between the IT/analytics and business. (Shuradze and Wagner 2016)

One noted capability in organizations performing highly in analytics, is ability to empower all the employees in analytics. Employees are trained to use simple analysis and software and have access to appropriate information and resources. Capability to support independent analytics to support decision making is one metric of analytics maturity. (SAS 2016)

Human and personnel related capabilities

One the most important capability of analytical competitor is its human resources. It is noted also that one of the biggest difficulties when executing the analytical strategies is the lack of right kind analytical people. Software and hardware by themselves cannot create the capabilities that analytical strategies need. Full adaptation of analytics needs some analytical skills on every level. From executives to frontline operative employees, every has part in making analytical competition successful. (Davenport and Harris 2007) Mikalef et al. (2018) supports this view and adds that overall analytics capability and capability to utilize analytics technologies and tools is vastly dependent on employees' skills and knowledge. These skills can be then divided into business analytics knowledge, technical skills, business knowledge and relational knowledge and skills. To acquire all these skills businesses should hire employees focused on different skills sets. For example, big data engineers and data architects usually accommodates more technical skill sets than business analyst. However, at least knowing the basics of other skills is necessary for every role for engaging analytics integration. (Mikalef et al. 2018) Table 2. further explains these skills.

Analytics knowledge and technical skills are required to build analytics capabilities (Bock 2008) These technical skills include at least the following abilities: programming and software development skills, management of project life cycles, data and network management and maintenance, capability in distributed processing or computing, and abilities with analysis models and methods (eg. statistical data analysis, data mining, data visualization, etc.). (Shuradze and Wagner 2016)

Analytics practitioners and specialists also require domain knowledge to be able to ask the relevant questions. This domain knowledge includes understanding the company's business policies in very high level, recognizing of the business problem where the analytics application is being developed, and understanding the business context and the markets of the business. Failure in understanding the domain, the analytics application development might not meet with the end user needs. Therefore, personnel expertise with two types of skills (technical and domain) is enabler of analytics capabilities. (Shuradze and Wagner 2016)

Managerial skills are also identified as important factor for building holistic analytics capabilities. Intelligence gained from analytics will have little use if managers fail to see the potential from these insights. Analytics managers should have ability to understand the current needs and predict future needs of different business units, customers, and other stakeholders. Good collaboration and relationships between analytics managers and function managers provides possibilities for development of analytics capabilities and thus leading to competitive advantages. (Mikalef et al. 2018)

Table 2. shows different skills mentioned in literature to build data and analytics capabilities. The table highlights the main categories and skills falling under these categories.

Table 2. Human skills and knowledge as part of analytics capabilities

Human skills and knowledge	**Characteristics**	**Source**
Technical knowledge	Programming languages Technical infrastructure management Big data infrastructure knowledge MapReduce Unstructured data management Data collection/integration Project management skills Distributed computing IT systems knowledge	(Shuradze and Wagner 2016), (Mikalef et al. 2018)
Business knowledge	Business strategy KPIs Business processes Change management IT and analytics personnel's high-level domain knowledge	(Shuradze and Wagner 2016), (Mikalef et al. 2018)
Analytics skills	Statistical analysis Forecasting Query and analysis (SQL) Predictive modeling Optimization Model management	(Kiron et al. 2014), (Shuradze and Wagner 2016), (Mikalef et al. 2018)

	Simulation and scenario development Business reporting/KPIs/dashboards Web analytics Social media analytics Interactive data visualization Text, audio, video analytics Data and text mining	
Managerial and relational skills	Communication skills Team building Analytics managers understand business the needs of other functional managers, suppliers, and customers Analytics managers anticipate the future business needs of other functional managers, suppliers, and customers Analytics managers understand and evaluate the output from analytics Collaboration between analytics managers and functional managers, suppliers, and customers Coordination of analytics in ways that they support other functional managers, suppliers, and customers	(Mikalef et al. 2018), (Gupta and George 2016)

Technical and technology related capabilities

Developing tangible data resources and ability to leverage from them is fundamental basis for building organizations overall analytics capabilities (Mikalef et al. 2018). Research made by Mikalef et al. (2018) categorizes these tangible data resources into three main aspects. First one is the data itself. Data quality is viewed as a key feature to build competitive advantage with analytics. The firm must have ability to ensure high quality data but also its availability, integrity, and security. High quality of data means that it is accurate, timely, reliable, and complete (Kiron et al. 2014) Capability to effectively capture data, clean data, integrate data, and visualize it is part of data quality capabilities (Mikalef et al. 2018). Ramakrishnan et al. (2012) support this view and note that for analytical success data quality and consistency are critical factors for analytical success.

Second aspect of data related resources to build overall analytics capabilities is the firm's process related infrastructure to store, share and analyze the data. Especially the scalability and connectivity are noted be important because the amount of data and its use cases increase rapidly. Other qualities of good analytics infrastructure are compatibility, modularity, agility, reliability, adaptability, integration, and accessibility. It has been also stated that this not major issue when building analytics capabilities since the technology itself has surpassed beyond the needs of analytics. (Shuradze and Wagner 2016) Infrastructure affects the range and reach of business opportunities available to firms. Thus, it is reasonable to include infrastructure as one of the aspects of data analytics capabilities. (Kiron et al. 2014) This data and analytics infrastructure denotes the firm's ability to provide the correct data for managers required for making important business decisions, ability to deliver customer insights for customer-facing employees to help them to drive sales and efficiency, and other capabilities to share data across functional silos or business units. (Mikalef et al. 2018)

Third aspect of resources and building block for analytics capabilities is the information systems (IS) and software for conducting all the data related activities. These include for example IS for managing and storing the data, processing and analyzing data, visualizing data, and systems for data security and risk management services. (Shuradze and Wagner 2016)

Research conducted by Shuradze and Wagner (2016) sees the infrastructure and the IT-tools as one entity since this infrastructure usually relies on commercial technologies like data warehousing and data base management systems, and Extract-Transform-Load tools. This infrastructure also denotes firm's analytical ability as a tools such IT-systems for ad-hoc queries, data visualization, forecasting trends, and statistical analyses. (Wamba et al. 2017) s.5057 Analytics infrastructure consists applications, hardware, data, and networks (Cosic et al. 2012)

Chen and Nath (2018) conducted research on business analytics maturity of firms. The research identified that there are also firm's overall capabilities to benefit from analytics in certain functions. These capabilities are firm's ability to enhance market trend identifying, ability to enhance business performance assessments, ability to enhance customer need anticipation, and organizations ability to enhance operational efficiency with analytical tools and practises. (Chen and Nath 2018)

Study conducted by SAS also remarks that organizations which understand their costs, and can deliver and proof analytics ROI, are able to perform better regarding analytics integration. (SAS 2016)

Synthesis of analytics capabilities

When building holistic data and analytics capability, organizations must consider many different aspects. Table 3. gathers all the discussed factors from the literature and clarifies the characteristics linked to each factor.

Table 3. Factors of data and analytics capability of an organization

Capability	Characteristics	Sources
Organizational capabilities	Ability to translate analytical insights into performance drivers Ability to execute strategy and manage performance Ability to redesign processes and adopt analytics applications Ability to handle analytics routines in structured manner Ability to govern analytics activities Ability to appropriately budget and schedule analytics projects Data strategy Data-driven culture Cooperation between analytics and functional organizations Analytical empowerment of employees Integration of analytics into process improvement and reengineering Capability to enhance market identification with analytics Capability to enhance business performance assessment with analytics	(Davenport and Harris 2007), (Shuradze and Wagner 2016), (Wamba et al. 2017), (Mikalef et al. 2018), (Gupta and George 2016), (Cao and Duan 2014), (Kiron et al. 2014), (DalleMule and Davenport 2017), (Chen and Nath 2018), (SAS 2016)

	Capability to enhance customer needs anticipation with analytics Capability to enhance operational efficiency with analytics Capability to enhance customer relationships with analytics Cost analysis and ability to proof analytics ROI	
Human and personnel capabilities	Analytics knowledge and skills Technical knowledge and skills Business knowledge Managerial and relational skills	(Davenport and Harris 2007), (Mikalef et al. 2018), (Shuradze and Wagner 2016), (Gupta and George 2016)
Technology and technology related capabilities	High data quality Ability to effectively capture, clean, integrate and visualize data Process related infrastructure to store, share and analyze data Technical infrastructure and information systems to store, share and analyze data	(Mikalef et al. 2018), (Kiron et al. 2014), (Shuradze and Wagner 2016), (Wamba et al. 2017),

2.4 Data and analytics maturity models

Due the complexity wide range of analytics identifying and reviewing analytics capabilities is not an easy task. To address this many domains, use maturity models (Tarhan et al. 2016). In addition, Chen and Nath (2018) empirically proved that analytics maturity has significant positive impact on overall analytics success. Maturity models are instruments used for assessing the level of development of organisational capabilities, processes, or resources (Cosic et al. 2012). Widely used Capability Maturity Model CMM (Paulk et al. 1993) was developed to guide organizations by assessing current process maturity and identifying the most critical qualities and process improvements. The model provides a roadmap for continuous improvement by determining evolutionary path that increases organizations maturity in stages. Maturity models help organizations to set goals for capabilities and track development of these capabilities. (Paulk et al. 1993)

Maturity can be assessed descriptively, prescriptively, and comparatively depending on the model. Descriptive model can be used to assess the as-is maturity in the organization. Prescriptive maturity model also includes guidelines for improving the maturity at each level. (Becker et al. 2009a) A prescriptive model that has been already used in several organizations can then be used with its historical data for comparative purposes (Cosic et al. 2012)

There are three distinct main types of maturity models: staged, continuous, and contextual. In staged model such as Capability Maturity Model (Paulk et al. 1993) each stage builds on the previous stage and the stages are characterised by set of criteria that must be met to fulfil that particular level of maturity. Continuous model is comparable to staged model, but the different factors of each level may develop at different rates. Contextual maturity models are like continuous maturity models expect the development of maturity can be nonlinear and different components can move forwards and backwards. This relates more closely to reality, but it is more complex. (Cosic et al. 2012)

There has been presented many different maturity models for data and analytics maturities in the literature. This subchapter will discuss different types of maturity models aiming to build a foundation for research question 3. and for the synthesis of maturity models presented later in this thesis.

The models were retrieved from digital libraries and due the rapidly developing subject only relatively new models were chosen. Due the scarcity of academic literary on the subject also commercial models were chosen for the review.

2.4.1 Characteristics of different stages of organizational analytics maturity

Organizations which have identified their analytics integration degree and maturity are better prepared to turn challenges into opportunities LaValle et al. (2011). LaValle et al. (2011) splits organizations by their analytics adoption into three groups. Group with lowest adoption rate is called aspirational organizations. These organizations are mainly focusing on automating current processes and have only few necessary components to incorporate and act on analytical insights. In the middle there is experienced organizations. These organizations are looking behind the cost management and are aiming for revenue growth with analytics. Experienced organizations are effectively acting on analytic insights. Most sophisticated organizations according to LaValle et al. (2011) are transformed organizations. These organizations use analytics to prescribe actions and use analytics as competitive differentiator. In this kind of organizations analytics guide the future strategies but also daily operations. The survey also notes that transformed organizations are three times more likely to substantially outperform their industry peers than the aspirational organizations. (LaValle et al. 2011) LaValle identifies 6 key areas where different companies in different stages of analytics maturity differentiate. These are motive, functional proficiency, business challenges, key obstacles, data management, and analytics in use.

Survey conducted by Kiron et al. (2013) uses similar kind of categorization of three distinct groups of organizations separated by their analytics adaptation. These groups are analytically challenged, analytical practitioners, and analytical innovators. The main areas where these groups differentiate according to the survey are the following:

- Adaptation of data-driven culture. Analytically challenged organizations rely mainly on experience than data analysis when making decisions. Analytical practitioners have identified the benefits from integrating data-driven culture on some level and have begun to develop this culture. Analytical innovators have integrated the data-driven culture and share belief that data is core asset and data can be used to improve operations, customer service, strategy, and marketing.
- Uses of data and analytics. Analytical challenged organizations focus on cost reduction with analytics. Analytical practitioners are using analytics to guide day-to-day operational work but not to drive innovation and change business. Analytical innovators are using strategic insights (eg. identifying target customers, improving customer experience, and establishing strategy etc.) much more and are driving innovation with analytics.
- Data quality. Analytically challenged organizations might have insufficient amounts of data, suffer from poor data quality, and have access issues. Analytical practitioners have made significant advances in these areas and are able to make use of their data resources but there is still room for improvement in terms of data proficiency. Analytical innovators see data as a core asset and place high value on data as an organization.

- Knowledge and skills of employees. Analytically challenged organizations have lack of appropriate skills related to utilizing analytics efficiently. Analytical practitioners have some knowledge and skills related capabilities but also problems with fragmented analytics ecosystem which slows down the integration of analytics. Analytical innovators have higher levels of data management, analytical knowledge, and other related skills.

(Kiron et al. 2013) s.4-17 & (Kiron et al. 2014) s.7

Analytics maturity indicators for different levels of analytics maturity

To drive integration and use of analytics applications Lismont et al. (2017) made quantitative study and clustered organizations by their analytics maturity stage. Based on these clusters they identified indicators of analytics maturity and provided growth path for companies in different stages of analytics maturity. The clustering of analytics maturity was based on surveyed analytics characteristics of different organizations. The characteristics of analytics were based on DELTA model (Davenport et al. 2010). (Lismont et al. 2017)

Four recognized cluster and their features were:

1. No analytics. Usually small companies (median of 10 employees) operating in local markets.
2. Analytics bootstrappers. Companies with relatively low application of analytics techniques and focus on online analytical processing and basic segmentation. Low application of HR analytics but common use of marketing, finance, and operations analytics. Moderate use of basic analytical techniques such as decisions trees and linear regression. Data quality is mostly not governed. Decision making is mostly based on intuition and there is lack of in-house analytical skills.
3. Sustainable analytics adopters. More common use of finance, marketing, and operations analytics. High adoption rate of basic analytical techniques but uncommon application of complex analytical techniques such as neural networks and survival analysis. Data governance is still an issue, but decision-making is less impacted by intuition.

