Bachelor Thesis

Economic evaluation of data sets: A descriptive approach to determine the value of data sets in the context of digitalization and Industry 4.0, by describing the concept of the Real Option Theory.

Hochschule für Angewandte Wissenschaften Hamburg
Faculty of Business and Social Sciences
Department of Business
Foreign Trade/International Management (B. Sc.)

Bachelor Thesis, Summer Term 2020
Prof. Dr. Christian Decker, Prof. Dr. Michael Gille
Date of Submission: 01.06.2020

Student: Jesse Julian Karathomas

[redacted] [redacted]

I Abstract

The present work illustrates if the Real Option Theory can be applied to determine the value of data and data sets within a corporate environment. This thesis will present the key to digitalization strategies and its consequence in Industry 4.0. It will illustrate the necessity of a reliable data foundation and point out the investments which have to be made in order to maintain the information quality drawn from the data. Determining the value of these investments by applying traditional discounted cash flow methods is difficult, as they do not offer suitable approaches /measures to include uncertainty and flexibility into the investment. The Real Option Theory in general does provide this potential and is therefore more applicable. Accordingly, this work will introduce the concepts of the Real Option. Furthermore, it will link this concept to investment projects in data respectively data sets.

Key Words: Digitalization, Industry 4.0, Internet of Things, Data, Data Quality, Data Quality Management, Real Option Analysis

JEL Classification: M20, M15, G31

II Outline

III List of Figures

IV List of Tables

V List of abbreviations

IoT - Internet of Things

CPS - Cyber-Physical-System

R&D - Research and Development

DQM - Data Quality Management

AI - Artificial Intelligence

ROA - Real Option Analysis

1 Introduction

1.1 Research Problem

The terms digitalization and Industry 4.0 have become omnipresent terms within corporate environment. Yet despite the awareness and the presence of these developments, which are seen as a potential cost savers, it is still questionable if entities emphasize it enough, or have the means to argue an investment into these developments within their structure respectively towards their stakeholders. Digitalization concepts as well as the development of an Industry 4.0 network require the identification and analysis of data to generate information and knowledge on which they can be based upon.

According to a survey conducted by the IW-Zukunftpanel 73% of the surveyed companies do not measure the value of the datasets which are available to the companies (Wirtschaft, 2019). The survey furthermore points out that half of the companies which determine the value of the available data are doing so only based upon the costs it creates to pull, maintain, use and save the data. This one-dimensional view on the value of data enforces an understanding which might not be beneficial and is potentially even harmful when moving forward into digitalization and Industry 4.0. Considering that assets are usually not only valued by its costs, but also by the potential of cost savings as well as potentially generated income. Generally the value of an asset is determined by traditional discounted cash flow methods. Problematic is, that these methods lack the opportunity to value the flexibility an investment project might obtain. But especially this flexibility and uncertainty is at the core of data investments. Therefore, applying a DCF method to this kind of project will not lead to an accurate evaluation, which consequently might lead to a false decision whether to execute a project, or not.

In order to address this inaccuracy, this paper will introduce an alternative method how to evaluate an investment. This method is based on the concept of financial options but is adjusted to the nature of real assets, such as machinery and other tangible assets, and is referred to as the Real Option Analysis. At the center of the analysis is the appreciation of uncertainty and flexibility.

The aim of this thesis is to analyze how the value of collected or to be collected data sets can be determined. A special focus will be on data sets that are gathered in the context of digitalization processes and Industry 4.0.

To tackle this research problem, this paper will start with an elaboration on digitalization and Industry 4.0. This also serves to argue why an investment in data is generally necessary. Following this elaboration the work will focus on introducing the basic concept of the Real Option Analysis. It will furthermore, provide a selective overview on which options do exist and additionally mention some examples for what kind of investment projects options have been developed and evaluated. Following this general overview, this thesis will discuss how data investment projects suit to be valued by Real Option Analysis. In the last chapter this work will be concluded by providing a summary, a critical acclaim as well as an outlook.

2. Digitalization and Industry 4.0

2.1 Digitalization

Digitalization originally described the process of transferring analogue values into digital values (Mertens, Barbian, & Baier, 2017, p. 35). Storing these information digitally is specifically undertaken to analyze and interpret them in order to ultimately generate knowledge and improve processes (Mertens, Barbian, & Baier, 2017, p. 35). This rather closed definition is nowadays widened and the term digitalization implies further understandings such as automation, more flexibility, and further individualization (Stumpf, 2019, p. 9).

Digitalization in a business context implies, the transformation of business models as well as products and services (Fleischmann & et. al. , 2018, p. 10). Subsequently to these trends, the analysis of Big Data and implementation of Cloud Computing, as well as an increased usage of mobile devices was and is observable (Stumpf, 2019, p. 9). Therefore, these developments do not only effect the professional life but also have an impact on the private life (Schildhauer & et al., 2019, p. 14; Huber, 2018, p. 2; Brauckmann, 2019, p. 73). The analysis of Big Data and the use of the internet in our private lives are interrelated, because Big Data means to process and analyse the data each of us generates while using the internet (Schildhauer & et al., 2019, p.

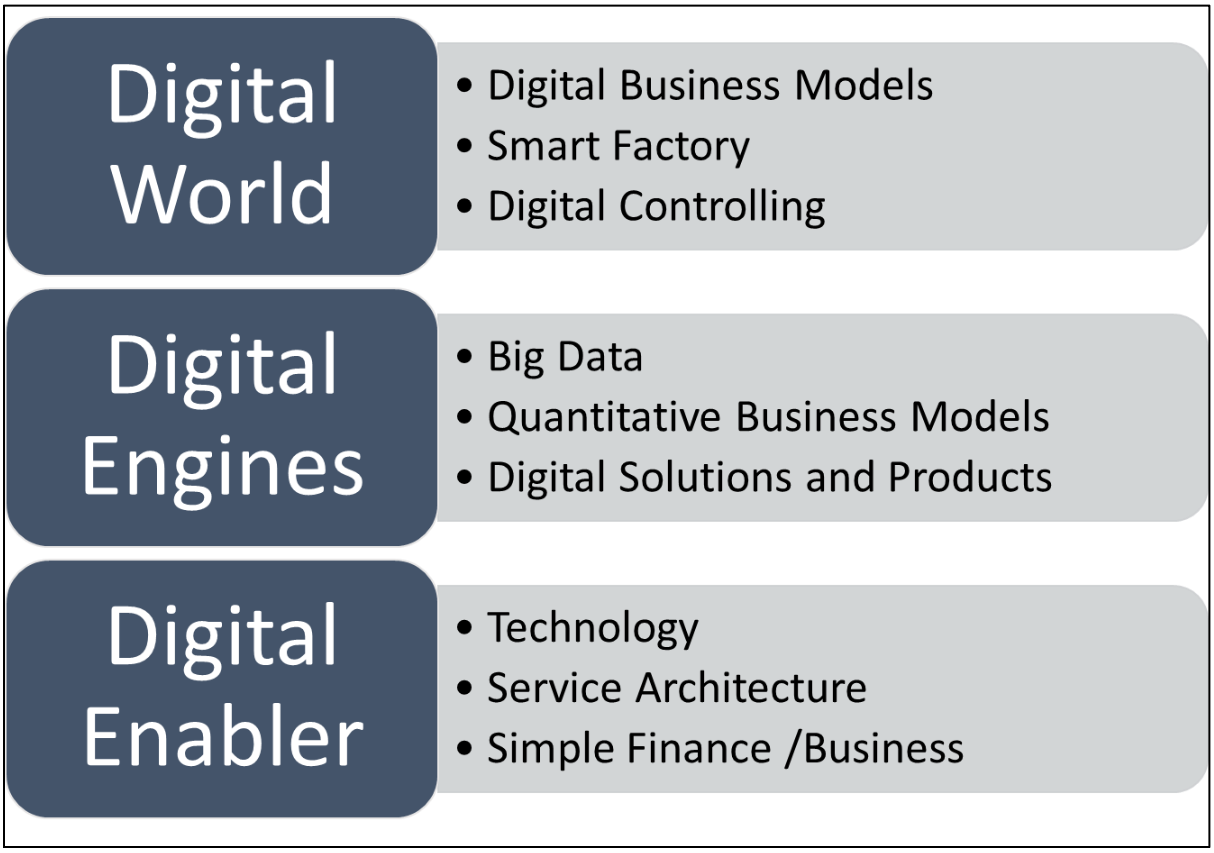
15). Accordingly, Big Data serves as the basis, for many partially not even foreseeable, potential uses (ibid.). While originally the accessibility to data was seen as a position of strength for a corporation, the now available, huge amounts of data do no longer constitute a point of difference for a company (Maisch & Palacios Valdés, 2018, p. 46). A closer look on Big Data will be provided in the following chapters.