4. Disruptive analytics innovators. This cluster has very high adoption rates of analytics in finance, marketing, and operations. HR analytics is practiced in high levels. There is high use of both simple and complex analytics techniques to drive analytical insights. However, there are issues with privacy, standardization, and documentation in data management. The key strength in this cluster is culture which embraces data and analytics, empowering the organization disrupt their strategic business processes.

(Lismont et al. 2017)

This research also noted that companies in different stages of analytics maturity conduct analytics in different types of teams or organizations. Less advanced companies conduct analytics more project-based and more analytical mature companies practice it more departmentally or organization wide. (Lismont et al. 2017)

2.4.2 Different analytics maturity models

Business Analytics Capability Maturity Model BACMM

The first steps towards analytics maturity models were presented Cosic et al. (2012) in their research paper. Need for this kind of model was justified with earlier recognised positive impact of analytics integration to business performance. Model identifies low-level business analytics capabilities which can be then assessed independently. (Cosic et al. 2012) The identified low-level capabilities can be seen in figure 3. This model does not specify the stages of independent capabilities and acknowledges this as a future research topic. Instead the model uses high-level general maturity scale to assess the maturity of capabilities. This is also descriptive model which does not give suggestions how to proceed towards higher levels of maturity. (Cosic et al. 2012) The five-level scale is initially defined as follows:

- Level 0 – Non-existent: the organization does not possess the capability
- Level 1 – Initial: the capability exits but is not well developed
- Level 2 – Intermediate: the capability is well developed but there is a lot of development possibilities

- Level 3 – Advanced: the capability is very well developed but the is little room for development
- Level 4 – Optimized: the capability is highly developed, and it is difficult to find more development possibilities. At this point the capability can be considered as fully mature.

(Cosic et al. 2012)

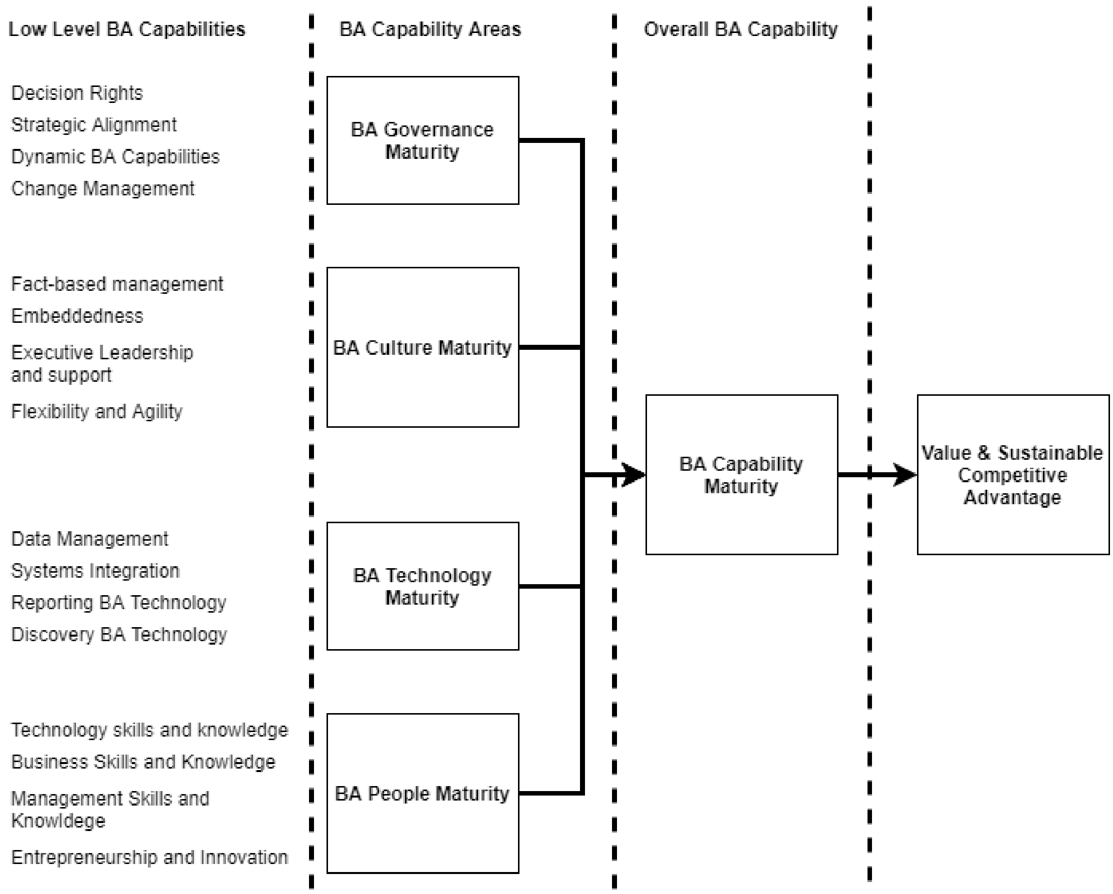

Figure 3. Business Analytics Capability Maturity Model (modified from Cosic et al. 2012)

DELTA+ model

Davenport and Harris (2017) developed an analytics maturity model to provide guidance for creating a roadmap for building organizations analytics capabilities and to reach competitiveness through analytics. This DELTA model highlights that significant changes in the organization must be made to achieve competitiveness through analytics. The DELTA stands for the main dimensions of analytics maturity. These are:

- Data. This denotes the raw material for analytics. High quality, diverse, and dynamic data is necessary for gaining precise insights from analytics. Data is seen as strategic asset and it must be managed to maximize its value for the organization.
- Enterprise. Organizations performing highly on analytics, manage and coordinate data and analytics related capabilities and resources on enterprise level, across the functions.
- Leadership. Leadership and committed leaders are main drivers for analytics integration and success. The importance of analytics is understood, and analytics is constantly utilized for data-driven decision making. Innovation, exploration of data, and continuous development is endorsed by the executives.
- Targets. The finite resources and capabilities are managed and coordinated to reach carefully specified targets to gain maximum benefits from analytics. These targets can be for example cutting costs, optimizing processes, improving customer satisfaction, increasing profitability, or scaling the business.
- Analysts. In addition to hiring a couple of talented analytical employees, much more is needed to build analytically mature organization. Analysts and data scientist are needed to build and maintain the analytical models, but data-savvy executives and decision makers are needed to oversee and benefit from analytics initiatives. Use of analytics should be organization wide.

(Davenport and Harris 2017)

In addition to these five dimensions of analytics Davenport and Harris (2017) added two dimensions more, based on their newer research and to answer the needs of big data and arrival of variety of new techniques. This new model is called DELTA+. These two new dimensions are:

- Technology. Robust and well-integrated technical architecture (data, software, processing power, and tools) is required for enable efficient analytics. The lack of this architecture may cause locked data in organizational silos and overlapping analytical work.

- Analytical Techniques. Techniques for analyzing the data come from wider variety of disciplines ranging from simple descriptive statistics to neural networks and genetic algorithm. Capability to choose the right technique for different occasions and ability to utilize that technique is basis for analytically mature organization.

(Davenport and Harris 2017)

DELTA+ model consists of five stages of analytics maturity. The model identifies that when progressing from the early stages, a detour might have to be taken. If there is no top management support, it must be gained by for example arousing interest with building successful use cases of analytics inside of smaller department. (Davenport and Harris 2017) Figure 4. shows the different stages of DELTA+ model and the possible detour to gain top management support.

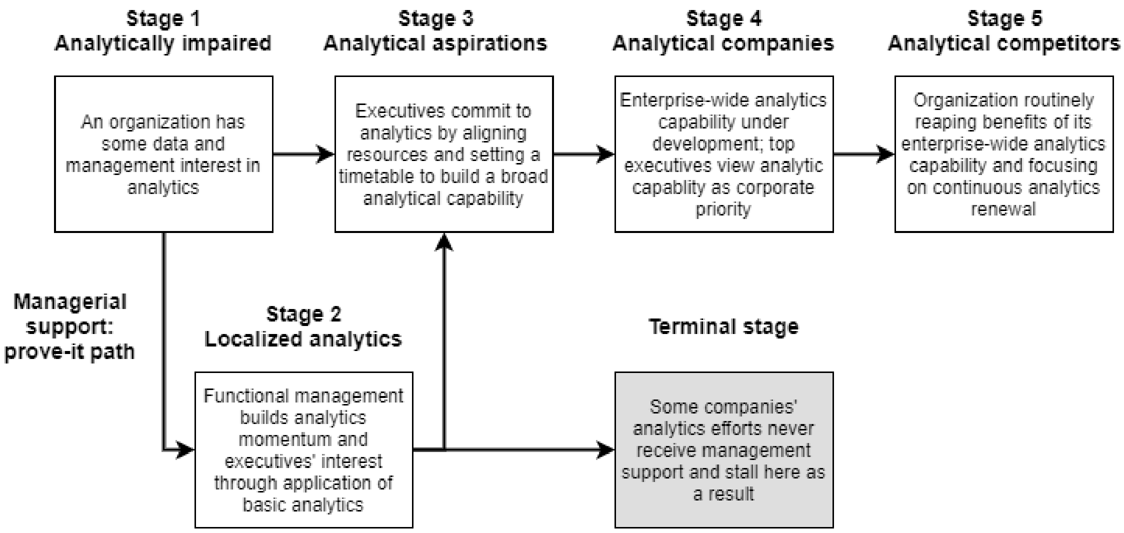

Figure 4. Road map to becoming an analytical competitor (modified from Davenport and Harris 2017)

PharmaVOICE & SAS Analytics Maturity Scorecard

To respond to the growing competition in the life sciences industry PharmaVOICE developed Analytic Maturity Scorecard together with SAS to help organizations to evaluate their analytics maturity and drive their competitiveness though analytics. Understanding where the organization is regarding every area of analytics maturity and setting the goals for these areas is the starting point for development of analytics. (PharmaVOICE 2014)

The scorecard contains five levels of maturity on each area of analytics maturity. The areas of analytics maturity are the following:

- Culture. This denotes the decision makers use of data and analysis. Ranging from analytically unaware to explorative where decision makers search actively new ways to use advanced analytics to support business decisions.
- Internal Process Readiness. In lower levels there is no defined analytic processes or data management processes. On the top level is continuously self-refining processes to data enhancement and analytic methods to optimize resources.
- Analytical Capabilities. This means the skill and capability to use analytical methods from simple reporting to advanced new techniques.
- Data Environment. Includes infrastructure and software for analytics. In the lower levels data projects are disorganized, overlapped, and the used software is consistent across the organization. In the higher levels of maturity, the projects are aligned to overall strategy and documented. There is also continuous improvement to support the most difficult business challenges.

(PharmaVOICE 2014)

IDC's Big Data and Analytics MaturityScape

A framework developed by International Data Corporation helps organizations to assess their Big Data and Analytics (BDA) competency, enable dialog across organization about goals and actions of BDA initiatives, and help define the short- and long-term goals for all areas of BDA maturity. (Vesset et al. 2015)

The framework is split into 5 distinct stages by how organizations conduct analytics as process. Also, the business outcomes or the goals of analytics are explained in each stage. The stages of analytics maturity and their business outcomes are:

1. Ad Hoc. Organizations conduct analytics in ad hoc manner. Often these are unbudgeted proof-of-concept pilots with no defined business case or goal. Value of analytics is concentrated in organization pockets with limited business outcomes. The main outcome is to provide decision makers with access to information. This can mean simple reporting, use of query, dashboard, or simply exposing the data itself to end user.
2. Opportunistic. In this stage the organization has learned lessons from the earlier analytics pilots and apply them business cases with project-specific budgets. However, resource allocation and project management are inefficient because lack of common analytics strategies. There are also problems with data quality and available technology. There might be lack of necessary skills and cross-organizational coordination. The primary goal in this stage is focusing the analysis part of whole from data to decision making process. This can cause problems without proper data management and preparation.
3. Repeatable. Organization in the repeatable stage conduct recurring, budgeted, and funded analytics projects to support business. The projects are documented and there are good project management practises in place. There should be business-unit level data and analytics strategy. Cost benefit analysis for analytics initiatives is not conducted in process-oriented manner. There is lack of good governance and security practices for data. Providing comprehensive insights based on varied data from internal and external sources is the main outcome of this stage.
4. Managed. In this stage the organizations have achieved cross-organizational BDA strategy and BDA program standards. There is enterprise wide budget for analytics and upper management support. Data guides actions in all levels of organizations. Data and technologies are monitored and tuned when necessary. There might be centralized technology group for BDA, but the analytics skills are still mostly decentralized. Primary goal in this stage is to produce actionable insights to all levels of organization. BDA is used to answer what happened and why it did.

5. Optimized. To reach this level organizations must have coordinated and continuous BDA improvement process. The BDA strategy is documented and accepted enterprise wide. There is budget for analytics operations but also ad hoc budget for unforeseen opportunities. Data quality is high, and it can be trusted. Wide range of software tools is utilized appropriately. There are all the necessary skills for data collection, management, analysis, dissemination, and management of BDA activities. There is also high level of automation in analytics for scalability. Data is seen as core asset and enabler for products and services. The business outcomes for analytically optimized organizations are providing foresights to all decision makers and to relevant external stakeholders. Analytics are integrated into business processes resulting predictive capabilities to capitalize on new opportunities and mitigate threats.

(Vesset et al. 2015)

The framework examines analytics maturity through dimensions of vision, people, process, technology, and data. These main dimensions have sub-dimensions to assess more detailed capabilities in different stages of maturity. This model also notes the effect of these dimension in different stages of BDA maturity. Vesset et al. (2015) argues that focusing on certain aspect is different stages helps the organization to move forward with analytical maturity. (Vesset et al. 2015) These dimensions and where they support organizations to move forward with the analytics maturity can be seen in figure 5.

CSC Big Data Maturity was developed based on IDC's Big Data and Analytics MaturityScape. This is a web platform for conducting the assessment. The survey questions are based on the main capability dimension as in the original framework but there is a feature to compare your results to industry average. (IDC 2020)

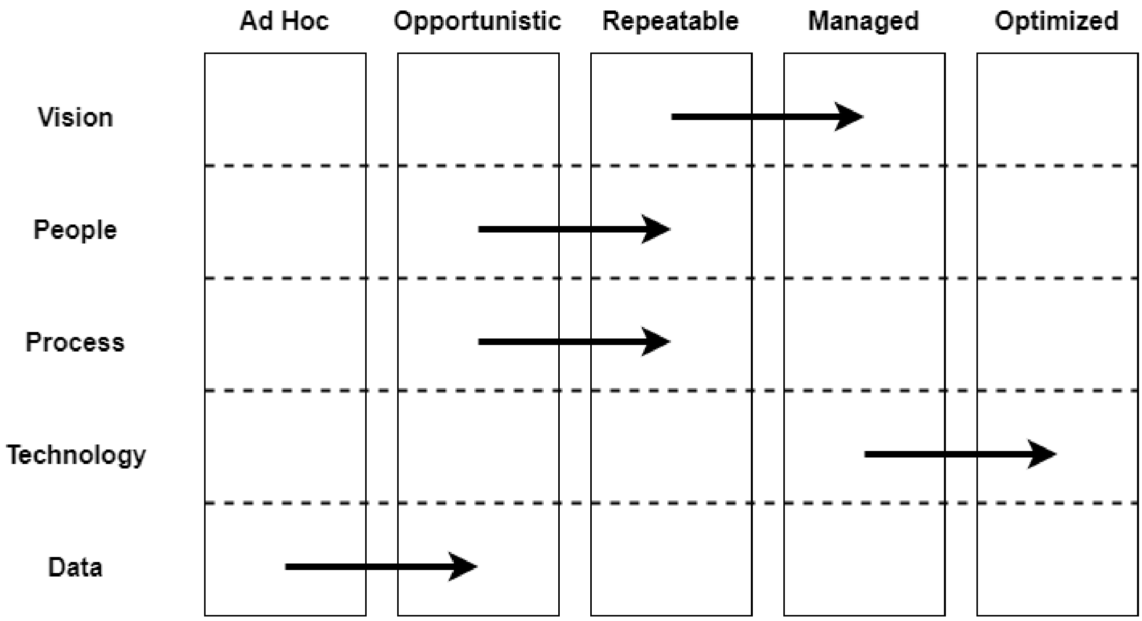

Figure 5. IDC's Big Data and Analytics MaturityScape with dimensions of analytics capabilities (modified from Vesset et al. 2015)

Analytical Processes Maturity Model (APMM)

Building analytical models is relatively new practice and there are only few common methodologies for establishing these models. To answer challenges generated from lack of common analytics methodologies Grossman (2018) has generated a framework to understand how capable an organization is in building analytical models that are:

- Statistically valid and completed in schedule.
- Able to be deployed into organizations operations, services, or products.
- Meeting the organization's goals for the model.

(Grossman 2018)

The stages of the APMM are categorized by organization's analytics process maturity. There are five levels and they are the following:

1. Build reports. In the beginning of analytics journey organization might be able to build reports and analyze data on very low level.

2. Build models. In the second level organizations can build, validate, and deploy models based on data.
3. Repeatable analytics. In this stage the organization has built system to repeatable build, deploy, and update analytic models. This process usually requires a efficient analytics governance.
4. Enterprise analytics. Analytics are used organization-wide and build and deployed with common infrastructure whenever possible. Outputs of the different models are integrated together to support the targets of the organization as whole. Analytics across the firm are coordinated from single governance structure.
5. Strategy-driven analytics. In the last stage organization has clear analytic strategy which is aligned with overall strategy. Analytic strategy is used to select the analytic opportunities and develop analytics to support the overall vision of the enterprise.

(Grossman 2018)

The model is descriptive but offers the main targets for the analytics key process areas. There is no guidance how to the organization should proceed in certain stages of maturity. These main targets give guidelines how the organizations should develop these key process areas. These key process areas are the following:

- Ability to build analytics models.
- Ability to deploy analytics models.
- Ability to manage analytic infrastructure.
- Ability to operate analytic governance structure.
- Ability to provide security and compliance for analytic assets.
- Ability to develop an analytics strategy.

(Grossman 2018)

TDWI Big Data Maturity Model

TDWI created an analytics maturity model framework for rising needs of big data and to guide organizations in different stages of analytics maturity. The purpose of the framework is to support the organizations to identify and define the goals around big data analytics and help to communicate that vision to the entire organization. It also serves as a tool to measure and track the progress of big data analytics adoption within the company. Firms can expect more value from their investments when they progress through the stages. Figure 6. presents these stages in visual form. (Halper and Krishnan 2013)

This model divides maturity into 5 distinct stages. These stages are the following:

1. Nascent. In this stage organizations might be unaware of big data and its value. The executive leadership is not currently supporting the development effort although there might be some scattered interest in big data in the organization. Some technical applications for analytics might have been taken into use but not fully integrated and coherently operated.
2. Pre-adoption. Big data and analytics have caught some interest in the organization and the organization is learning about the subject. Planning of implementation of big data applications is usually led by IT department rather than business.
3. Early adoption. At this point there might be one or two proof of concepts which are being integrated to production. Some executive sponsorship is being committed to analytics. Infrastructure and data management practices are being built.

   This stage takes usually relatively long time because of the chasm. This chasm means time-consuming obstacles which slows down organizations data efforts. These obstacles can be for example political issues about data ownership and analytics vision, or lack of correct skill set for advanced analytics. This usually happens when the organization starts to achieve big benefits and business transformation from analytics.
4. Corporate adoption. After the chasm has been crossed the end users typically get more involved, gain more insights from the analytics, and the business is being transformed by analytics and big data. There is strong understanding that analytics is a competitive differentiator. Data and analytics are seen as core value for innovation.
5. Mature/visionary. When the organization has achieved a level where the analytics and big data projects are organized and executed smoothly and effectively the

organization is analytically mature. All the key elements of big data analytics are highly tuned, and the culture embraces analytics.

(Halper and Krishnan 2013)

The TDWI maturity model identifies five key dimensions of big data maturity. These dimensions are the following:

- Organization. How much strategy, culture, leadership, and funding support analytics in the organization.
- Infrastructure: To what extend does the infrastructure and architecture support the analytics initiates. What are the technologies in place to support analytics?
- Data management. How well the data quality is ensured? How is the data managed?
- Analytics. The level of technical skills and knowledge of analytics in the company. Ability to deliver analytics applications.
- Governance: Coherence of organizations data governance strategy to support of its big data analytics.

(Halper and Krishnan 2013)

TDWI has also created a web platform for assessing organizations big data maturity and to benchmark maturity results. The survey is based on TDWI Big Data Maturity Model. (TDWI 2020)

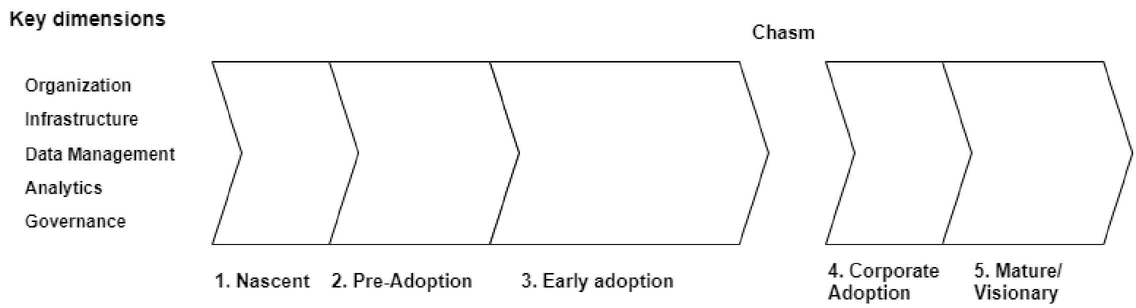

Figure 6. Stages of maturity in the TDWI Big Data Maturity Model with key dimensions (modified from Halper and Krishnan 2013)

INFORMS Analytics Maturity Model

INFORMS uses online platform to evaluate analytics maturity of organizations. The maturity assessment contains three main sections. Since the results from the assessment can be compared to industry averages, the model is comparative maturity model. These are the following:

- Analytics capability. Does the organization have the services, methods, and models to perform analytics?
- Organizational capability. Does the organization have culture and practices to enable effective use of analytics?
- Data and infrastructure. Is the organizations data usable and sufficient for appropriate analytics?

(Burciaga 2013, INFORMS 2020)

The model differentiates from the usual maturity models by having 10 levels of maturity but three main stages of maturity. The main stages are beginning, developing, and advanced. The organization can then gradually develop the capabilities inside of the main stages before jumping to the next stage. Higher resolution in assessing the maturity level makes the model more generalizable. (Burciaga 2013, INFORMS 2020)

Industrial Analytics Maturity Model (IAMM)

In recent years manufacturing has become increasingly more data-intensive and there has been recognized many benefits (for example operational efficiency, process innovation, environmental impact, strategic improvements etc.) from utilizing this growing amount of data and insights generated from it. However, there has been challenges to identify areas for improvement and challenges to measure current analytics capabilities. O’Donovan et al. (2016) created a multi-dimensional maturity model to help with these problems and for assessing industrial analytics capabilities. This model considers the characteristics of industrial domain and is considered to be manufacturing domain specific. (O’Donovan et al. 2016)

The model inspects analytics maturity through five main dimensions and their subdimensions. From these five main dimensions are:

- Open standards. Technologies and protocols based on standards. Promotes interoperability between stages of analytics lifecycle.
- Operation technology. Technology to support the acquisition and processing of data.
- Information technology. Infrastructure to share, store, and transmits the data.
- Data analytics. Knowledge, skills, and overall capability to build and deploy analytical models.
- Embedded analytics. Capability to embed analytics applications to operations and drive real-time decision-making.

(O'Donovan et al. 2016)

IAMM uses three-staged scale to assess maturity of every sub-dimension. The stages are nonexistent, partial, and fully existent. (O'Donovan et al. 2016)

Maturity model for big data analytics in airline network planning

Hausladen & Schosser (2020) developed a big data readiness maturity model to address major organizational and strategic challenges of newly available big data for airline network planners. Traditionally logistics and especially airlines have invested in collecting, processing, and analyzing data. However, there is still gap of analytics utilization in network planning. This model aims to address this issue. The research acknowledges that the model is highly specialized in network planning and management. (Hausladen and Schosser 2020)

The six-staged model inspects analytics maturity through the four following main domains:

- Strategic alignment. Considers formulation of specific big data strategy, strategic alignment of business and IT functions, availability of resources for strategy execution but also culture and "level of change readiness".

- Organization. This domain denotes organizations structure, roles, and responsibilities regarding analytics initiatives. Also, employees' skills and knowledge are under this main domain.
- Data. Data quality and processes of data management.
- Information technology. IT structure to integrate data sources and tools to analyze it. Also, includes the used IT tools and their capabilities to support analysis.

(Hausladen and Schosser 2020)

2.4.3 Summary of data and analytics maturity models

There are many different analytics maturity models from many different perspectives. Most of the models describe the same subjects only with bit different focus or point of view on the subject. However, to gain holistic understanding about all the aspects of analytics maturity it is useful examine all the models and their components. Table 4. shows a summary of the examined models with purpose of the model, key dimensions of the model, and the number of maturity levels in said model. Repeating key dimensions in these models are strategy, organizational capabilities, ability to benefit from analytics, people and culture, data quality and management, and technical infrastructure.