As mentioned, the digitalization of a corporation is accompanied by a creation of new processes and thereby a transition of their business model into completely new models, which constitutes the definition of a disruptive development (Schlotmann, 2018, p. 1; Brauckmann, 2019, p. 81). Companies face the challenge, to understand that their ability of previous innovativeness will not necessarily equal to the future ability to be inventive and stay state of the art (Schlotmann, 2018, p. 4). Previously, manufacturers have focused on cost efficiency throughout their manufacturing process in each cost center, that focus often caused a negative effect for the customer, as it causes increasing lead times within production lines, if an individual article is required by the customer (Brauckmann, 2019, p. 82). Digitalization is based upon the implementation of sensors, as these sensors generate data with an increasing accuracy to provide new information and furthermore, they connect machines to one another (Brauckmann, 2019, p. 89). Digitalization is therefore the basis to a fourth industrial revolution, ultimately resulting in Industry 4.0 (Wiegand, 2018, p. 2; Brauckmann, 2019, p. 6). These several layers of digitalization are presented in the following Chapter.

2.2. Layers of Digitalization

The following Figure 1 reflects the three layers of digitalization with its respective components. This figure shall visualize key components of which digitalization requires as well as the impacts it will have. While the bottom layer digital enablers predominantly reflect the prerequisites. The layer digital engines refers to the components which are further enforcing digitalization strategies. The top layer, the digital world presents what the two previous layers can lead to.

Figure 1 Layers of Digitalization



Source: Own Graf Based upon Klingebiel, 2019, p. 164

As already described, the figure reflects three evolutionary steps of digitalization, containing three components each. Some of these components will be further and more detailed elaborated on throughout the following chapters. Among other key components to enable digitalization processes the infrastructure within an entity in form of hardware as well as service architecture is essential. The following section refers to tools and ideas which will further push digitalization processes once the infrastructure is implemented as described in the previous section. Especially the term Big Data will be further explained in the following chapters. The last section herein named the digital world, refers to new concepts which will be developed due to further digitalization. This refers to new business models which will be developed, as well as the implementation of smart factories which indicate the development of Industry 4.0. Industry 4.0 will be explained in detail throughout the proceeding chapters.

2.3 The fourth Industrial Revolution (Industry 4.0)

The term Industry 4.0 refers to the outcome of the fourth industrial revolution (Wiegand, 2018, p. 2; Obermaier, 2019, p. 6). Industry 4.0 first appeared and got major recognition at the Hannover Messe 2011 (Hanover Industrial Fair) (Frederick, 2016, p. 9; Huber, 2018, p. 7; Obermaier, 2019, p. 5). Initially, the term was used by the German government as a name for the project to digitally connect traditional producing industries (Obermaier, 2019, p. 6). Accordingly, it is a predominantly German term. Nevertheless, all major industrial nations have set up similar programs which might have different names (Huber, 2018, p. 1). For instance, Internet of Things (IoT) for North America (Huber, 2018, p. 7). At the center of this fourth revolution, following the first three major revolutions of mechanization, automation and digitalization, is the connection of industrial infrastructures (Obermaier, 2019, p. 3). A further elaboration on the IoT will be undertaken in the following chapter.

It has become obvious that this development is not a movement, which will be gone after a little while, but it is going to have major impact on the structure of our economic world, not only within production, but also on research and development (R&D) as well as aftersales services and other sections within businesses (Huber, 2018, p. 1). Predominant technologies , such as but not limited to, robotics, artificial intelligence, cloud computing or big data are going to be effected (Frederick, 2016, p. 10). Accordingly, completely new value chains and business models are being developed (Huber, 2018, p. 1). Instead of managing a single production factor, the focus will be on the complete value creating process (Brauckmann, 2019, p. 6). Therefore, manufacturing processes will become increasingly complex (Ebert, 2018, p. 269). According to (Brauckmann, 2019, p. 8), Industry 4.0 will reshape the value chain towards the needs of the consumer rather than focusing on single manufacturing stages. A key aspect of this fourth industrial revolution is that it is based on freely accessible and exchangeable data and information (Frederick, 2016, p. 10). In 2015 the international exchange of data has contributed more to the global economic growth than traditional trade of goods (Huber, 2018, p. 4). Furthermore, accessible data is not only restricted to one matter, but covers various which can lead to the revelation of new patterns and correlations (Frederick, 2016, p. 11). Especially, because the entire process will be represented and therefore can be optimized digitally (Brauckmann, 2019, p. 6).

2.4. The Internet of Things (IoT)

As pointed out in the previous chapter, the Internet of Things (IoT) describes the process of connecting all physical things with one another and let them communicate and exchange information in form of data (BmBF, 2017, p. 7). Thereby, data is being generated and gathered without time delays and no need of human interference (Otto & Österle, 2016, p. 5). According to (Frederick, 2016, p.10) the term is not precise as it also includes the potential connection of humans respectively their bodies to the internet, and should therefore, be named "The Internet of all Physical Things". Therefore, the basis for the IoT as well as the internet of all things are Cyber-Physical Systems (CPS) which sensor their environment, save and evaluate data, communicate with each other and impact the physical world through actuators, with the help of embedded systems, (Hofmann J. , 2018, p. 6). Hence, one prerequisite to actually connect things with each other is to implement sensors and actuators (Sendler, 2018, p. 37). Additionally, humans will be connected through the CPS with machines (Hofmann J. , 2018, p. 6; Obermaier, 2019, p. 4). Accordingly, even though the core of the IoT is the Machine-to-Machine-communication (Stumpf, 2019, p. 17), it is not only the machines which are communicating with one another (Wiegand, 2018, p. 2). The machines are also communicating with the customer, that provides the customer with the opportunity to make individual changes to the product throughout the process of manufacturing and even shortly prior to delivery (Wiegand, 2018, p. 2). For that reason, it will reshape the traditional production planning inevitably (Obermaier, 2019, p. 6). Potentially, each component of the final product constantly provides a status update throughout the complete value chain (Wiegand, 2018, p. 2). That indicates while previous data analytics focused on certain points of production, it is shifting to an analysis of the complete manufacturing process (Otto & Österle, 2016, p. 6). The connected machines will not only generate the data themselves, but are also capable of processing the received data (Otto & Österle, 2016, p. 6). Therefore, the ability to manage data is essential to the success of a company (ibid.).

2.5 Data Management

2.5.1 Data quality

Data has overtaken a dominant role within corporations (Huber, 2018, p. 3). As described in chapter 2.2 the huge amounts of freely accessible data and the expected enormous amounts of data, which will be generated within the IoT, provide various opportunities. Yet, for corporations it is more efficient to keep their data sets rather small as those smaller sets are easier to maintain up to date (Frederick, 2016, p. 11). Especially, because corporations need to decide on which data is actually usable and provides information for the issue at hand (Sendler, 2018, p. 37). Therefore, it is not only the accessibility to data, which is crucial for a business's success, but also the quality of the available data (Engelmann & Großmann, 2011, p. 3). The quality of data indicates to the extent to which the data is useable for specific processes (Otto & Österle, 2016, p. 11; Morbey, 2011, p. 16). The quality of data is assessable through four dimensions; completeness, accuracy, consistency, and actuality (Heinrich & Klier, 2011, p. 47; Hinrichs, 2002, p. 6). Furthermore, two basic requirements have to be fulfilled; scientific correctness as well as usability in a practical context (Heinrich & Klier, 2011, p. 49). Evaluating the data quality is an ongoing process, to be able to determine if the quality of a specific data set is no longer meeting the requirements, and potentially make changes to the respective data set (Otto & Österle, 2016, p. 20). Especially, because data quality is subject to change as data reflects the reality and reality is constantly changing (Otto & Österle, 2016, p. 31). A constant reevaluation of the data quality itself requires a data quality management (DQM) (Otto & Österle, 2016, p. 18). An indicator for a general lack of data quality in corporations is that entities spend on average one to five percent of their annual turnover on software, but often companies are not able to answer basic questions, such as what the size of their product portfolio is (Otto & Österle, 2016, p. 14). Therefore, a high degree of data quality leads to better and more reliable information (Eppeler & Helfert, 2004, p. 312).

From an economic point of view, information is essential to set the right targets in various perspectives (Engelmann & Großmann, 2018, p. 4). The information can be used for manufacturing, to develop competitive strategies, or in general preparation of decisions which will have to be made (Engelmann & Großmann, 2018, p. 4). The goal is to use stored information to generate more knowledge (Engelmann & Großmann, 2018, p. 8). Information, can be acquired through external channels, or be generated within a corporation (Engelmann & Großmann, 2018, p. 20). The terms data, information, and knowledge are interrelated and need to be looked at simultaneously (Engelmann & Großmann, 2018, p. 11). Yet, the minimum basis of reliable information is the completeness and accuracy of the underlying data (Rohweder & et al., 2018, p. 24).

2.5.2 Data Quality Requirements

The following Table 1 will provide a brief introduction of the in chapter 2.3 mentioned dimensions data and datasets have to meet. While the first column mentions the respective dimension, the second column points out the rational of each requirement.