Table 4. Summary of maturity models

Maturity model, source, and publication year	Purpose of the model	Key dimensions	Number of maturity levels
TDWI Big Data Maturity Model (Halper and Krishnan 2013)	Answer the needs of big data and analytics. Identify and define goals, measure the progress, and communicate analytics vision.	Organization, Infrastructure, Data management, analytics, Governance	5 + chasm between stages 3. and 4.
Business Analytics Capability Maturity Model BACMM (Cosic et al. 2012)	Early work for analytics maturity models. Identifies BA capability areas, low-level capabilities, and maturity levels.	Governance, Culture, Technology, People	5

PharmaVOICE & SAS Analytics Maturity Scorecard (PharmaVOICE 2014)	Helps life science industries to evaluate their analytics maturity and drive competitiveness.	Culture, Internal Process Readiness, Analytical Capabilities, Data Environment	5
IDC's Big Data and Analytics MaturityScape (Vesset et al. 2015)	Helps organizations to assess BDA competency, enables dialog, and helps with defining short- and long-term goals of analytics. Explains focus areas in different stages of analytics maturity.	Vision, People. Process, Technology, Data	5
DELTA+ Model (Davenport and Harris 2017)	Provides guidance for creating a roadmap for building organizations analytics capabilities.	Data, Enterprise, Leadership, Targets, Analysts, Technology, Analytical techniques	4+1 (prove-it detour)
Analytical Processes Maturity Model (Grossman 2018)	Helps to understand how capable organization is to build statistically valid models, deploy the models, and meet the goals for the models.	Building models, Deploying models, Analytical infrastructure, Analytical governance, Data security, Analytics strategy	5
INFORMS Analytical Maturity Model (INFORMS 2020)	Online platform to evaluate analytics maturity.	Analytics capability, Organization, Data and Infrastructure	Three main stages and 10 levels. 1-3 Low, 4-7 Medium, 8-10 High
Industrial Analytics Maturity Model (IAMM) (O'Donovan et al. 2016)	Assess analytics capabilities in manufacturing domain.	Open standards, Operation technology, Data analytics, Embedded analytics	3
Maturity Model for BDA in airline network planning (Hausladen and Schosser 2020)	Address major organizational and strategic challenges of newly available big data for airline network planners.	Strategic alignment, Organization, Data, Information technology	6

2.5 Success factors for data and analytics adaptation

Since developing the analytics maturity is complex matter and there are many chokepoints to slow down the development (Vidgen et al. 2017), it is important to understand these factors that might slow down the development, or the other way around, accelerate, the progression of analytics maturity.

Chen and Nath (2018) noted that a good foundation for analytics development is firms and especially its leader's positive view of IT and its benefits. If the IT is seen as a strategic capability with significant effect on firm's performance, adaptation of analytics is also easier. The same study also confirmed that there is a correlation between perceived benefits of IT and success of different analytics maturity factors which then have a correlation with analytics success. (Chen and Nath 2018)

However, when speaking of analytics not only the IT and data aspects are necessary for analytical success. Skilled analysts and strategic positioning are also needed. Firms need to invest in state-of-the-art tools, quality data and data-savvy people who understand not only the relevant technologies but also the business side. (Grover et al. 2018)

A survey conducted by Kiron et al. (2014) showed that analytically developed companies were drastically more likely to have been investing into analytics technologies and analytics-related human resources in the past 12 months and were planning to make investments in the next 12 months than less analytically developed competitors. The same survey also displayed that there is stronger pressure from the senior management to become data-driven in companies that are more advanced in analytical maturity. (Kiron et al. 2014)

To achieve benefits for the business and the end users of the analytics, the analytics initiatives and development should always be made with clear business goals. Pursuit of higher analytics maturity is not the end goal. It is only necessary for reaching the business goals where true value lies. High level acuity of analytics is the only way to reach the full potential of analytics and success with analytics. (Ali et al. 2018) Research made by Chen & Nath (2018) support this and states gaining process level benefits as one of the main essential success factors for analytics. Value assessment helps to achieve more success on analytics development and initiatives. However, calculating the benefits of analytics is very difficult since analytics affect and is affected by many different factors. (Grover et al. 2018)

Chokepoints slow down the development and should be avoided when possible. To do this it is important to understand what these chokepoints are. Davenport and Harris (2017) s.185-186 state that the main challenge of organization analytics development is acquiring and deploying the needed human and financial resources. They also list at least the following as a factor that could slow down the progression:

- Too much focus on only one dimension of analytic capabilities. Or investing too much or little in any analytic capability compared with demand.
- Collecting data without plans to use it.
- Attempting to do everything at once.
- Not prioritizing the analytics initiatives based on business value.
- Automating decision-based applications without monitoring the outcomes and other maintenance.
- Not fully understanding the problem when developing analytical solution for it.

To sustain competitive advantage with analytics, the company must assess, renew and develop the analytical capabilities continuously. Firms that successfully compete on analytics have analytics capabilities that are:

- Hard to duplicate. IT applications and other resources are easy to copy but analytics culture and processes which bring value are difficult to copy.
- Unique. The capabilities are built for the company and can’t be directly used in other businesses.
- Adaptable to many situations. Organization is capable to apply the analytics in different and changing situations.
- Better than competitors. The companies use analytics wider and, in more detail, than their competitors.
- Renewable. Analytics are under continuous development and renewal to create more value.

Davenport (2017)

2.5.1 Four fundamental success factors for competing on analytics

Davenport & Harris (2017) note that they have found out during their studies four different common key characteristics for analytical success for competing on analytics. First one is support of a strategic, distinctive capability. This means that the analytic efforts of the organization should be focused on the primary strategic capability. Though not every firm has this kind of distinct main strategic capability. For example, primary focus at Netflix is on predicting customer viewing preferences and at Walmart the main effort is on supply chain analytics. After some analytical maturity has been achieved the organizations analytical efforts should spread to other functions as well. (Davenport and Harris 2017)

Second key characteristic is an enterprise-level approach to and management of analytics. This denotes that the analytics development and activities should be managed, governed, and guided from a unit that covers the whole enterprise. However, there should be analytical capability and use of analytics inside of departments, but these should be governed centrally. This allows the organization to achieve the analytical goals set for the enterprise, but also promotes the principle of singe sourced truth and decreases the amount of overlapping analytical work. (Davenport and Harris 2017, Hausladen and Schosser 2020)

Third main factor for analytical success is commitment of senior management. Davenport & Harris (2017) mention that during their extensive researches they did not find a single company that could compete on analytics and did not had commitment and broad support from the executive level. It is almost impossible to make necessary cultural changes to truly adopt analytics without drive from CEO or from another C-level executive. (Davenport and Harris 2017) Factor analysis for analytics maturity conducted by Chen & Nath (2018) supported this and listed analytics integration and management support as a one underlying aspect for analytics maturity.

Fourth key success factor from Davenport and Harris research (2017) is large-scale ambition. One common factor for analytically competitive firms is their success with analytics-based strategies. To truly achieve competitive advantage from analytics, the scope and scale of the analytical targets should be large enough to disrupt the business. Incremental, tactical use of analytics will result in minor improvements. Strategic, competitive use of analytics will result major advantages. (Davenport and Harris 2017)

These four characteristics from Davenport & Harris studies (2017) are tightly intertwined and all four are needed for reaching a level where your organization can compete on analytics. However, these factors also support each other. If executive leadership is committed to analytical development and has built a analytics-based strategy around organizations main strategic capability, it is likely that the organization will then adopt enterprise-wide approach and the results from the analytics will reflect on the strategic orientation. (Davenport and Harris 2017)

2.5.2 Short-term and long-term planning as success factors for analytics

Vesset et al. (2015) argue that one important factor for gaining success in analytics and maximizing the value potential of analytics is good planning and setting short-term and long-term goals for analytics development. Their study recommends planning actions for three different time periods. First period is now. Organization should make actions to develop their analytics capabilities as soon as possible. Second planning period is the next budget cycle. The last planning period is the next three to five years.

For starting analytics development Vesset et al. (2015) recommends that the organization should assess their business and analytics situation “as is” as soon as possible. The organization should recognize all the relevant technologies available and identify the already had analytical capabilities. Organization should identify opportunities to use existing data, technology, and analytics in new ways. Experimentation and creating proof-of-concepts is important part of this. (Vesset et al. 2015, Davenport & Harris 2017) Building data infrastructure is one of the first main steps for starting the development of analytics (Arunachalam et al. 2018).

For the next budget cycle Vesset et al. (2015) recommend that organizations should aim to make quantifiable wins to demonstrate the benefits of analytics and justify budget allocations. The organization’s analytics capabilities should be assessed continuously.

Chokepoints of development should be recognized and actions regarding those should be planned accordingly. Expand technical architecture and develop governance capabilities.

In the next three to five years Vesset et al. (2015) suggest that organizations should aim to achieve high level governance and performance management of analytics. Also, the organization should ensure that experimentation and discovery use cases are supported with appropriate technology, data, processes, staff, and funding. Organization should be able to re-engineer its business processes in response to new insights from analytics. Resources are balanced and prioritized across all dimensions of analytics capabilities. The capabilities are systematically reviewed and adjusted to match evolving requirements.

2.6 Synthesis of the literature review

This chapter synthesizes the theory foundation from previous chapters and merges processed information as holistic analytics capability maturity model. The construction of the maturity model follows roughly an approach to build maturity models proposed in *Developing Maturity Models for IT Management* (Becker et al. 2009b). Firstly, the problem was defined in earlier chapters. Then a comparison between existing models was conducted and now in this chapter the model is built. Later, the model is reviewed through empirical interviews and necessary iterations are implemented.

The model is built by taking all the relevant capabilities and maturity items from existing models and adding any missing capabilities addressed in chapter 2.3. *Data and analytics capabilities.* Then the maturity stages are determined based on the existing models and the description of the different stages are populated with combination of description of stages from existing models.

Building of the maturity model aims to follow widely used design-science research guidelines defined by Hefner et al. (2004). These guidelines are the following:

1. Design as an artifact. Output of the research should be an artifact in the form of feasible construct, method, or model.
2. Problem relevance. Objective is to produce solution for important and relevant business problem.

3. Design evaluation. The quality and usefulness of the model must be tested with evaluation methods.
4. Research contributions. Research should provide clear contribution in the areas of design artifact, foundation, and/or design methodologies.
5. Research rigor. Research relies on various methods to build and evaluate the artifact.
6. Design as a search process. solutions must be iteratively proposed, refined, evaluated, and, if necessary, enhanced.
7. Communication of research. Research must be presented effectively to technology-oriented but also management-oriented viewers.

(Hevner et al. 2004)

2.6.1 Key Dimensions and maturity levels for synthesized framework

Existing maturity models use different key dimensions but there are common subjects being handled behind these different dimensions. Five dimensions were chosen for synthesized model. These five dimensions were chosen by their recurrence in the existing literature and to build holistic view of analytics maturity. Table 5. shows how different maturity models relate their analytic capabilities and maturity items to these chosen dimensions. Since *Maturity model for big data analytics in airline network planning* (Hausladen and Schosser 2020) is targeted for airline network planning, only generalizable parts of the model were used. IAMM (O'Donovan et al. 2016) was left out completely since the model was too domain specific.

Table 5. Key dimensions for synthesized framework and related sub-dimensions from existing maturity models

Chosen key dimension	Related sub-dimension from existing maturity models	Maturity Model
Strategy	Analytics strategy	TDWI Big Data Maturity Model, IDC MaturityScape, Analytical Processes Maturity Model
	Alignment of analytics initiatives to the business strategy	BACMM, Analytics Maturity Scorecard,

	Targets	DELTA+
Organization	Budgeting	TDWI Big Data Maturity Model, IDC MaturityScape
	Organizations' change management capability	BACMM
	Justification of projects	IDC MaturityScape
	Performance Management	IDC MaturityScape, INFORMS Maturity Model, DELTA+
	Collaboration	IDC MaturityScape
	Governance Groups	IDC MaturityScape, Analytical Processes Maturity Model
	Governance and policies	TDWI Big Data Maturity Model, INFORMS Maturity Model, DELTA+, Maturity Model for BDA in airline network planning
	Analytical organization	IDC MaturityScape
Analytics	Techniques	TDWI Big Data Maturity Model, INFORMS Maturity Model, DELTA+
	Overall analytics process	Analytics Maturity Scorecard, IDC MaturityScape, INFORMS Maturity Model
	Model building process	Analytical Processes Maturity Model
	Model deploying process	Analytical Processes Maturity Model
Data Management	Standards of data	TDWI Big Data Maturity Model, Maturity Model for BDA in airline network planning
	Transparency on data requirements	Maturity Model for BDA in airline network planning
	Process	TDWI Big Data Maturity Model, BACMM, Analytics Maturity Scorecard, IDC MaturityScape, IDC MaturityScape, INFORMS Maturity Model, Maturity Model for BDA in airline network planning
	Overall quality	TDWI Big Data Maturity Model, IDC MaturityScape, INFORMS Maturity Model, DELTA+

	Access	TDWI Big Data Maturity Model, INFORMS Maturity Model, Maturity Model for BDA in airline network planning
	Security and privacy	TDWI Big Data Maturity Model, Analytical Processes Maturity Model
	Data completeness, and variety	TDWI Big Data Maturity Model, IDC MaturityScape, Maturity Model for BDA in airline network planning
	Trust	IDC MaturityScape
	Timeliness	IDC MaturityScape, Maturity Model for BDA in airline network planning
	Data ownership and traceability of used data	INFORMS Maturity Model, Maturity Model for BDA in airline network planning
	Availability of external sources	Maturity Model for BDA in airline network planning
	Transparency on available data	Maturity Model for BDA in airline network planning
People and culture	Executive support	TDWI Big Data Maturity Model, BACMM, Analytics Maturity Scorecard, IDC MaturityScape, INFORMS Maturity Model, DELTA+, Maturity Model for BDA in airline network planning
	Perceived value of analytics	TDWI Big Data Maturity Model, Analytics Maturity Scorecard, INFORMS Maturity Model, Maturity Model for BDA in airline network planning
	Formal technical and analytical skills	TDWI Big Data Maturity Model, BACMM, Analytics Maturity Scorecard, IDC MaturityScape, INFORMS Maturity Model, DELTA+, Maturity Model for BDA in airline network planning
	Analytical skills of non-analytical employees.	Maturity Model for BDA in airline network planning
	Mindset and attitude towards analytics	TDWI Big Data Maturity Model, BACMM, IDC MaturityScape,

		INFORMS Maturity Model, Maturity Model for BDA in airline network planning
	Fact-based management	Analytics Maturity Scorecard, IDC MaturityScape
	Domain knowledge and business skills of analytics specialists	BACMM
	Management skills for analytics initiatives and projects	BACMM
	Training	IDC MaturityScape, DELTA+, Maturity Model for BDA in airline network planning
Technology	Development of infrastructure	TDWI Big Data Maturity Model, IDC MaturityScape, Analytical Processes Maturity Model, Maturity Model for BDA in airline network planning
	Technologies	TDWI Big Data Maturity Model, BACMM, INFORMS Maturity Model, DELTA+
	Architecture and deployment of technologies.	TDWI Big Data Maturity Model, BACMM, Analytics Maturity Scorecard, IDC MaturityScape, Analytical Processes Maturity Model, INFORMS Maturity Model, DELTA+
	Performance of technologies	IDC MaturityScape
	Functionality of technologies	IDC MaturityScape
	Flexibility to add new data sources	Maturity Model for BDA in airline network planning

To create holistic understanding what are the capabilities and maturity items of analytics, the framework will also address analytics capabilities examined in chapter 2.3.1 *Different types of data and analytics capabilities*. Only the capabilities which are not already included in earlier maturity models were integrated. These capabilities were:

- Organizations capability to redesign and integrate new processes

- Ability to appropriately budget and schedule analytics projects
- Cost analysis and ability to proof analytics ROI
- Integration of analytics into process improvement and reengineering
- Capability to enhance market identification with analytics
- Capability to enhance business performance assessment with analytics
- Capability to enhance customer needs anticipation with analytics
- Capability to enhance operational efficiency with analytics
- Capability to enhance customer relationships with analytics

Five distinct stages were chosen for the model since it was the most frequently occurring amount among the examined models.