Table 1 Dimensions for Data Quality

Dimension:	**Rational:**
Correctness	The Data needs to correspond to the actual object it is describing.
Consistency	Data on the same object for instance stored at different locations need to be aligned.
Completeness	All attributes of a dataset need to be available.
Actuality	The data must be up to date at any point in time, with respect to the status of the object.
Accessibility	The user must have access to the data at any time.

Source: Own table based upon Otto & Österle, 2016, p. 31.

The goal of fullfilling the above mentioned dimensions, is to generate reliable Information. Therefore, the data quality is essential to the resulting quality of the information (Rohweder & et. al, 2015, p. 25). Crucial to the degree of the quality is the clarity of the presentation (Rohweder & et. al, 2015, p. 26). Furthermore, if the correctness to the analysed data is mentioned, stakeholders can assess how trustworthy the offered information is (Morbey, 2011, p. 19).

2.5.3 Data Quality Management

Data Quality Management (DQM) refers to the function of a corporation, to constantly maintain and increase the quality of the available data (Otto & Österle, 2016, p. 1). As previously described the enormous amounts of data, which are generated through the IoT, need to be valuated and interpreted, yet in order to draw the right conclusions it is essential that the underlying data is correct (ibid.). The data quality is organized on two levels; whereas the DQM itself the result, and the DQM-Organization and the processes for DQM serve to reach that goal (Otto & Österle, 2016, p. 23). A working DQM requires measurability, that can be achieved through the implementation of key performance indicators (KPI), which are interrelated to the KPIs of the underlying business processes (Otto & Österle, 2016, p. 24). Furthermore, DQM needs to be organized on a superior level as data is an overlapping topic, which crosses the borders of departments and divisions of an entity (ibid.). This cross-divisional organization of data management requires various roles within the different divisions (ibid.). Hence, members of the different divisions are responsible to maintain the quality of the available data, which is generated within their division (ibid.). Additionally, essential to a successful DQM is transparency to where the data is being stored, as well as through which channels the data flows (ibid.). In general, DQM can be divided into two perspectives (Otto & Österle, 2016, p. 32). On the one hand, a proactive perspective where the underlying idea is to avoid false data to enter the dataset. And on the other hand, the reactive approach which tries, to correct already existing data defects (ibid.). While the proactive approach carries unproportionally increasing marginal costs, it also reduces the overall risk, because the better the data quality the lower the risk of spending preventable costs (Otto & Österle, 2016, p. 34). Therefore, to prevent increasing marginal costs but also to reduce the risk of false information, the goal is to have a combination of measures, resulting from both perspectives (Otto & Österle, 2016, p. 34). Data quality becomes even more important as the amount of available data is rapidly increasing and are being analyzed through Big Data.

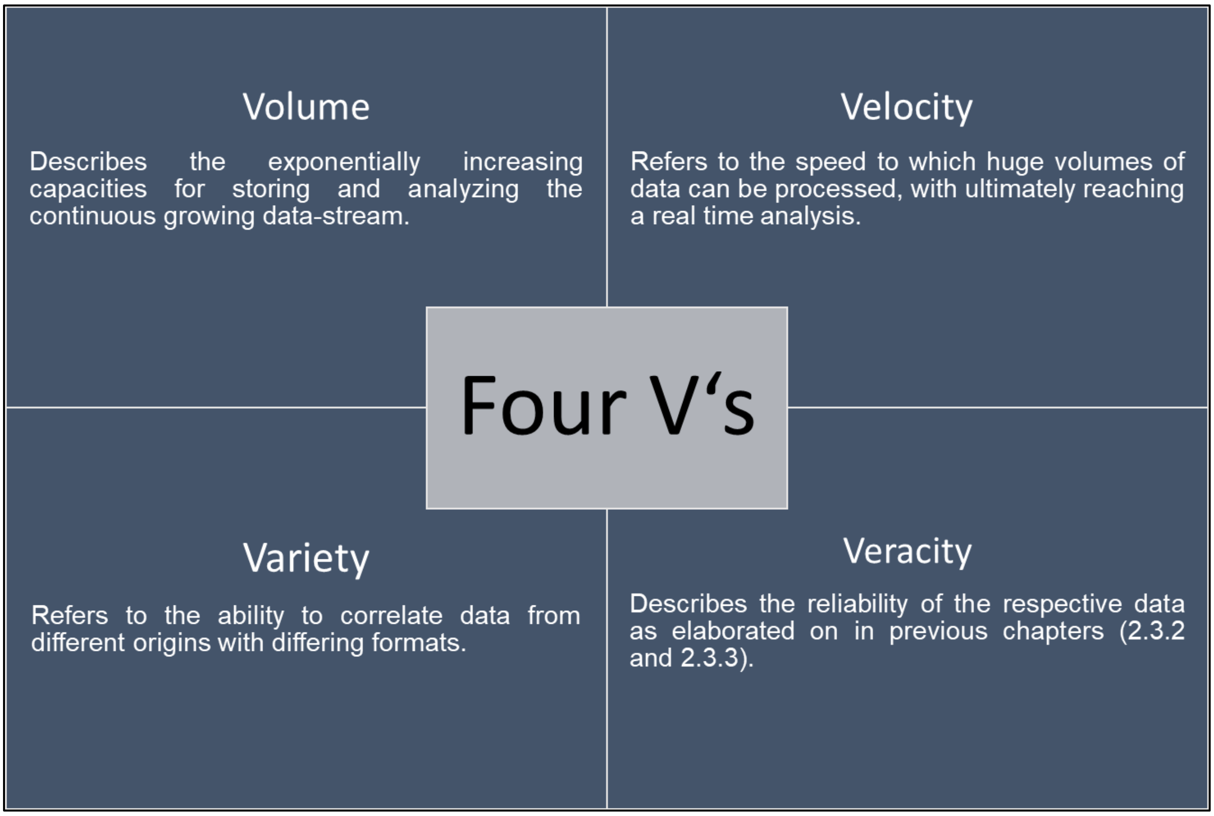
2.5.4 Big Data

The term Big Data, refers to massive amounts of unstructured data, which are generated in various industries, such as the financial sector, health sector or also aviation (Bendel, 2018; Brauckmann, 2019, p. 9). Big Data describes the process of analyzing mass data of various sources, containing complex and unstructured information through algorithms and artificial intelligence (AI) (Knorre, Müller-Peters, & Wagner, 2020, p. 6; Brauckmann, 2019, p. 9). Algorithms, describe mathematical concepts, which are able to solve specifically defined problems, by applying predefined processing regulations (Knorre, Müller-Peters, & Wagner, 2020, p. 6). In Comparison AI is able to not only apply the given regulations, but is capable of adjusting to a new situation, and adapting the existing forms by realizing the presence of new patterns within the new acquired data, and consequently, developing new solutions in real time. Accordingly, the terms Big Data, algorithm respectively Artificial Intelligence are inevitable interrelated (ibid.). As previously mentioned, the goal of Big Data is to generate usable information. The validity of the information can be determined by the information life cycle (Coyne & et al., 2018, p. 154). The basis for applying Big Data are four V's: Volume, Velocity, Variety and, Veracity (IBM, 2020; Ortner, Papp, & Meir-Huber, 2019, p. 5). The following chapter will discuss these four V's.

2.5.5 The four V's of Big Data

The following Figure 2 will provide a brief overview on the four V's in the context of Big Data. Each term is accompanied by a short explanation. The exponentially increasing volumes refer to Moors law, to which every 12 to 24 month, the storing capacities are doubled while simultaneously the respective computer chips are shrinking in physical size with an increasing performance (Knorre, Müller-Peters, & Wagner, 2020, p. 6). The second section "velocity" specifies the speed of with which data being processed, analyzed, and transformed into information, which can ultimately result in a real time analysis. Segment three, is specifically pointing out that data will originate from various sources, with various presentations, and originally differing intentions to be generated, yet, through AI can be put into correlations and used for the development of new perspectives. The last section "veracity" indicates that the analyzed data needs to be reliable in its quality, which has been at the center of the chapters 2.3.1 Data Quality and 2.3.3 Data Quality Management.

Figure 2- The Fours V's-of Big Data



Source: Own Figure based upon Knorre, Müller-Peters, & Wagner, 2020, p. 6; Ortner, Papp, & Meir-Huber, 2019, p. 5

These key characteristics of Big Data, lead to Big Data actions, which can be concluded in a Big Data Life Cycle in order to make the data usable and draw information (Coyne & et al., 2018, p. 157). The following chapter will picture, and explain the respective steps of the Big Data Life Cycle.