The model is populated with stages from earlier maturity models. Generalized levels from BACMM (Cosic et al. 2012) is used when there are no previously defined stages for a certain capability.

The synthesized model with modifications based on the empirical interviews can be found in Appendix 1.

3 EMPIRICAL STUDY

This section describes the empirical study that was conducted to gain comprehensive understanding on the themes of the thesis. Insights from analytics practitioners are used to fulfil and support earlier findings from the literary. Research methodology, interviewees, and analysis of the interview are also presented in this chapter.

3.1 Materials and methods

3.1.1 Data collection and interviews

In all sciences major advances in understanding usually require experimental and observational data. One way to gather this data is interviewing. (Weller and Romney 1988) Therefore, this thesis will also use interviewing in addition to literature study.

Qualitative half structured theme interview was used since the field of study is relatively less studied and formulation of exact and precise questions is difficult and theme interview leaves room for clarifying questions. In addition, half structured explorative study may lead to findings that the researchers did not anticipate before the interview.

The set goal for the interviews was "to gain valuable information from practitioners and researchers to, complement earlier findings from the literature or identify completely new findings not discussed in the existing literature, and validate the composed analytical maturity framework".

The interviewees were selected by their knowledge in the field of analytics. The positions of the interviewees were mostly head of data & analytics, Chief Data Officer (CDO), data business designer, data scientist, or similar positions related management of data and analytics. All the interviewees had long history in the field, and most had over 10 years of experience of working with data and analytics. Aim was to gather interviewees from four distinct types of organizations. These types were small and medium enterprise (SME), large companies, analytics consultant companies and research organizations. Organizing the interviews was relatively easy and interviewees seemed to be really interested about the topic. Table 6. show information about the interviewees and their organization.

Table 6. The interviews held for this thesis

#	Representative title	Type of the company	Industry
1.	Business Excellence Manager	SME	Safety products and services
2.	CDO	SME	E-commerce
3.	Head of data and analytics	Large company	Transportation
4.	Head of Data Analytics & AI in Advisory	Consulting company	Auditing, consulting
5.	Head of data and analytics	Large company	Vehicle sales
6.	Data Business Designer	Consulting company	IT, consulting
7.	Senior Scientist, Technical Manager	Research organization	Product development, research
8.	Research Director	Research organization	Research
9.	Business Director	Consulting company	IT, consulting
10.	Head of Data	Large company	Transportation

Interviews were mostly conducted via teleconferencing software. Typically, one interview took 40-60 minutes, and the interview was recorded. Interviewees had the opportunity to familiarize themselves with the interview template and the built analytical maturity model before the actual interview. The interview started with introduction to the subject, type of the interview, and the terminology that would be used in the interview. After that, the questions were handled in systematic order. Before reviewing the analytical maturity model with the interviewees, short introduction to models' goals and background was held. Finally, the interviewees were asked to comment any other issues regarding the subjects discussed and the interview itself. All the interviews were held in august 2020.

3.1.2 Structure of the interviews

The interview protocol was built with principles defined by Susan et al. (1988) in their book *Systematic Data Collection*. First principle is to clarify the domain that is being inspected and that interviewees speak with same terminology. Second principle is to try not to ask close ended questions that can be answered with yes or no. This may lead to more productive answers. Third principle is to strive to formulate simple and easily understandable questions. (Weller and Romney 1988)

First principle was handled in the questionnaire with clarifying the terminology about analytics and by asking the interviewees what kind of terminology they use, so that the interviewer can use terms familiar to them. Second and third principle were dealt with formation of the questions.

Questionnaire was formed based on the earlier made literature review. Firstly, the goal and the targets for the interview were set and questions were formed to reach the set targets. In the first part of the questionnaire, the basic information is collected (i.e. name, role, experience, business function and its size). The second part of the questionnaire handles interviewees view on the on data and analytics capabilities. The third part examines the development of analytics capabilities in the interviewee’s organization. Finally, the analytics maturity model is assessed by the interviewee. The questionnaire contained in total 16 questions (appendix 2.).

4 RESULTS AND DISCUSSION

This master's thesis is a qualitative research consisting a literary review and empirical study to build a holistic understanding about the research topic. The *literature review chapter* built a theorical foundation for the study and reviews the research questions through the previous research. In the *empirical study* chapter, the used empirical study methods and the interviewees were presented, and the research questions were assessed with interviews to obtain practical insights. The goal was also to validate the synthesized framework for analyzing data and analytics capabilities and other findings from the literature. After this the results from the literature review and from the empirical study were presented. In addition, analyses of the results were conducted this chapter. Also, a review of the differences between results from the literature and from the empirical studies was held.

In this chapter these findings and analyses are composed to answer the research questions and brief conversation about the results is held. Finally, in the last chapter *Conclusion*, summary of the research is composed, and a generalizability and the reliability of the study is assessed. In addition, theoretical contribution and managerial implications are given in this chapter. Finally, future studies about the research topic is discussed.

4.1 Interview results

4.1.1 Data and analytics capabilities

Since the terminology in data related activities is rather incoherent and unclear used terminology was discussed first in the interviews to clarify any misunderstandings. At the same time understanding of used terminology in the field was built. "Data and analytics" or "analytics" were used mostly as an umbrella term to describe all the activities related to gathering and using the data. There were also organizations which used "business intelligence", "business analytics", "value creation from data" or did not use any umbrella term. There was some discussion that "AI" as a term was used sometimes to describe also less advanced analytics techniques.

All the interviewees saw data and analytics as important factor for future development, and most of the interviewed organizations had identified some analytics capabilities. Systematic identification was mostly linked to existence of data, analytics, or AI strategy or development plan. Otherwise the identification of the capabilities was mostly informal and tacit knowledge. Organizations mainly understood analytics capabilities as wide and complex matter including for example organizational issues and culture. However, technology consultancies and research organizations linked the analytics capabilities more heavily into technical skills and technologies. Overall, the recognition of the analytics capabilities was seen as important for overall analytics development. During the interview, there was recognized two capabilities that did not occur in the studied literature these were:

- Capability to maintain analytics applications. As analytical applications are launched into production there is still a need to maintain these applications and the data that these applications use. This capability was brought up in many interviews and it was seen as a slowing factor in overall development. Planning for production and automation of repetitive tasks related to the maintenance was said to help with this.
- Capability to identify relevant use cases for analytics. This also includes the ability to recognize and measure the value of the use cases. To achieve success with analytics it is important to recognize the analytical opportunities with real business value.

4.1.2 Developing data and analytics capabilities

All the interviewed companies were developing or planning to develop analytics capabilities. However, systematicity of the development varied a lot between the organizations. Some of the companies had strictly governed paths for analytical development and most had data, analytics, or AI strategy or had linked the development of analytics heavily into their business strategy, but couple of the organizations did not have any formal plan how to develop these capabilities. Overall, the interviewed organizations estimated their analytical capabilities to be a bit better than the average in their industry.

Frequently appearing target for overall analytics development was ability to execute analytics in every level of organization in a relevant way and data-driven culture. Usage of real-time data was mentioned to be one goal. Consultancies emphasized that their capabilities must meet the needs of the market and development targets are set based on the said needs.

During the interviews there was mentioned many success factors for developing analytics capabilities. Brought up factors were the following:

- Measurement of ROI and other value from the analytics initiatives. Analytics initiatives are more easily sold internally and justified when there is concrete proof from the benefits. This also helps to prioritize the analytics initiatives. One high level advisor from a consultancy firm said “it is important that we do not only develop capabilities. We need to simultaneously and quickly bring up successful use cases and justification that there is business value. These are vital for internal sales and getting more investments, and approval that it is beneficial to develop these capabilities.”
- Critical mass in an analytical organization or team. Team has enough technical and business knowledge to solve the emerging problems efficiently.
- Iterative building of analytical capabilities. Since analytics maturity is a long journey it should be build a piece by piece.
- Communication and data awareness. Data awareness has been noted to lessen the usual problems regarding the change resistance and communication is need for this.
- Quality of training material from technology partners. Material for learning new technical skills varies a lot. Material from big and mature organizations has been noted to be better quality and this way supporting the faster development of technical skills.
- Commitment of the C-level executives and allocated budget. Gaining enough resources from significant development of analytics is difficult if there is no allocated budget for it.
- Software development practices. Agile working has been noted to be beneficial also for data and analytics initiatives.

There was also discussion about chokepoints of analytics development. The following subjects were mentioned to slow down the development of overall organizational analytics capability:

- Availability of key personnel. Collaboration between analytics and other functions requires time from time from employees from these other functions. Problem is that they are usually very busy and not easily accessible, and analytics initiatives take long time to complete.
- Availability of data and analytics workforce. There is higher demand than supply for analytics professional which causes problems to recruit and keep skilled employees.
- Gap between analytics know-how and business. One interviewee said "often the analytics organization might be left disconnected from the rest of the organization. Therefore, the development of the culture is also important. If the business does not understand the benefits of analytics, does not speak the same terminology, and analytical and business knowledge are not integrated, the collaboration is really difficult and business value from analytics is rarely achieved."
- Data management, quality, and availability. There are still basic data quality issues in many data sources, and it takes time to fix these.
- Data privacy issues. Modern data privacy legislation such as GDPR cause difficulties when analyzing data containing personal information.
- Priority of analytics initiatives. Lot of resources are wasted on analytics initiatives which do not have much business value and take lot of effort. One interviewee claimed that in his opinion there is enough technical know-how in Finland to execute all the relevant business cases, but the priority of the initiatives is very poor. He also added that this is because of general lack understanding of analytics and how much effort analytics initiatives take.

Analytics organization structure was also frequently occurring topic during the interviews. Many of the interviewees identified three main types of how analytics should be organized: central analytics organization, widespread organization with limited or no governance, and a hybrid organization where more technical aspects were managed centrally (e.g. data and infrastructure) but analytical processing capability was spread to different functions. There was no conclusion which one would be best, and the different type structures suited well different types of companies.

4.1.3 Reviewing the developed maturity model

All the interviewed representatives saw maturity models and built model as useful tool for development of analytics maturity. Analytically most developed companies had already used similar analytics maturity assessment tool. Generally, the main dimensions of the model were seen important and one representative said "it is good that the model has the organizational, strategy, culture and people aspects as well. Analytics is not only use of technologies."

The main uses for the analytics maturity models were the following:

- Charting the present state.
- Setting a goal for development.
- Planning the development roadmap or strategy.
- Tracking the development.
- Communicating the goals and areas of development.

These uses for maturity models were brought up by almost every interviewee. One consulting advisor added that maturity models are good way to explain and educate broad and complex concepts to individuals who are not familiar with the subject.

Even though the maturity model was seen useful, it was seen also as long and complex. It was advised that the some of the models subdimensions should be merged and there should be a summary which only includes 10-20 most important dimensions.

There was also discussion about feasibility of the model in certain organizations. The model was generally seen useful for big organizations but suitability for small and low hierarchical organization was questioned. Some of the subdimensions were not relevant to organizations where there are only small amount employees, and the organizations analytical capabilities are held by only few employees. Also, there was comments that the model did not suit research and public organizations very well. However, most of the dimensions in the model were seen important for every type of organization.

In addition to the interview, the interviewee's reviewed every subdimension of the built model and estimated its relevancy and importance for their organization. The final modifications and adjustments of the model are done based on these reviews. Averages of the importance scores were then calculated for every sub dimensions. Final model can be seen in appendix 1.

4.2 Synthesis of the interviews and the results

The findings from the interviews followed mostly the same topics that the earlier literature review had already discussed. However, there was also some completely new findings that had not been examined in the literature that this thesis covered. Also, one noteworthy finding was that there was no coherent terminology or language for the subject. Terms like AI, analytics, data science, data, etc. were used to describe same phenomena.

Two new analytics development success factors were found in the interviews. These was proofing the value of the analytics initiatives to justify the development and use of software development practises in data and analytics context. Also, two slowing factors was highlighted, the gap between business and analytics, and low analytics awareness.