2.5.6 The Big Data Life Cycle

The following figure on page thirteen shows the Big Data Life Cycle, by picturing its necessary steps. Starting at the top, it shows that initially an assessment of the needs is required. Essential, to the assessment is that a corporation has to name the data and information that is mandatory to answer a certain question (Coyne & et al., 2018, p. 155). Following this initial step, the respective data needs to be collected. The source for the data can be various, it is essential that due to the big amounts of data which potentially can be collected, a precise valuation to which sources might be most beneficial, to answer the question at hand, is undertaken (Coyne & et al., 2018, p. 158). The next step "sifting" refers to evaluating the collected data and excluding data which might not be necessary to analyze since that data brings no further knowledge, as well as excluding the data which might be incorrect or incomplete (ibid.). Subsequently, to the sifting the data can be ingested into the corporation's information system (ibid.). Synchronization now refers to the step in which the newly generated data is being combined with already existing data (ibid.). After synchronizing the data can be stored (Coyne & et al., 2018, p. 159). The step preprocessing is not obligatory, especially, if the respective data is already suitable for analysis this step is not required (ibid.). Nevertheless, preprocessing can help to reduce further workloads, by for instance, removing duplicates (ibid.). Considering the enormous amounts of data which are constantly being generated, a corporation has the need to regularly refresh their data in use, to not rely on false information, that also requires constant interpretation of the respective data, which are the next two steps (ibid.). The following three steps, monitoring, analytics and reporting now reflect the steps, which are creating the actual additional value to the corporation (ibid.). The term monitoring hereby, indicates that instead of looking at data from time to time it is an ongoing process which needs to be performed constantly (Coyne & et al., 2018, p. 159). Analytics refer to the use of software like AI to analyze the data (ibid.). Whereas it is essential to the Data Life Cycle, as it is expressing the step in which the generated additional information is brought to the decision makers in order to steer the corporation carefully prepared (Coyne & et al., 2018, p. 160). The ultimate task is to dispose the data which is no longer needed, to keep the data up to date and clear occupied storage capacities (Coyne & et al., 2018, p. 158).

Figure 3 – The Big Data Life Cycle

Source: Own Figure, based upon Coyne & et al., 2018, pp. 154-155

The above mentioned steps can, furthermore, be gathered in three superordinate categories. According to (Coyne & et al., 2018), those are: creation which combines the steps, needs assessment up to synchronization. Secondly, maintenance which summarizes the steps storage to interpretation, plus disposition. The third category, use, which includes the three steps: monitoring, analytics and reporting (ibid.).

2.6 Cloud Computing

2.6.1 Definition

As previously pointed out, cloud computing contributes an essential component to digitalization strategies. The following chapters provide a short introduction to this topic. Cloud Computing is not describing a single technology yet rather refers to several progresses with respect to technology as well as user expectations (Barton, 2014, p. 41) Cloud Computing describes a service, that offers the usage of web based IT-Infrastructure, applications and various other possible uses (Baun, Nimis, Kunze, & Tai, 2011, p. 1). Accordingly, Cloud Computing is based on the internet as a global communication platform, which accessibility from all over the world especially through mobile devices (Barton, 2014, p. 41). The term cloud herein describe the possibilities that a user of a cloud service can simultaneously offer their own service within that cloud (Baun, Nimis, Kunze, & Tai, 2011, pp. 1-2). Cloud services are enormously scalable and very dynamic (Baun, Nimis, Kunze, & Tai, 2011, p. 2).They can provide additional resources and applications when actually needed without causing long term expenses (Baun, Nimis, Kunze, & Tai, 2011, p. 2). Accordingly if the demand increases the cloud service can provide that additional need by its automatically growing infrastructure (Baun, Nimis, Kunze, & Tai, 2011, p. 2). The growing infrastructure, is especially reflected by the progress made within the information technology as the several components of infrastructure for instance tablets, smartphones and other mobile devices become more efficient but cheaper at the same (Barton, 2014, p. 41). Furthermore, this includes that the devices are increasingly able to process data due to decreasing costs for storage capacity (Barton, 2014, p. 41). Therefore, the customer is at any point able to use the capacity which is currently needed and is only charged for the capacity actually used (Baun, Nimis, Kunze, & Tai, 2011, p. 2). And the goal of Cloud Computing is to no longer own a specific amount respectively capacity but to make use that capacity as a service (Barton, 2014, p. 44) Hence in case the cloud is not used costs do not occur (Baun, Nimis, Kunze, & Tai, 2011, p. 2). Consequently, Cloud Computing has an enormous economic potential as costs can be significantly reduced by its flexible structure (ibid.). The following table presents an overview of the key characteristics of cloud computing.

Table 2 – Cloud Computing Characteristics

Characteristic	Explanation
Service provision by demand	The service is provided only if demanded by the consumer and does not require any further human interaction.
Web based accessibility	Services can be used through standardized technologies.
Resource pooling	Resources are consolidated within pools and are able to provide its services to a number of customers simultaneously but meeting each of the customers individual needs.
Elasticity	Capacities are individually provided and thereby offer a scalability which seems to be unlimited for the user.
Measureable quality of service	The offered services are measureable in quality as well as quantity to bill the service in accordance to the actual usage.

Source: Own Table based upon Baun, Nimis, Kunze, & Tai, 2011

2.6.2 Costs and Benefits of Cloud Computing

Data storage as well as data processing is predominantly no longer undertaken through in-house technological infrastructure, yet is being performed within cloud services (Borges, 2018, p. 1). The dynamic adaptability of IT-Resources within Cloud Computing, and the consumption based billing creates the opportunity for corporations to reduce the fix costs and account them incurring costs as operative costs (Baun, Nimis, Kunze, & Tai, 2011, p. 117). This flexible billing method causes an elasticity which reduces various risks, such as overprovisioning, and underprovisioning (Baun, Nimis, Kunze, & Tai, 2011, p. 118). The following table will briefly introduce these terms.

Table 3 – Risks reduced through Cloud Computing

Risk	**Brief Explanation**
Overprovisioning / Underutilization	Resources > Demand
Underprovisioning / Saturation	Demand > Resources

Source: Own Table based upon Baun, Nimis, Kunze, & Tai, 2011, p. 118

Cloud Computing offers an elasticity which enables the provider to increase or decrease capabilities within minutes to meet the actual demand (Baun, Nimis, Kunze, & Tai, 2011, p. 118). In theory Cloud Computing even offers an unlimited scalability (Baun, Nimis, Kunze, & Tai, 2011, p. 119). The price as well as the billing very much depends on the function actually used by the user, while some services are predominantly billed on a monthly basis such as an generalized service fee others can be booked and are accordingly billed if needed (Barton, 2014, p. 46). For instance the transfer of data can be billed as a single event which is not included within the obligatory service fee (Barton, 2014, p. 47). Accordingly, additional costs often occur for extra processing power, as well as access to additional data (Barton, 2014, p. 47). Nevertheless, the collective usage of resources creates an enormous efficiency advantage, and thereby reduces the overall costs (Borges, 2018, p. 1). Despite the mentioned advantages Cloud Computing also carries specific risks, which need to managed, especially because the data is being accessed through the internet and thereby offers potential weak spots for attacks (Borges, 2018, p. 2).

2.7 Costs of Data and Digitalization

As described in chapter 2.1 digitalization does not only refer to the implementation of sensors within a manufacturing process, but potentially changes the entire business model of an entity (Fleischmann & et. al. , 2018, p. 10). Therefore, costs do not only arise from technical upgrades which have to be performed, but more so are the result of necessary changes within a corporation's infrastructure (Brauckmann, 2019, p. 90). That also implicates that a shift to new key positions exercising new key activities within the organization will be undertaken, those key activities will include data-analytics, software development, as well as data protection and security (Häckel & Übelhör, 2019, p. 174). Yet the costs for that transfer process are not reliably plannable, especially, because it is very uncertain what the economic consequences will be, which do play a severe role within traditional book keeping, but for instance customer satisfaction is not yet measurable in monetary units (Brauckmann, 2019, p. 90). Costs furthermore, occur to maintain the data quality, but also are the result of faulty managerial decisions based on data with poor quality (Haug, Zachariassen, & van Liempd, 2011, p. 170). Even though, due to innovations with respect to the hardware which is being used to save the data, and the respective declining prices, reasonable costs occur for storing the data especially in the software sector (Tallon, 2013, p. 34). As described in chapter 2.3.6 it is not always possible for a corporation to know which data is useful and which is not. Therefore, an investment into data which might not be suitable can occur (Coyne & et al., 2018, p. 158).