From the interviews there was identified couple novel analytics capabilities that were not discussed in the literature. These were capability to maintain analytics applications in operation and capability to identify new applications for data and analytics. These capabilities will be used in the final version of the organizational analytic maturity model. The built organizational analytics maturity model was seen generally very useful, but it was seen cumbersome. Modifications to the model is done based on these comments. The final model can be seen in Appendix 1.

There was identified some differences between organization type and the development of analytics maturity. Smaller organizations did not see the analytics maturity model as important as larger organizations. Also, smaller organizations rated budgeting and governance related capabilities less important.

4.3 Key findings

Firstly, the thesis explored the concepts and definitions of data, analytics, and analytics capabilities answering the research question 1. The different definitions and to some extent similar concepts indicate that there are no clear de facto definitions and concepts agreed among the researchers and practitioners. Different disciplines had only slight differences and seemed to describe same phenomena only from bit different perspective. For example, Larson and Chang (2016) argued that business intelligence is the umbrella term of data and analytics related activities, whereas for example Everitt and Skrondal (2010) claimed that data science is the right term to use. Other sources such as Cao (2017) then state that business intelligence and data science are completely different disciplines.

However, main finding from subsection was clarification on vague terms such as big data and analytics. This was also somewhat necessary for clarifying what is being discussed later in this thesis. This notion was supported by the data from the interviews. Almost every organization had their own terms to discuss these topics. It was mentioned that bringing this topic up in the interviews was a good idea to clarify any misunderstandings.

Thesis also studied what is understood when talking about analytics capabilities. Variety of different capabilities were identified through the literary review and interviews. The study demonstrates that analytics capabilities are considered to be rather vast field of different distinct subjects, ranging from organizations collective technical skills to organizations culture and ability to change. Diverseness could be seen from the earlier studies. For example, Industrial Analytics Maturity Model made by O'Donovan et al. (2016) considered the analytics maturity from more technical point of view, whereas DELTA+ model made by Davenport & Harris (2017) examined analytics maturity form management point of view. This thesis combines all the discussed capabilities from the literature but also from empirical interviews.

Second research question was to understand what the factor acceleration are the development of organizational data and analytics capabilities. This was answered through examining prior studies and during empirical interviews. There were new success factors for developing analytical capabilities identified through empirical interviews.

To answer the final research question, this study used discovered analytics capabilities and already established analytics maturity models to synthesize a more complete maturity model for analyzing organizations' analytics capabilities and maturity. Maturity model was supplemented with comments from experts in the field. The model brings together all the mentioned dimensions and fulfills them based on earlier studies and the empirical study. The built maturity model is the most complete from all the models reviewed. Though this causes some limitations. This study is answering the need for updated maturity model especially since IT field is developing rapidly and there were no recent analytics maturity models published. Noteworthy mention was that during the empirical interviews all the interviewees saw the maturity model as a useful tool for mapping and developing organizations analytical capabilities.

The results of this study were mostly expected. However, it was interesting to notice that the terminology varied a lot in field but in also in the literature. This might be caused by different disciplines studying and developing somewhat same topics but not discussing together to agree universal terminology.

This study discusses analytics capabilities and analytics maturity as complex, wide, and holistic matter. The extreme extent of the subject was somewhat unexpected. Though, it is not unanticipated that when organizational change is in question, things get complicated since change in organizations is connected to people, processes, and systems.

It is noteworthy to mention that based on the empirical interviews, generally the development of analytics maturity is still in very early stages in Finnish organizations. The interviewed organizations were chosen by their expertise but none of them would be considered highly advanced in analytics maturity. It was noted that in average organization is barely collecting data, and not necessary using it in any way. More detailed research could be done based on this information.

4.4 A critical evaluation of the study

There are some limitations, which could be excellent topics for future studies, regarding the built analytics maturity model. The number of conducted interviews was quite narrow which may cause some biases to the results. Especially when all the interviewed organizations were from one country. All though many of them were international companies. Also, the commercial models that appreciated by the professionals were behind paywalls and thus inaccessible for the researcher and left out of the study. Second limitation was the heavy focus on literary review. This might cause some constraints in real-world usability which was one goal of the model. Also, the complexity of the model might also reduce its real-world applicability if the model is overfitted.

It is difficult to estimate of how well this study and the built maturity model answers the fundamental goal of developing the analytics maturity, which is eventually producing useful insights supporting decision making in every level of the organization. The risk is that activities focus too much on acquiring these analytics capabilities which are not valuable by themselves if not used to support the underlying goal to bring useful insights.

Like stated previously, this study was conducted in two parts, with literature review and empirical interviews. The findings from literature review can be considered as credible since this thesis used only sources that are sufficiently new to discuss the topic with the modern understanding, and most of the sources were peer-reviewed and published in respected journals. Used commercial maturity models were considered to be trustworthy by the practitioners. And most of the findings were verified through interviewing researchers and practitioners of the field. However, some of the sources were relatively old which may affect the reliability and validity of that specific citation.

5 CONCLUSION

The exploding amount of data and computing power, globalization, increased competition, and emerging new technologies are transforming the markets and businesses. Utilizing data and analytics can improve organizations competitiveness significantly. However, use of data and analytics and organizations analytical maturity is still very low. This is primarily because gaining success with analytics is very complex matter and analytics capabilities cannot be bought. It takes time to gradually develop these capabilities. This thesis discussed how these resources and capabilities can be managed, understood, and developed to pursue better utilization of these new technologies.

The research questions were formulated to scope the previous question to better examinable objects. Firstly, the basic concepts of the subjects were clarified. Then the analytical capabilities and analytical development success factors were discussed. Finally, based on these findings and previous studies, a maturity model for assessing organization's analytics capabilities was built. The findings were then verified and complemented with practical insights with empirical interviews.

Main findings of this thesis are the mapping of the terminology used to describe the subject, listed analytics capabilities and success factors of analytics development, and the formulated analytics maturity model for assessing organizations analytics maturity. These findings give an up-to-date view on these subjects. This is important because the discipline is relatively new, and the technologies develop rapidly. One notable conclusion from the empirical interviews was that the development of analytics maturity in most organizations has just began and is just taking its first steps. There is still lot of benefits to be gained from data and analytics.

From managerial point of view this thesis gives comprehensive overview about terminology about data, analytics, and other related disciplines. The built model is a good starting point for development of analytics. It can be used to assess the as-is situation which is the first step of development (Vesset el al.2015), to plan future development, and to communicate it through the organization. Also, the success factors of analytics development are surely interesting from managerial point of view.

Findings from the empirical interviews could be examined and researched more thoroughly since the findings were completely novel and the sampling group was rather small. This might be interesting topic for future research. Also, one could doubt the stated importance of data and analytics since all the interviewees were working with this topic and might have had a bias about perceived importance of the subject.

5.1.1 Future research

To counter some of the limitations of this study, the research could be continued. Empirical part of this study was conducted only on Finnish organizations or Finnish departments of global companies and on rather small sample size. It would be useful to conduct the same study with larger and more international sample.

Immaturity of an organization is linked to inefficient operations, extension of schedules and budgets in projects, and bad quality of products due to unrealistic expectations (Paulk et al. 1993) s.19. To help with this problem, it would be interesting to research could the built framework be used to estimate maturity of a supplier. Also, outsourcing some of the capabilities is a possibility. It would be interesting to study further which ones and to which extend.

Focus of this thesis was on data and analytics maturity and from which capabilities it was formed. It would be also interesting to study how and in which order organization should develop these capabilities. In other words, how to build your data and analytics maturity based on the knowledge from this thesis.

6 REFERENCES

Ali, A., Mancha, R., and Pachamanova, D., 2018. Correcting analytics maturity myopia. Business Horizons, 61 (2), 211–219.

Alpaydin, E., 2020. Introduction to machine learning. MIT press. pp. 1-4, 267-268

Arunachalam, D., Kumar, N., and Kawalek, J. P., 2018. Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. Transportation Research Part E: Logistics and Transportation Review, 114, 416–436.

Aydiner, A. S., Tatoglu, E., Bayraktar, E., Zaim, S., and Delen, D., 2019. Business analytics and firm performance: The mediating role of business process performance. Journal of business research, 96, 228–237. pp. 234

Ayodele, T. O., 2010. Types of machine learning algorithms. New advances in machine learning, 19–48. pp. 19

Bassellier, G., Reich, B. H., and Benbasat, I., 2001. Information technology competence of business managers: A definition and research model. Journal of management information systems, 17 (4), 159–182. pp. 178

Becker, J., Knackstedt, R., and Pöppelbuß, J., 2009a. Developing maturity models for IT management. Business & Information Systems Engineering, 1 (3), 213–222. pp. 216-221

Becker, J., Knackstedt, R., and Pöppelbuß, J., 2009b. Developing maturity models for IT management. Business & Information Systems Engineering, 1 (3), 213–222.

Bharadwaj, A. S., 2000. A resource-based perspective on information technology capability and firm performance: an empirical investigation. MIS quarterly, 169–196. pp. 171

Bock, M. E., 2008. Statistics: harnessing the power of information. Journal of the American Statistical Association, 103 (484), 1331–1333. pp. 1331

Breiman, L., 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16 (3), 199–231. pp. 199-215

Burciaga, A., 2013. INFORMS Analytics Maturity Model, INFORMS Annual Meeting October 2013 [online]. Available from: https://www.informs.org/Resource-Center/Video-Library/INFORMS-Meetings-Videos/INFORMS-Annual-Meeting-Presentations/2013-INFORMS-Annual-Meeting-Minneapolis/Analytics-Maturity-Model [Accessed 16 Jun 2020].

Cao, G. and Duan, Y., 2014. A path model linking business analytics, data-driven culture, and competitive advantage. pp. 13

Cao, L., 2017. Data science: a comprehensive overview. ACM Computing Surveys (CSUR), 50 (3), 1–42. pp. 31-32

Chen, H., Chiang, R. H. L., and Storey, V. C., 2012. Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165–1188.

Chen, L. and Nath, R., 2018. Business analytics maturity of firms: an examination of the relationships between managerial perception of IT, business analytics maturity and success. Information Systems Management, 35 (1), 62–77. pp. 73

Chevalier-Roignant, B. and Trigeorgis, L., 2011. Competitive strategy: options and games. MIT Press.

Codd, E. F., 2002. A relational model of data for large shared data banks. In: Software pioneers. Springer, 263–294.

Cosic, R., Shanks, G., and Maynard, S., 2012. Towards a business analytics capability maturity model. In: ACIS 2012: Location, Location, Location: Proceedings of the 23rd Australasian Conference on Information Systems 2012. ACIS, 1–11. pp. 3-5

DalleMule, L. and Davenport, T. H., 2017. What’s your data strategy. Harvard Business Review, 95 (3), 112–121. pp. 1-2

Davenport, T. H., 2018. The AI advantage : how to put the artificial intelligence revolution to work. London: The MIT Press. pp. 9-17

Davenport, T. H., Harris, J. G., and Morison, R., 2010. Analytics at work: Smarter decisions, better results. Harvard Business Press.

Davenport, T. H. and Prusak, L., 2000. Working knowledge: How organizations manage what they know. Ubiquity, 2000 (August), 6. pp. 2

Davenport, T. and Harris, J., 2007. Competing on Analytics: The New Science of Winning. Harvard Business School Press. pp. 28-29, 79-80, 115, 177-181, 316

Davenport, T. and Harris, J., 2017. Competing on analytics: Updated, with a new introduction: The new science of winning. Harvard Business Press. pp. 46-60 157-181, 222-244

Dhar, V., 2013. Data science and prediction. Communications of the ACM, 56 (12), 64–73. pp. 64

Efron, B. and Tibshirani, R., 1991. Statistical data analysis in the computer age. Science, 253 (5018), 390–395. pp. 390

Everitt, B. S. and Skrondal, A., 2010. The Cambridge Dictionary of Statistic Fourth Edition. Cambridge: Cambridge University Press. pp. 125

Fang, H., 2015. Managing data lakes in big data era: What's a data lake and why has it become popular in data management ecosystem. In: 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE, 820–824. pp. 1

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., 1996. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39 (11), 27–34. pp. 28-29

Fey, C. F. and Birkinshaw, J., 2005. External sources of knowledge, governance mode, and R&D performance. Journal of Management, 31 (4), 597–621.

Google Trends, 2020. Interest over time [online]. Google Trends. Available from: https://trends.google.com/trends/explore?date=all&q=data%20analytics,data%20science,business%20intelligence,data%20mining,statistical%20analysis [Accessed 12 Aug 2020].

Grossman, R. L., 2018. A framework for evaluating the analytic maturity of an organization. International Journal of Information Management, 38 (1), 45–51. pp. 45-50

Grover, V., Chiang, R. H. L., Liang, T.-P., and Zhang, D., 2018. Creating strategic business value from big data analytics: A research framework. Journal of Management Information Systems, 35 (2), 388–423. pp. 418-419

Gupta, M. and George, J. F., 2016. Toward the development of a big data analytics capability. Information & Management, 53 (8). pp. 38

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L., 2019. Multivariate data analysis. Edition Eight. Cengage Learning, EMEA. pp. 7-9

Halper, F. and Krishnan, K., 2013. TDWI big data maturity model guide interpreting your assessment score. TDWI Benchmark Guide, 2014, 2013. pp. 6-15

Han, J., Haihong, E., Le, G., and Du, J., 2011. Survey on NoSQL database. In: 2011 6th international conference on pervasive computing and applications. IEEE, 363–366. pp. 364

Han, J., Pei, J., and Kamber, M., 2011. Data mining: concepts and techniques. Elsevier. pp. 8-21, 327, 443

Hand, D. J., 2007. Principles of data mining. Drug safety, 30 (7), 621–622. pp. 621-622

Hausladen, I. and Schosser, M., 2020. Towards a maturity model for big data analytics in airline network planning. Journal of Air Transport Management, 82, 101721. pp. 51-54

Hevner, A. R., March, S. T., Park, J., and Ram, S., 2004. Design science in information systems research. MIS quarterly, 75–105.