3 The Real Option Analysis theory

3.1 Definition

Real Options Analysis (ROA) serves as a tool to value non-financial or 'real' investments with flexibility and uncertainty (Moriarty & Palczewski, 2019, p. 8). The essential core of option pricing techniques is to determine the ratio between the value of the investment program and the value of the project once completed (Majd & Pindyck, 1986, p. 10). In the context of financing, an option describes the reserved right to accept or deny a specific contractual offer (Heldt, 2018). Accordingly, if a manager has an option he obtains the right yet not the obligation to act (Jun, 2008, p. 13). The Real Option Theory is an evolution of the initially developed financial option theory (Jun, 2008, p. 13; Lee, 2011, p. 4445).The difference is that the Real Option refers to real assets whereas the financial option applies to financial assets (Tong & Reuer, 2007, p. 5). The rational of the Real Option Theory is to include the value of future flexibility into economic decisions (Schiel, 2019, p. 1). Therefore, a

Real Option model can provide managerial flexibility (Lee, 2011, p. 4445). If the designing and execution process of a project contains flexible options, additional value can be generated (Jun, 2008, p. 3). Hence, an option adds value to the rather passive net present value of a project (Benaroch & Kauffman, 2000, p. 200) Nevertheless, these options require appropriate assessment (ibid.). The price for an option is determined by the exercise price for the future investment needed to acquire an asset (Myers, 1977, p. 155). In the context of Real Options, dividends are the projects inherent opportunity costs if the project is being deferred (Villani, 2009, p. 2). Accordingly, the options value is defined by the value it carries when it expires as well as whether it is being exercised (Myers, 1977, p. 155). The application of the Real Option Theory, potentially, leads to the understanding, that risks associated with uncertainties may also represent an advantageous opportunity, which can be utilized (Li, 2007, p. 68). This risk originating from uncertainty is, within the Real Option Theory, expected to follow a stochastic process (Schiel, 2019, p. 104). The suitability of the process differs depending on the type of the uncertain parameter (ibid.).

The underlying question of a Real Option Analysis (ROA) is whether to invest, and if so, when to invest (Schiel, 2019, p. 98). Additionally, the ROA shall serve as a tool to precisely determine the value of an investment project (Lee, 2011, p. 4447). More specifically, of an investment project where the outcome is due to changing input parameters highly uncertain (Schiel, 2019, p. 95). Therefore, the ROA is particularly feasible for projects that require high managerial flexibility, or contain high overall risks (ibid.). Especially, because the ROA compares the value of managerial flexibility, to an irreversible investment option and takes uncertainties into account (Schiel, 2019, p. 96).

These uncertainties, and the respective valuation of the managerial uncertainties, reset investment barriers, by introducing an option theory to the evaluation of a project the upper price to perform an investment is in general increased, whereas the lower price to terminate an investment is decreased due to the consideration of the value of an option (Tauer, 2006, p. 340). Accordingly, the expected return of an investment needs to be even higher in order to convince an investor to execute the option (make the investment). Whereas the return also needs to be even lower to convince the investor to terminate the investment (ibid.). Therefore, the creation of an option as well as the option's value are severely influenced by the volatility of prices and thereby determine the costs to exit or enter an investment (ibid.).

As described, the value of an option is highly influenced by the expected returns the potential investment contains that indicates that an accurate reconstruction of all possible future scenarios with their respective probabilities of occurrence is essential (Rambaud & Pérez, 2016, p. 10). One of the most frequently used methods to undertake the simulation is the Monte-Carlo-Simulation (Schiel, 2019, p. 96). One of the major advantages of this simulation is, that it is rather flexible and therefore, can be adjusted to the specific needs of the application (Schiel, 2019, p. 106). That is especially, feasible for applications which do not fully comply with the characteristics of a standard financial option (ibid.).

3.2 Types of Real Options

3.2.1 List of Real Options

The current value of a company is determined by the option to make future investments (Myers, 1977, p. 148). These future investments also refer to specifically irreversible spending by firms, such as advertising, sales, improving efficiency, and incorporating new technologies (Myers, 1977, pp. 155-156). Seven types of Real Option can in general be identified, option to defer, time-to-build option, option to alter operating scale, option to abandon, option to switch, option to growth, and interaction among multiple Real Options (Lee, 2011, p. 4445). A brief explanation on each type is presented in the following table.

Table 4 – Seven types of Real Options

Type	**Rational**
Option to defer	Gives the management the opportunity to postpone to exercise an option, for instant on purchasing an asset.

Time-to-Build option	Throughout the process of building e.g. a production plant a reevaluation whether and how to continue can be undertaken at any stage.
Option to alter operating scale	Changing market conditions may cause an acceleration, expansion or even in case of negative developments a stop and restart of current production.
Option to abandon	In extreme cases, management can choose the option to sell operations.
Option to switch	Changing demands might lead to a change of the output or input to meet market requirements.
Option to growth	Describes creating the opportunity for future growth by investing for example in R&D.
Interaction among multiple Real Option	Enhancing options and protective options are often related and exercised at the same time to maintain financial flexibility.

Source: Own Table based on Lee, 2011, p. 4447

The following chapters will elaborate on selected options more precisely in order to provide a more detailed insight to what extend the specific option creates additional value. The option to abandon, along with the time to build option, and the option to defer, are the Real Options most used in business practice (Rambaud & Pérez, 2016, p. 2). Therefore, these options are among those which will be more closely elaborated on in the following.

3.2.2 Option to switch

An option to switch arises when two or more sources are available and can be substituted, that usually occurs, when the infrastructure for substitution is already given (Gatfaoui, 2015, p. 270). Whether to switch is to be determined by evaluating both options relatively to each other over a sufficient time horizon (ibid.). The switching decision serves as a tool to avoid higher costs than necessary, for instance an investor can switch between oil or gas as energy sources to utilize benefits from low prices and thereby reduced costs (Gatfaoui, 2015, p. 270). Accordingly, an innovative company framework provides the opportunity to save money by creating chances to utilize cost benefits resulting from options to switch (Gatfaoui, 2015, p. 279). Hence, providing the option to switch and thereby saving costs, this option creates a comparative advantage (ibid.). Furthermore, this comparative advantage, potentially translates into a strategy advantage (ibid.). Obviously, the more costs can be saved to the utilization of reduced costs, the higher the comparative advantage is (ibid.). Nevertheless, as stated the option to switch requires the respective infrastructure, this infrastructure needs to be build which depend upon an initial investment. Thus, the benefits resulting from the option to switch need to exceed the initial investment made in order to provide that option. (Gatfaoui, 2015, p. 280).

3.2.3 Time to Build Option

Traditional discounted cash flow (DCF) criteria, usually are inadequate to evaluate a spending project, as these projects only very rarely contain a fixed spending pattern which yet are required by those DCF methods (Majd & Pindyck, 1986, p. 8). Besides, DCF methods do not provide a realistic description of investment behavior under uncertainty (Majd & Pindyck, 1986, p. 8). The essential problem of a real project lies within the heterogeneity with respect to the units of capital that must be installed in a sequential matter and can be considered to be unproductive, therefore, do not generate and return until the project is completed (Majd & Pindyck, 1986, p. 10). Hence, it is necessary to introduce a scheme to evaluate the project at any stage and implement a decision rule which enables the decision maker whether to invest further or to exercise an alternative investment (ibid). Obviously, that decision rule needs to take into consideration what the underlying value would be today if the project was completed. Furthermore, it needs to assess the remaining expenditure necessary for completion, and the risks a potential completion delay contains with its respective costs as well as opportunity costs (ibid.). The results of that assessment will lead to the market value the respective project contains, that market value is the basis for the investor to evaluate the complete investment program and it is what a company with a value maximizing approach would pay for, and thereby execute the option (Majd & Pindyck, 1986, p. 14). Applying a concept of Real Option and implementing the time to build option, provides the opportunity, to adjust the patterns of expenditure, as new information arrive and thereby create greater flexibility, yet, that additional flexibility in general requires a higher investment (Majd & Pindyck, 1986, p. 25).

3.2.4 Option to defer

Introducing mechanisms to provide flexibility when facing uncertainties is essential to managers who are facing complex market conditions, but once introduced the option to defer a project becomes an alternative for managers (Henao & et al., 2017, p. 194). Within the concept of traditional DCF methods, managers are expected to take a passive position once they have made the decision to execute an investment project (Henao & et al., 2017, p. 195). This assumption does not hold true in a real world project, as managers are usually provided with the flexibility to reevaluate their investment decision and adjust their strategy in accordance to new information they receive and thereby changing the initially forecasted variables (Henao & et al., 2017, p. 195). Accordingly, reflecting this flexibility in project calculations makes sense when following prerequisites are fulfilled (Henao & et al., 2017, p. 195).

Table 5 – Prerequisites for managerial flexibility

1. The investment is expected to contain high costs as well as high benefits.
2. The investment made will be at least partially irreversible.
3. High amounts of the expected benefits rely on variables which are subject to high uncertainty.

Source: Own Table based upon Henao & et al., 2017, p. 195

As previously described an investment project can be split into several investment stages. Therefore, the value of the option must be determined at each stage (Henao & et al., 2017, p. 197). The determination of that value is the basis to further decide whether it is the right time to proceed with the respective investment or if it is more beneficial to defer the option to invest, until the conditions are more beneficial (ibid). Hence, a precise determination of the value of the option is essential, and the more planned stages of the project are covering the option to defer the more accurate evaluation can be made (ibid.).Furthermore, structuring the project in stages also provides the opportunity to invest prior to the expiration of the option (ibid.). The binominal tree is one tool which can serve to precisely value the option to defer at each stage until the conditions are appropriate (Henao & et al., 2017, p. 203). That also includes that decreasing uncertainty leads to the decision to not further invest in a project as it is expected to be no further beneficial (Henao & et al., 2017, p. 206). Therefore, introducing a Real Option concept to evaluate a project, compared to a traditional DCF-based approach may lead to a more precise evaluation and also to potentially better judgement by the managers even if it causes to ultimately abandon a project (Henao & et al., 2017, p. 206).

3.2.5 Option to Abandon

Taking advantage of the uncertainties that lie within a investment project by identifying and valuing those can increase the value of the investment itself (Rambaud & Pérez, 2016, p. 1). Hence, the value of a company can be increased through this valuation as well, especially, when these consideration of Real Options lead to strategic opportunities (ibid.). Even more so when these options provide the chance to act proactively (ibid.). The identification of these options yet is essential, as an option can only be executed if identified (ibid.). But once identified, these Real Options affect various parameters of the project and occur on several stages (ibid.). One of the respective presented option can be the option to abandon a project (Damaraju, Barney, & Makhija, 2015). The option to abandon a project, would mean to entirely liquidate an investment project, and thereby generating a cash flow representing the residual value of the project (ibid.). Accordingly, this residual value is considered to be the amount which can be regained throughout the projects lifespan, because by abandoning the project it can be sold (Rambaud & Pérez, 2016, p. 10). Thereby, the value of the project increases and the difficulties to terminate a project are reduced if it becomes a unbeneficial investment (Rambaud & Pérez, 2016, p. 2). It is within the nature of the option to abandon, as with an increasing maturity of the option the value rises (Rambaud & Pérez, 2016, p. 11).

Each of the previous chapters dealing with the different types of Real Options stated that a precise and diligent identification within a project is necessary to identify and evaluate the present Real Option. Therefore, the following chapters will present a variety of projects where Real Options can occur.

3.3 Scopes of Application

3.3.1 Environmental Investment Projects

One of the goals of investments in environmental projects is to reduce the environmental impact of the business operation (Schiel, 2019, p. 67). For instance, a corporation that indents to reduce the risk of future retrofit measures could oversize the investment into the emission reducing technology, and thereby cause a large scale of sunk costs (Schiel, 2019, p. 69). Furthermore, investments into environmental projects are often not profitable, in absolute numbers yet are still necessary, due to legal and political regulations (Schiel, 2019, p. 86). Hence, the decision is often not whether to invest or not, but rather in which of the options shall be invested and when (Schiel, 2019, p. 118). Accordingly, it is necessary to precisely evaluate the different options (ibid.). Traditional evaluation methods may not be suitable, as the revenues, are often not big enough to apply those traditional methods (Schiel, 2019, p. 91). Furthermore, these traditional measures, assume that all input parameters also those describing the future are certain (Schiel, 2019, p. 92). Yet, due to various external influences on environmental projects, for instance political influences the Real Option Analysis can especially be suitable (Schiel, 2019, p. 117). Within the scope of environmental investments, the theory can be used to evaluate when the respective investment, which is economically not advantageous has the least negative impact on the entity (Schiel, 2019, p. 118).

3.3.2 IT Projects

Technological projects, similar to projects of other nature, do not immediately generate a return after making an investment decision (Kauffmann, Liu, & Ma, 2015,

p. 157). Technological projects sometimes even tend to take an uncertain amount of time until they are completed (ibid.). Moreover, IT investments often carry various risks which need to be identified and managed properly (Benaroch, Lichtenstein, & Robinson, 2006, p. 4). Though for certain is that there will always be a time span between the first investment into the project and moment it generates the first benefits (Kauffmann, Liu, & Ma, 2015, p. 157). For managers it can be valuable to defer the decision to make an investment, as new information with respect to the value of the project will reveal over time (Kauffmann, Liu, & Ma, 2015, p. 164). Nevertheless, the longer management defers to make an investment decision, potential first mover benefits decrease (ibid.). Therefore, to make a decision whether to defer a decision or not needs to be carefully evaluated if there will be a difference on the potential returns of the respective project (ibid.).

3.3.3 Real estate and Infrastructure projects

There are various examples for options within real estate and construction projects, some of these are portrayed within the following. A building can be expanded if there is a major demand for a change by the architect or owner (Francis & Björnsson, 2004, p. 473). To save costs the option to reduce could be a possibility (ibid.). Utilizing new inventions could lead to the option to switch material if a better one becomes available (ibid.). If environmental circumstances change, the option to abandon a project could be taken into account (ibid.). Also, the option to defer is probable, especially if there are influences, which were not foreseeable (ibid.). Maintaining an infinite lifespan after having completed an investment project, especially a construction project, requires to keep the maximum rate of an investment to be flexible, and thereby providing the option to continuously invest (Majd & Pindyck, 1986, p. 14). This option should be accounted for when evaluating the project (ibid.).

3.3.4 Research and Development (R&D) Projects

Evaluating R&D projects with a Real Option Analysis instead of a conventional net present value approach provides the advantage, that the maintained managerial flexibility can be taken into consideration (Villani, 2009, p. 2). One key problem for managers in evaluating R&D projects, is that costs can often not be predetermined reliably, since as the project develops, the necessary funding might fluctuate as well

(Villani, 2009, p. 2). With changing market conditions, investing in R&D becomes more and more important to stay competitive, therefore, evaluating those projects reliably and accurately is of increasing importance as well (Villani, 2009, p. 2). Yet, usually investment opportunities into R&D projects are not available to one firm only but are accessible to various competitors (Villani, 2009, p. 16). On the one hand it is generally beneficial to generate a first mover advantage, by being the first market participant to invest in a R&D project, on the other hand it can be advantageous to wait and generate knowledge and avoid mistakes (Villani, 2009, p. 3). Especially considering that a product generally follows the concept of a product life cycle, which also includes a phase where demand is decreasing (Bollen, 1999, p. 9). Hence, assuming that a demand for a product will constantly grow decreases the option to contract or abandon a project (ibid.). Furthermore, it overvalues the option to expand for instance production capabilities (ibid.). R&D projects provide the opportunity, to increase its value with every investment stage, as it potentially generates information and new knowledge with every stage additionally that every reached stage reduces the uncertainty inherent in a R&D project (Majd & Pindyck, 1986, p. 8)

4 Evaluation of Datasets using the concept of the Real Option Theory

4.1 Investing in Data

According to Myers, 1977, p. 170, all discretionary future expenditure increase the value of a company. As many technology investments carrie severe risks with a serious potential of loss, managers seek for a way to control that risk and simultanouisly create IT driven business value (Kauffmann, Liu, & Ma, 2015, p. 154). An option not exercised must be valued as opportunity costs (Lee, 2011, p. 4445). To determine the value of an R&D investment, is rarely immediately possible because the potential return usually occurs later (Lee, 2011, p. 4447). The potential profit resulting from a R&D project is usually being realized in subsequent investment oportunities consequential from a successful R&D project (Lee, 2011, p. 4447). Furthermore, due to not being able to express the value of information and data on the balance sheet corporations have not introduced a methodologie how to accurately measure the value of their data respectively information (Tallon, 2013, p. 33). Especially, because IT hardware as well as software often are recognized as commodities, whereas the inherent information of the data emerges to be the actual asset (Tallon, 2013, p. 38). The determination of the value additionally requires to track how the value is developing over time (Tallon, 2013, p. 33). Nevertheless, the generated mass data is the key ressource and essential basis for future business models (Häckel & Übelhör, 2019, p. 174). This generated Data, needs to be stored and analyzed in a cloud based structure (Häckel & Übelhör, 2019, p. 174). The key is to invest in a sequential fashion and thereby benefiting from decreasing uncertainties (Tong & Reuer, 2007, p. 3).

4.2 Determining the potential benefits of analyzing data

In general, data only contains a value if the respective data is also being used (Otto & Österle, 2016, p. 36). The goal of analyzing data is to generate new knowledge and thereby, utilize a benefit, which reduces costs, helps to access new markets or provides a better basis for quicker decisions (Papp, 2019, p. 15). Furthermore, a sophisticated data basis serves to develop new digital based business models (Häckel & Übelhör, 2019, p. 167). Yet, for entities, it is especially difficult to quantify the underlying costs for data and data quality as these costs often originate from indirect sources (Eppeler & Helfert, 2004, p. 311). Moreover, determining whether these indirect costs are actually the result of low data quality or have other origins is particularly difficult (Eppeler & Helfert, 2004, p. 311). For industrial corporations though analyzing data contains the potential benefit to transfer the existing business model of offering a product to a model where they offer complete solutions, based upon the data they have generated through the IOT and analyzed respectively (Häckel & Übelhör, 2019, p. 169). This transfer will gain even more importance, due to constantly decreasing margins which are caused by an increasing competitive pressure originating from further automation and globalized value chains (Häckel & Übelhör, 2019, p. 169). Therefore, the value of analyzing data and thereby generating new knowledge does not necessarily lie within the possibility of selling the existing product. But proactively developing a concept to target the customer with a data based service and thus generate a competitive advantage increases the value (Häckel & Übelhör, 2019, p. 169). Furthermore, not only the existing customer base can be targeted, but based on the collected data, a corporation can reach out to new target groups (Häckel & Übelhör, 2019, p. 172). These multi-sided markets provide entities with the opportunity to have several target groups due to their data base

(Häckel & Übelhör, 2019, p. 173). For instance, a producing company can issue their energy usage data to an energy company to provide them with information to optimize their energy supply (Häckel & Übelhör, 2019, p. 173). The Implementation of a data based business model also leads to a dependency on data created by their customers, and that the resulting information is reliable (Häckel & Übelhör, 2019, p. 169).