Hilbert, M. and López, P., 2011. The world's technological capacity to store, communicate, and compute information. Science [online], 332 (6025), 60–65. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.1200970 [Accessed 13 May 2020]. pp. 64

Hogan, O., Holdgate, L., and Jayasuriya, R., 2016. The Value of Big Data and the Internet of Things to the UK Economy [online]. London. Available from: https://www.sas.com/en_ie/offers/16q1/cebr-big-data-internet-of-things.html [Accessed 16 May 2020]. pp. 32

Holsapple, C., Lee-Post, A., and Pakath, R., 2014. A unified foundation for business analytics. Decision Support Systems, 64, 130–141. pp. 133-134

Hong, T.-P., Kuo, C.-S., and Chi, S.-C., 1999. Mining association rules from quantitative data. Intelligent data analysis, 3 (5), 363–376. pp. 2

IDC, 2020. CSC Big Data Maturity Tool [online]. Available from: http://csc.bigdatamaturity.com/ [Accessed 18 Jun 2020].

INFORMS, 2020. INFORMS Analytics Maturity Model [online]. Available from: https://analyticsmaturity.informs.org/ [Accessed 16 Jun 2020].

Inmon, W. H., 2005. Building the data warehouse. John wiley & sons. pp. 13-16

Intel, 2017. Planning Guide: Getting Started with Advanced Analytics.

Jayawardene, V., Sadiq, S., and Indulska, M., 2015. An analysis of data quality dimensions. pp. 8

Khatri, V. and Brown, C. v, 2010. Designing data governance. Communications of the ACM, 53 (1), 148–152. pp. 149-150

Kiron, D., Ferguson, R. B., and Prentice, P. K., 2013. From value to vision: Reimagining the possible with data analytics. MIT Sloan Management Review, 54 (3), 1.

Kiron, D., Prentice, P. K., and Ferguson, R. B., 2014. The analytics mandate. MIT Sloan management review, 55 (4), 1. pp. 10-11, 20

Ladley, J., 2019. Data governance: How to design, deploy, and sustain an effective data governance program. Academic Press. pp. 11

Larson, D. and Chang, V., 2016. A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36 (5), 700–710. pp. 21

Laskowski, N., 2016. Data lake governance: A big data do or die. URL: http://searchcio. techtarget. com/feature/Data-lake-governance-A-big-data-do-or-die (access date 28/05/2016).

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., and Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. MIT sloan management review, 52 (2), 21–32. pp. 22-24

Lismont, J., Vanthienen, J., Baesens, B., and Lemahieu, W., 2017. Defining analytics maturity indicators: A survey approach. International Journal of Information Management, 37 (3), 114–124.

Marsland, S., 2015. Machine learning: an algorithmic perspective. CRC press. pp. 4-6

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., and Barton, D., 2012. Big data: the management revolution. Harvard business review, 90 (10), 60–68. pp. 4-9

Mikalef, P., Pappas, I. O., Krogstie, J., and Giannakos, M., 2018. Big data analytics capabilities: a systematic literature review and research agenda. Information Systems and e-Business Management, 16 (3), 547–578. pp. 561-568

Miloslavskaya, N. and Tolstoy, A., 2016. Big data, fast data and data lake concepts. Procedia Computer Science, 88 (300–305), 63. pp. 302

Negash, S. and Gray, P., 2008. Business intelligence. In: Handbook on decision support systems 2. Springer, 175–193. pp. 177-178

Newbold, P., Carlson, W. L., and Thorne, B., 2013. Statistics for business and economics. Pearson Boston, MA. pp. 22

O'Donovan, P., Bruton, K., and O'Sullivan, D. T. J., 2016. IAMM: A maturity model for measuring industrial analytics capabilities in large-scale manufacturing facilities. International Journal of Prognostics and Health Management, 7. pp. 1-10

Organisation for Economic Co-operation and Development, 2003. The OECD glossary of statistical terms. Organisation for Economic Co-operation and Development. pp. 154-170

Paulk, M. C., Curtis, B., Chrissis, M. B., and Weber, C. v, 1993. Capability maturity model, version 1.1. IEEE software, 10 (4), 18–27. pp. 18-22

Perkins, L., Redmond, E., and Wilson, J., 2018. Seven databases in seven weeks: a guide to modern databases and the NoSQL movement. Pragmatic Bookshelf. pp. 10-11

PharmaVOICE, 2014. Five Steps to Analytical Maturity. pp. 1-5

Posavec, A. B. and Krajnović, S., 2016. Challenges in adopting big data strategies and plans in organizations. In: 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 1229–1234. pp. 1234

Provost, F. and Fawcett, T., 2013. Data science and its relationship to big data and data-driven decision making. Big data, 1 (1), 51–59. pp. 57-58

Ramanathan, R., Philpott, E., Duan, Y., and Cao, G., 2017. Adoption of business analytics and impact on performance: a qualitative study in retail. Production Planning & Control, 28 (11–12), 985–998. pp. 994

Ramsay, J. O., 2004. Functional data analysis. Encyclopedia of Statistical Sciences, 4. pp. 9

Reinsel, D., Gantz, J., and Rydning, J., 2018. The digitization of the world from edge to core. IDC White Paper. pp. 3

Richards, G., Yeoh, W., Chong, A. Y. L., and Popovič, A., 2019. Business intelligence effectiveness and corporate performance management: an empirical analysis. Journal of Computer Information Systems, 59 (2), 188–196. pp. 188

Russom, P., 2011. Big Data Analytics. TDWI best practices report, 19 (4), 1–34. pp. 6-12

Santhanam, R. and Hartono, E., 2003. Issues in linking information technology capability to firm performance. MIS quarterly, 125–153. pp. 148-151

SAS, 2016. Assessing Your Business Analytics Maturity: Eight Metrics That Matter. pp. 2

Shuradze, G. and Wagner, H.-T., 2016. Towards a conceptualization of data analytics capabilities. In: 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 5052–5064. pp. 5057-5063

Souza, G. C., 2014. Supply chain analytics. Business Horizons, 57 (5), 595–605.

Tarhan, A., Turetken, O., and Reijers, H. A., 2016. Business process maturity models: A systematic literature review. Information and Software Technology, 75, 122–134.

TDWI, 2020. TDWI BIG DATA MATURITY MODEL AND ASSESSMENT TOOL [online]. Available from: https://tdwi.org/pages/maturity-model/big-data-maturity-model-assessment-tool.aspx?m=1 [Accessed 18 Jun 2020].

Tonidandel, S., King, E. B., and Cortina, J. M., 2016. Big data methods: Leveraging modern data analytic techniques to build organizational science. Organizational Research Methods, 21 (3), 525–547. pp. 2

Tukey, J. W., 1977. Exploratory data analysis. Reading, Mass. pp. 55-7

Vesset, D., Girard, G., Feblowitz, J., Versace, M., Burghard, C., O'Brien, A., Olofson, C., Schubmehl, D., McDonough, B., and Woodward, A., 2015. IDC Maturity Scape: Big Data and Analytics 2.0. International Data Corporation. pp. 1-6

Vidgen, R., Shaw, S., and Grant, D. B., 2017. Management challenges in creating value from business analytics. European Journal of Operational Research, 261 (2), 626–639. pp. 23-24

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., and Childe, S. J., 2017. Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 70, 356–365. pp. 3

Wang, R. Y. and Strong, D. M., 1996. Beyond accuracy: What data quality means to data consumers. Journal of management information systems, 12 (4), 5–33. pp. 6-21

Ward, J. S. and Barker, A., 2013. Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821. pp. 1-2

Weller, S. C. and Romney, A. K., 1988. Systematic data collection. Sage publications.

Wu, J., Ping, L., Ge, X., Wang, Y., and Fu, J., 2010. Cloud storage as the infrastructure of cloud computing. In: 2010 International Conference on Intelligent Computing and Cognitive Informatics. IEEE, 380–383. pp. 380-382

Xia, B. S. and Gong, P., 2014. Review of business intelligence through data analysis. Benchmarking: An international journal. pp. 301

Appendix

Appendix 1. The final organizational analytics maturity model

Model was improved based on the insights from the interviews. Modified and added sections are marked with light blue color.

Dimension	Maturity Stage					Source
	1.	2.	3.	4.	5.	
Strategy						
Analytics strategy	No strategy.	Department level.	Business unit level.	Across multiple business units.	Enterprise wide.	(Vesset et al. 2015)
Alignment of analytics initiatives to the business strategy	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	(Cosic et al. 2012), (PharmaVOICE 2014)
Targets	No targets for analytics.	Multiple disconnected targets, typically no strategic importance. Analytics projects have poorly defined scope and objectives.	Analytical efforts coalesce behind a small set of important targets.	Analytics centered on a few key business domains with explicit outcomes.	Analytics integral to the company's distinctive capability and strategy. Analytics strategy is used to select appropriate analytic opportunities.	(Davenport and Harris 2017), (Grossman 2018), (INFORMS 2020)

Dimension	Maturity Stage					Source
	1.	2.	3.	4.	5.	
Organization						
Budgeting	Localized, ad hoc funding secured for each new project.	Decentralized budgets based on department's plans.	A mix of business unit–level and localized budgets; no	A mix of centralized and localized periodic budgets supplemented	Centralized and localized budgets governed by enterprise wide policies.	(Vesset et al. 2015)

			ad hoc funding.	by ad hoc funding		
Justification of projects	No formal investment justification required.	Investment requires defined business problem.	Investment requires defined business problem and expected cost savings.	Investment requires defined business case, expected cost savings, and benefits at the project level.	Investments made only based on standard enterprise wide guidelines and processes that include specific business case, cost-benefit analysis, and cost method(s).	(Vesset et al. 2015)
Performance Management	KPIs are not defined.	Measurements are unclear or qualitative.	KPIs measure success of a technology initiative, but not impact on the organization.	Metrics for evaluating process quality, results of analysis, and business outcomes success have been established.	Ongoing assessment, revision, and learning built into decision making across the organization, and business benefits can be quantitatively tied to initiatives. Systematic and broad-based rewards tied to analytics-based metrics.	(Vesset et al. 2015), (INFORMS 2020)
Collaboration	Project-based collaboration on an as-needed basis.	Collaboration is encouraged but technology and processes to do so are lacking.	Collaboration technologies enable sharing of data, metrics, and best practices among internal groups, but is not widely used.	Collaboration technology and processes enable sharing of relevant data, metrics, and best practices among internal groups; the process of collaboration is reviewed periodically.	Enterprise wide collaboration universally accepted and enforced by governed processes and tools for data, metrics, analytics, and best practices. Business and IT are working together to innovate new business opportunities with new technologies.	(PharmaVOICE 2014), (Vesset et al. 2015)
Governance Groups	No governance.	Departmental governance.	Development of enterprise-wide governance team.	Enterprise-wide governance team.	Analytics program management office or similar in place.	(PharmaVOICE 2014)

Governance and policies	No formal governance processes exist.	Departmental governance practices. policies exist for a single domain such as data or technology or security.	Development of enterprise-wide governance practices. Policies exist for multiple domains such as data or technology or security, but not for all the domains. Partially unclear for employees.	Enterprise-wide governance practices. Policies exist for most domains such as data, technology, and security, but their guidance is not necessarily followed by the organization.	Well defined strategy in place to guide governance. Policies exist for all domains such as data, technology, and security, and they guide the organization's analytics. Fully transparent to employees.	(PharmaVOICE 2014), (Vesset et al. 2015), (Grossman 2018), (Hausladen and Schosser 2020)
Analytical organization	Location of the staff with needed skills unknown.	Skilled staff reside within one group or area of the organization.	Skilled staff are distributed among IT, line of business, and analytics groups.	Staff are distributed among IT, line of business, and analytics groups and their performance are measured.	Staff are distributed among IT, line of business, and analytics groups based on strategic requirements and staff members' core competencies.	(Vesset et al. 2015)
Organizations capability to redesign and integrate new processes	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	(Davenport and Harris 2017), (Cosic et al. 2012)
Ability to appropriately budget and schedule analytics projects	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	(Gupta and George 2016)

Cost analysis and ability to proof analytics ROI	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	(SAS 2016), (Chen and Nath 2018)
Use of software development practises in data and analytics context	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	The empirical interviews.
Capability to maintain analytics applications in operation	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	The empirical interviews.