Nevertheless, the risk of evaluating an option based upon the false assumption, due to weak data maintains (Tong & Reuer, 2007, p. 9). In order to reduce that risk, companies need to get to know their customers, and the individual needs of the customers even more thoroughly (Häckel & Übelhör, 2019, p. 173). This required increasing communication with the customer is possible through intensified digital based interaction, which is established by the nature of the digital service offered (Häckel & Übelhör, 2019, p. 173).

4.3 Real Option and Data Quality

A key characteristic of a Real Option Analysis is the absence of at least one or more parameter, respectively, the uncertainty of that parameter (Schiel, 2019, p. 104). The risk resulting from that uncertainty needs to be evaluated and managed, for which the Real Option Analysis can serve as a method (Benaroch, Lichtenstein, & Robinson, 2006, p. 4). Within an environment of sequential decision making processes the ROA serves as a useful tool (Francis & Björnsson, 2004, p. 480). Entities, often struggle to quantify the return they can expect on an investment they have made into a data quality initiative (Eppeler & Helfert, 2004, p. 311), or IT projects, as they often require a high amount of flexibility for managers to make the most economically beneficial decision which, therefore, also requires a flexible evaluation method (Francis & Björnsson, 2004, p. 480). The traditional methods of capital budgeting are problematic, as they do not take the possibility of subsequent decisions after the initial decision has been made into account (Francis & Björnsson, 2004, p. 473). Yet one key aspect of investing into data is the subsequent evolution of costs throughout the project (Eppeler & Helfert, 2004, p. 312). Determining the value of such a project is precisely difficult, as the return resulting from a data quality initiative predominantly originates from costs, which will have been saved due to the increased quality in data and are, therefore indirect costs (Eppeler & Helfert, 2004, p.

311). To measure these cost savings proactively is extremely difficult and is usually based upon assumptions (Eppeler & Helfert, 2004, p. 311). Nevertheless, it can be expected that if the costs to improve data quality, then the indirect costs as a consequence from poor data quality decrease (Eppeler & Helfert, 2004, p. 319).

4.4 Real Option and Data Sets

With an increasing potential of future knowledge and information, the value of an option raises (Benaroch & Kauffman, 2000, p. 202). That indicates, that if the underlying assets value contains a high uncertainty the options value is respectively higher (ibid.). Therefore, the value of the option lies within the asymmetry of the right to capture the upside, without the obligation to bear the downside of an investment (Francis & Björnsson, 2004, p. 473). Furthermore, the value of uncommitted financial resources is determined by what a company can gain by waiting, and by the risk that a firm loses the option to invest to a competitor (Leahy, 1993, pp. 1105-1106). According to (Leahy 1993, p. 1106), the option to remain is even worthless in a market of perfect competition, because if the investment is put on hold it will immediately be eliminated by the competitors. Accordingly, while the initial model that an option maintains its value if put on hold, proves to be right for some investment projects, it is not applicable for highly competitive industries where the option will be executed by the competition (Tauer, 2006, p. 340). As previously pointed out the value of data is extremely hard to determine prior to analyzing it and therefore contains a high degree of uncertainty. The value of the option lies within the risk that is managed by the option, subsequently, if there is no risk the option does not carry any value (Benaroch, Lichtenstein, & Robinson, 2006, p. 10).

4.5 Real Option and Data Sets cost components

The value of a Real Option increases with an increasing uncertainty, regardless of the origin the uncertainty has (Schiel, 2019, p. 116). The Real Option is defined by five major components (Lee, 2011, p. 4448). The following table lists these five components. Furthermore, it provides a short explanation of each of the components, when the third column where that component is seen in a project investing into data.

Table 6 – Real Option Components in Data Sets

Real Option Component	Explanation	Data Sets
The Underlying Price	Describes the present value of the assets which need to be acquired to realize the project	Are reflected within the costs to acquire and store the data
Exercise Price	Refers to the investment needed to transfer the investment opportunity into the underlying asset	Reflects the investment, which needs to be made to make the data useful.
The Time to Maturity	Defines the maximum time the project can be postponed without losing the opportunity	The time until which the new knowledge will bring, for instance, a competitive advantage.
Risk-Free-Rate	Generally, the 10-year-government bond describes the risk-free-rate of return	n.a.
Volatility	Refers to the uncertainty of the future expected cash flows generated from the projects	The Investment into data contains a high degree of uncertainty with respect to the cash flows

Source: Own Table based upon Lee, 2011, p. 4448; Coyne & et al., 2018, pp. 154-155

4.6 The Value of Data applying the concept of a Real Option

According to (Lee, 2011, p. 4449), the value of a project increases if the underlying price, time to maturity, risk-free rate, and volatility increase whereas they cause a decrease of value if the exercise price is raising. While managers, often use options to mitigate negative risks, the Real Option can also provide an impact on realizing positive options by presenting potential beneficial impacts of a follow up investment for a current project (Benaroch, Lichtenstein, & Robinson, 2006, p. 22). The value of the underlying assets partially determines the value of the respective option in addition to the time until that option expires (Benaroch & Kauffman, 2000, p. 202). If the underlying assets value is flexible, then it can be more beneficial to not exercise the option (ibid.).

Throughout, the process of making decisions, it can be expected, that managers have the potential to learn more about the uncertain amount, the project will have as a return. This change in value describes the value of the deferral option (Benaroch & Kauffman, 2000, p. 202). Therefore, the decision to defer an investment needs to be revalued with any further information that is provided respectively learned (Benaroch & Kauffman, 2000, p. 204).

The entity can either hold the option or it can hold the underlying asset, which means that they have exercised the option (Benaroch & Kauffman, 2000, p. 204). The precisely executed evaluation of a Real Option can lead to contradicting conclusions, to those, which have been drawn by traditional evaluation methods (DCF methods). That means, that an investor might decide not to invest into a project even though the calculated Net Present Value (NPV) of the respective project is positive, because the embedded option to defer the project is sufficiently valuable, and therefore, the opportunity costs of investing in the current period are significant (Tong & Reuer, 2007, p. 10). Following this rational the conclusion drawn by the investor when the NPV is negative might not necessarily be to deny the investment, because the embedded growth options which could be exercised can be sufficient and therefore, beneficiary (Tong & Reuer, 2007, p. 10).

A further benefit of the Real Option Theory is that it is capable of recognizing asymmetric payoff structures of an investment. The several options which are by nature included in a project enable firms to reduce potential downside risks while at the same time utilizing upside opportunities (Tong & Reuer, 2007, p. 10). These options are the origin for the potential asymmetric performance outcomes of the respective project. Because, following that logic it means that an increasing uncertainty with the therefore, increasing number of options respectively decisions is creating an increasing potential value to the holder of the option, while downside losses are contained (Tong & Reuer, 2007, p. 11). Hence, a substantial proportion of the value of a project is potentially generated by its underlying uncertainties, and the thereby caused flexibility. Accordingly, the theory emphasizes that maintaining the flexibility is very valuable (Tong & Reuer, 2007, p. 11). Nevertheless, the value of these options significantly differs across firms and industries. Therefore, it is essential to precisely identify the origin of the respective options (ibid.). Yet, once that assessment has been conducted with the required due diligence, the Real Option Analysis can lead to a reevaluation of an entities resource allocation (ibid.).

The Real Option Theory contains the opportunity to value sequential decisions, whereas traditional investment models lack this usability (ibid.). Accordingly, valuing a project provides the management with an opportunity to incorporate strategic reality into capital budgeting. Usually, corporations especially show incapability to manage unsuccessful projects, rather than successful projects, Real Options though provide the opportunity to for instance abandon a project (ibid.). Therefore, the Real Option can at least serve as an analysis to define the value of a specific uncertainty with its containing options the corporation has due to that uncertainty (Tong & Reuer, 2007, p. 19). Obviously, valuing a project with the concept of Real Option Analysis might therefore, lead to a more positive view, as it statues uncertainty as the source of value which cause a more constructive approach (Tong & Reuer, 2007, p. 19).

5 Conclusion

5.1 Summary

This bachelor thesis' goal was to assess whether the Real Option analysis and its concept can be applied to evaluate data and data sets. Therefore, an assessment of the necessity of data is crucial and has within this project been in depth undertaken throughout the chapter 2 "Digitalization" with its referring subchapters. Hence, a detailed representation of the ideas, concepts, and components is provided. It is pointed out, that digitalization is not limited to the process of transferring existing analog business processes and models into a digital presentation. If digitalization is understood in a broader context, it needs to be valued as a disruptive technology, it includes to question and even more so to rethink the current processes and models to develop a new business model. Obviously, developing a new business model which is the eventual goal of digitalization, is a process itself. Therefore, the chapter 2.2.1 “Layers of Digitalization” elaborates on the different stages of this process, which indicates that the initial step is to implement an infrastructure that actually allows the entity to collect the necessary information on which that new model can be based upon. The subsequent two layers describe the means, how to process the data and eventually to the development of new knowledge.