Dimension	**Maturity Stage**					**Source**
	1.	**2.**	**3.**	**4.**	**5.**	
Analytics						
Techniques	A few analytic methods and techniques used on an experimental basis. Mostly simple math, extrapolation, and trending etc.	Basic statistics, segmentation, database querying, tabulations. Not defined or managed.	A broad range of analytic methods and techniques are defined, used, and standardized. Simple predictive methods.	Broad range of advanced analytic methods and techniques are defined, used, standardized, and measured.	All kinds of data is used with most sophisticated techniques when necessary. Use of analytical methods is monitored and measured; metrics-driven results drive continuous improvement.	(Halper and Krishnan 2013), (Vesset et al. 2015), (Davenport and Harris 2017)

Overall analytics process	No defined analytics processes.	Development of department-level processes. Ad hoc activities.	Processes defined but not completely integrated. Used in parts of the organization, but not across the entire organization.	Processes deployed company wide.	Continuously refining and optimizing the processes.	(Halper and Krishnan 2013), (Chen and Nath 2018)
Model building process	No capability	Organization can build reports.	Organization can build and validate individual models on an ad hoc basis.	Organization follows repeatable process for building models.	Analytic models are built with common process whenever possible.	(Grossman 2018)
Model deploying	No capability	Organization can leverage reports.	Organization can deploy individual models on ad hoc basis	Organization follows repeatable process for deploying models.	Analytic models are deployed with common process whenever possible. Outputs of the analytical models are integrated together.	(Grossman 2018)

Dimension	**Maturity Stage**					**Source**
	1.	**2.**	**3.**	**4.**	**5.**	
Data Management						
Standards of data	No naming standards and storage structures are minimally defined. Data is collected in different file formats.	Minimal naming standards, storage structures are minimally defined. Data is collected in different file formats	Departmental naming standards and storage structures defined.	Enterprise-wide standards defined but not fully implemented.	Enterprise-wide standards for structured and unstructured data. Fully implemented.	(Halper and Krishnan 2013)
Transparency on data requirements	No data requirements recorded	Some data requirements are trans-parent but no systematic tracking in place.	Data requirements are transparent, but data properties are not systematically processed.	Data requirements and properties are transparent but no tracking what data are used in what	Data requirements and properties are fully transparent which makes it possible to quickly identify what data are used in what decision-making processes and why.	(Hausladen and Schosser 2020)

				decision-making processes and why.		
Process	No defined data management. Data quality managed within specific projects and individual groups.	Departmental or project-specific data management practices. No data life cycle management.	Data quality processes defined and documented for individual business units. Effort to identify useful data.	Solid data management and governance plan enterprise wide.	Data quality addressed enterprise wide, with ongoing monitoring, correction, measurement, and proactive issue prevention.	(Vesset et al. 2015), (Halper and Krishnan 2013), (Hausladen and Schosser 2020)
Overall quality	Inconsistent, poor quality, quality unknown.	Standardized, and structured data. Islands of data.	Data quality metrics established.	Integrated, accurate, and common. Little unique data.	Data quality metrics used for scoring data source health and made available to all organizational users.	(Davenport and Harris 2017), (INFORMS 2020)
Access	Data is stored only locally without access from network or other devices.	Localized data access provided disparately by IT personnel.	Centrally managed access to multiple types and sources of data.	All data are centrally stored and available across the organization.	Integrated access to all data types for use in all business areas (with individual usage permissions).	(INFORMS 2020), (Hausladen and Schosser 2020)
Security and privacy	Organization not aware of security issues.	Organization is aware of security issues, but the security issues are not identified.	Organization starts to identify security issues.	Structures in place to cover data security and other security issues. Room to improve.	Security infrastructure and strategies in place. State of art security and privacy practices. Continuous improvement on security issues.	(Halper and Krishnan 2013), (Grossman 2018)
Data completeness and variety	Data from one or a few limited types of internal systems.	Data from most internal systems and some external sources of limited types. Not all information needs can be satisfied.	Data from most internal systems and most relevant external sources of different types. Most information needs can be satisfied.	Data from all relevant internal and most external sources of different types. All information needs are satisfied with at least one kind of data.	All relevant internal and external data, at necessary strategic granularity. Relentless search for new data. All information needs are satisfied with the best possible data.	(Vesset et al. 2015), (Davenport and Harris 2017), (Hausladen and Schosser 2020)
Trust	Data definitions and lineage known to few users.	Data definitions and lineage are known within a	Data definitions and lineage for multiple	Data definitions and lineage are documented	Data definitions and lineage are known, documented, governed, and well understood.	(Vesset et al. 2015)

		group or for a specific project.	projects are documented.	and governed.		
Timeliness	Data rarely available on time for relevant uses. No real-time data feeds (neither intern-ally nor externally).	Data collected at fixed intervals and available at fixed intervals.	Data collected at fixed intervals and available on demand or within relevant workflows.	Continuously processed data available at fixed intervals. Some real-time data feeds available.	Continuously processed data available on demand and for relevant workflows throughout the enterprise. Real-time data available.	(Vesset et al. 2015), (Hausladen and Schosser 2020)
Data ownership and traceability of used data.	Data has no ownership.	Data traceability is limited within the system where the data is found.	Some decisions and business processes can trace their underlying data back to its source.	Data has ownership and usually it can be tracked.	Complete ownership. Responsibility for master data established. Full traceability back to original source from strategic decisions, management decisions, measures, and business processes.	(INFORMS 2020), (Hausladen and Schosser 2020)
Availability of external sources	No external data gathered.	External data gathered sporadically on need-basis.	Some external data sources used regularly, but mostly on ad hoc basis.	Regular and well-established use of external data sources.	Regular and well-established use of external data sources. Relentless search for new data.	(Hausladen and Schosser 2020), (Davenport and Harris 2017)
Transparency on available data	No transparency on data gathered by other functions	Little transparency on data gathered by other functions.	Basic transparency on data gathered by other functions.	Satisfactory transparency on internal data from other functions.	Full transparency on internal available data from other functions.	(Hausladen and Schosser 2020)

Dimension	**Maturity Stage**					**Source**
	1.	**2.**	**3.**	**4.**	**5.**	
People and culture						
Executive support	Executives unaware of the power of analytics. Lack of interest.	Some awareness. One or no executive sponsors.	More than one executive is interested. Emphasis on data-driven culture but not enough	Value of analytics is understood across the board. Leadership strongly	Executives see analytics as a critical standard for conducting business. Leadership mandates and	(Halper and Krishnan 2013), (PharmaVOICE 2014), (Davenport and Harris 2017),

			resources provided.	emphasizes a data-driven culture and assessment methods.	incentivizes the use of data, analytics, and technology.	(Hausladen and Schosser 2020)
Perceived value of analytics	Little to no value seen in analytics.	Value of analytics seen by individuals. Full benefits poorly understood.	Value of analytics is seen mostly in cost reduction instead of gaining competitive advantage.	The importance of evidence-based operations and decision making is stressed at all levels.	The importance of big data and analytics is an organizational value that all should know and embrace.	(Halper and Krishnan 2013), (Vesset et al. 2015), (Hausladen and Schosser 2020)
Formal technical and analytical skills	Lack of analytical skills; poorly organized, reactive.	Individuals with analytical and technical skills scattered around the organization. Some use of external staff. Skills generally unmanaged.	Effort to acquire more analytical and technical skills. Analytical skills are still siloed departmentally.	Analytical talent centralized; best practices shared. Enterprise's skill set is periodically reviewed centrally and enhanced as needed.	Enterprise wide skill set fed by continuous processes and recruiting to maintain broad internal and external skills inventory.	(Halper and Krishnan 2013). (PharmaVOICE 2014), (Vesset et al. 2015), (Davenport and Harris 2017)
Analytical skills of non-analytical employees.	Staff lack of awareness of analytics	Staff have mainly a personal interest in analytics but lack the required skills to track the fast-paced technological evolution.	Individual experts develop deep knowledge on analytics tools and topics.	All staff are fully engaged with analytics technology and tools.	All staff feel empowered to experiment with analytics tools beyond the formal definition of their role.	(Hausladen and Schosser 2020)
Mindset and attitude towards analytics	Staff is unaware about big data and analytics.	Skepticism around analytics. Attitude is entrenched in a negative way towards IT-driven innovation.	General interest in analytics through the organization.	Most of the company sees analytics as a competitive differentiator. Positive attitude.	The company is continuously determining new ways to use analytics and create value from it.	(PharmaVOICE 2014), (Hausladen and Schosser 2020)
Fact-based management and use of analytics	Analytics is not used to support decision making.	Use of analytics to support decision	Evidence based decisions and judgement calls occur	Most decisions are evidence-based and	All major decisions are evidence-based and grounded in data, and all	(Halper and Krishnan 2013), (PharmaVOICE

		making is inconsistent.	with similar frequency.	grounded in data and decision makers are trained sporadically to use and interpret data.	decision makers are trained to use and interpret data on a regular base.	2014), (Chen and Nath 2018), (Hausladen and Schosser 2020)
Domain knowledge of analytics specialists	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	(Cosic et al. 2012) Modified based on the comments from empirical interviews
Business skills of analytics specialists	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	(Cosic et al. 2012) Modified based on the comments from empirical interviews
Managerial skill in the analytics orgnization	Non-existent: the organization does not have this capability.	Initial: the capability exists but is poorly developed.	Intermediate: the capability is well developed but there is much room for improvement.	Advanced: the capability is very well developed but there is still a little room for improvement.	Optimized: the capability is so highly developed that it is difficult to envision how it could be further enhanced. At this point the capability is considered to be fully mature.	(Cosic et al. 2012) Modified based on the comments from empirical interviews
Training	No formal training on technology, data, or analytics.	Training provided on specific technology as needed for specific projects.	Training on technology, data, and analytics provided at regular intervals.	Training needs and outcomes on data, technology, and analytics are provided and assessed periodically.	Training on data, technology, and analytics incorporates world-class best practices across internal groups and external sources.	(Vesset et al. 2015)

Dimension	**Maturity Stage**	**Source**

	1.	2.	3.	4.	5.	
Technology						
Development of infrastructure	Some projects have defined scope and objectives. No identification of advanced analytics technologies.	Project-driven, often reactive; no best-in-class sharing; completeness unknown. Available advanced analytics technology identified only incidentally.	Clear project life cycles and processes. Plan for development of infrastructure. Sporadic market screening for available advanced analytic tools.	Projects aligned to strategy; documented best practices. Infrequent but regular market screening for available advanced analytic tools.	Continuous improvement/learning to support the most difficult business challenges. Market screening for available advanced analytics tools integrated in normal corporate planning cycles. Able to address any analytics request with in-house development when commercial alternatives are not viable.	(PharmaVOICE 2014), (Davenport and Harris 2017), (Hausladen and Schosser 2020), (INFORMS 2020)
Architecture and deployment of technologies.	Misunderstanding the need to differentiate the operational infrastructure and analytics related infrastructure. Ad hoc deployment of siloed technologies; no defined architecture.	Each department selects its own methods, software, and hardware. Some planning on the architecture.	Infrastructure and software indexed and retrievable, desire for new features. Architecture is reviewed and modified on occasions.	Architecture is unified. Enterprise-wide standards for installation, configuration, and maintenance. Architecture governed by a central architecture board.	Centralized approach to select methods, software, and hardware for various problems. Architecture is flexible and centrally governed to easily adapt new user needs.	(Halper and Krishnan 2013), (Vesset et al. 2015), (INFORMS 2020)
Performance of technologies	Poor performance, no monitoring and management processes and skills.	Moderate performance requiring manual management; no monitoring capability.	Satisfactory performance with some monitoring and management processes, skills, and tools.	Optimized performance that requires substantial manual effort for processes and tools.	High level of automation in systems management resulting in optimized performance and dynamic scalability.	(Vesset et al. 2015)
Functionality of technologies	Limited data management. and analysis functionality for one specific use case.	Some data management. and analysis functionality for several use cases.	Data management and analysis functionalities for many use cases.	A broad range of data management and analysis functionalities to address most use cases. Capability to	A proactively updated, comprehensive range of governed data management and analysis functionality addresses all use cases. Capability to	(Vesset et al. 2015), (Hausladen and Schosser 2020)

				handle some forms of unstructured data.	handle all forms of unstructured data.	
Flexibility to add new data sources	No corporate IT architecture in place.	IT architecture existing, but no integration of new data.	Sporadic integration of new data sources on a case-to-case base.	Flexible integration of new data sources as needed.	Fully event-driven network planning architecture, capable to add any required data source.	(Hausladen and Schosser 2020)

Appendix 2. The empirical interview questions

Data & analytics maturity, Questions for the interview

The term *analytics* is defined in this interview as follows: "The extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions". (Davenport 2017)

Goal of the interview is to gain valuable information from practitioners and researchers to, complement earlier findings from the literature or identify completely new findings not discussed in the existing literature, and validate the composed analytical maturity framework.

Basic information

1. Basic information of the interviewee
    a. Name
    b. Role in the organization
    c. Experience, years
    d. Business department/team
    e. Number of employees in your department

Data and analytics capabilities

2. What kind of terminology your organization uses to describe data and analytics related activities (e.g. Data mining, statistical analysis, data science, business analytics, etc.)?
3. Does your organization clearly identify distinct analytical capabilities?
    a. If yes, what are these capabilities?
4. Does your organization see that it is important to identify and review these capabilities to develop the overall analytics ability?

Developing data and analytics capabilities

5. How are you currently developing your analytics capabilities?
    a. How is the development measured?
6. What are the targets for analytics development in your organization?
    a. How are these targets communicated inside and outside of your department/team?
7. How mature you perceive your organization's overall analytical ability to be?
    a. What are the areas where your organization is good at?
    b. What are the areas that need more development?
8. What are the main challenges when developing overall analytics capability?
    a. How would you solve these challenges?
9. Any other comments regarding analytics development?

Data and analytics maturity model

10. Would you see analytics maturity models as beneficial for your organization and development of analytics?
11. Are you already utilizing any maturity models in your organization?
    a. If yes, what model and how the model is used?
12. What subdimensions would you prioritize from the created analytics capability maturity model?
13. Are there any dimensions you would modify, add, or remove?
    a. If yes, what, and why?
14. Any other comments regarding the analytics capability maturity model?