One of the predominantly discussed evolutions of digitalization, is the implementation of Industry 4.0 respectively the development of a Cyber Physical System. As pointed out throughout the chapters 2.2 and 2.2.1. These networks enable each component within the line of production to communicate with each other. The real time communication of for instance machines with one another provides the entity to satisfy consumers with individual solutions while still utilizing scale effects and thereby offering competitive prices. In order to enable the machines to communicate within a cyber physical system they need to be equipped with the necessary hardware as well as software. This Internet of Things will generate data which will have to be managed at various stages. Therefore, the chapters 2.3 and following elaborate on data quality and data quality management.

Essential to data quality management is to define standards which data and data sets have to fulfill to be of high quality. These factors are various, and each data set has to be continuously controlled to ensure the accuracy of the information drawn from the data analysis.

Managing that data quality, furthermore, requires a process that can be transparently followed, especially, as the amounts of data get even bigger and result in an analysis which is referred to as Big Data. Big Data enables a system by implementing AI to analyze unstructured and unconnected data to generate new information. But again, for the information to be accurate, the underlying data must be of expectable quality.

Generally, the quality of data is to be valued within the context of its usage. Therefore, the quality of data is determined by the information the data provides. As pointed out the amounts of data are enormous, and storing these amounts traditionally would require significant investments, therefore, part of digitalization often includes the implementation of Cloud Computing which can considerably reduce the costs for storage and furthermore, provide additional benefits throughout the process of analyzing data. Nevertheless, maintaining the data quality respectively information quality requires an investment. This investment needs to be valued in order to determine its profitability.

Evaluating these projects, is especially difficult as traditional DCF methods do not account for the flexibility and uncertainty investment projects like these carry. Therefore, the chapters 3 and 4 are elaborating on an alternative evaluation method, the Real Option Analysis. The Real Option Analysis is an evolution of the financial option, yet instead of investing in a financial product the respective entity invests in a real asset.

By pointing out the different scope of applications chapter 3 illustrates the flexibility and applicability of the Real Option Analysis. It shows that even though it is a complex analysis it can be applied to various investment projects. Yet the chapter also illustrates that applying this concept requires an extremely thorough analysis of the project.

Subsequent to the more general overview Chapter 4 is more specifically pointing out how the Real Option Analysis can help to determine the value of a data and whether it is profitable or not. It again refers to the essentials of data sets. Ultimately it shows, how the different key aspects of the Real Option Analysis are met within an investment project into data and data sets in order to generate new information and knowledge. Conclusive to the examined analysis the Real Option Theory can serve as a suitable tool to more accurately determine the value of data and data sets.

5.2 Critical acclaim

This thesis elaborates on some aspects of digitalization, Industry 4.0, and data. Yet, the complexity of each of the topics is not and cannot be portrayed in this work. Therefore, this work does not claim to reflect a complete picture. Especially the closely related topic of data security is not concerned within this paper. Nevertheless, when developing a new data driven business model data security compliance should be one major aspect of the process, as ignoring it might lead to a complete failure of the model due to legal requirements.

The elaboration on applying a Real Option Analysis to the given problem is also a rather artificial one within this work. The goal was not to develop a mathematical formula through which the value of a respective investment could be determined. Therefore, the work solely reflects on the underlying concept of the the ROA. However, in order to make this analysis applicable for businesses a precise mathematical basis is essential. Hence, this work might be suitable to constitute a starting point for further research on how to determine the value of a data set by applying the Real Option Analysis, but not a framework of how it can be applied.

On top of that, this work does not reflect on why the Real Option Analysis is rarely applied within the business world, even though it was initially developed within the 1980’s and ever since been continuously improved. Hence, a theoretical applicability does not necessarily lead to a real implementation by businesses. In order to solve this discrepancy an in depth analysis why it is not applied by businesses needs to be undertaken.

5.3 Outlook

Throughout the research for this paper, and especially within the chapter on digitalization it has become clear that data and the analysis of specific data will become even more essential to a business than it already is. In fact, the importance will raise to a level where it will no longer be the question what the benefits of an analysis will be, but the question will be, if the company can afford not to invest into data but if the company will be able to survive if it does not invest into data. Nevertheless, an actual further digitalization requires an infrastructure, which is capable of transmitting those huge amounts of data. Therefore, prior to investing in tools which will be capable of analyzing the huge amounts of data such as AI, Cloud Computing and Big Data solutions, the investment in for instance 5 G technology is essential.

Obviously, especially within Germany data security plays a huge role and will also have a real impact in the future on any development. Among others, these factors, the reluctant political will to support and build the digital infrastructure as well as the general high emphasize on data security, with its partially misunderstood meaning of data security have been hindering to new developments and innovative ideas. Yet writing this paper in times of corona, and thereby experiencing an increase of willingness to open up to these topics, the crisis could function as a catalyst. Hence, we might experience a digital based innovativeness throughout the coming years, and ultimately understand and benefit from its disruptive nature.

VI List of references

Anderie, L. (2018). *Gamification, Digitalisierung und Industrie 4.0.* Liederbach: Springer Gabler.

Barton, T. (2014). *E-Business mit Cloud Computing.* Worms: Springer Vieweg.

Baun, C., Nimis, J., Kunze, M., & Tai, S. (2011). *Cloud Computing Web-basierte dynamische IT-Services.* Heidelberg: Springer-Verlag.

Benaroch, M., & Kauffman, R. (2000, June 15). Justifying Electronic Banking Network Expansion Using Real Option Pricing: An Empirical Illustration. *MIS Quarterly*, pp. 197-225.

Benaroch, M., Lichtenstein, Y., & Robinson, K. (2006). *Real Option in IT Risk Management: An Emprirical Validation of Risk-Option Relationships.* Syracuse: MIS Quarterly.

Bendel, O. (2018, 02 19). *Gabler Wirtschaftslexikon*. Retrieved from Gabler Wirtschaftslexikon: https://wirtschaftslexikon.gabler.de/definition/big-data-54101/version-277155

BmBF. (2017). Industrie 4.0 Innovationen für produktion von morgen. *Industrie 4.0 Innovationen für produktion von morgen*. Berlin, Berlin, Deutschland: Bundesministerium für Bildung und Forschung.

Bollen, N. (1999). Real Opütions and Product Life Cycles. *Management Science*, pp. 670-684.

Borges, G. (2018). Einführung: Herausforderungen an das Identitätsmanagement im Cloud Computing. In G. Borges, & B. Werners, *Identitätsmanagement im Cloud Computing* (pp. 1-11). Berlin: Springer .

Brauckmann, O. (2019). *Digitale Revolution in der industriellen Fertigung - Denkansätze.* Lüdenscheid: Springer Vieweg.

Coyne, E. M., & et al. (2018, 05 02). Big Data information Governance by accountants. *International Journal of Accounting & Information Management*, pp. 153-170.

Damaraju, N., Barney, J., & Makhija, A. (2015, February 01). Real Options in Divestment Alternatives. *Strategic Management Journal*, pp. 728-744.

Decker, C., & Werner, R. (2016). *Academis research and writing a concise introduction.* Frankfurt am Main: iCADEMICUS GmbH.

Deckert, R. (2019). *Digitalisierung und Industrie 4.0.* Hamburg: Springer Gabler.

Ebert, B. (2018). *Prozessoptimierung bei Industrie 4.0 durch Risikoanalyse.* Groß-Zimmern: Springer Vieweg.

Engelmann, F., & Großmann, C. (2011). Was wissen wir über Informationen. In K. Hildebrand, & e. al., *Daten- und Informationsqualität* (pp. 3-23). Wiesbaden: Springer Vieweg.

Engelmann, F., & Großmann, C. (2018). Was wissen wir über Informationen. In K. Hildebrand, & et. al., *Daten und Informationsqualität* (pp. 3-23). Freising: Springer Vieweg.

Eppeler, M., & Helfert, M. (2004). *A Classification and Analysis of Data Quality Costs.* Lugano: International Conference on Information Quality.

Erner, M. (2019). *Management 4.0 - Unternehmensführung im digitalen Zeitalter.* Heilbronn: Springer Gabler.

Fleischmann, A., & et. al. . (2018). *Ganzheitliche Digitalisierung von Prozessen.* Wiesbaden: Springer Vieweg.

Francis, P., & Björnsson, H. (2004, June). Using real option and decision analysis to evaluate investments in the architecture, construction and engineering industry. *Construction Management and Economics*, pp. 471-482.

Frederick, D. E. (2016, May). Libraries, data and the fourth industrial revolution (Data Deluge Column). *Library Hi Tech News*, pp. 9-12.

Gatfaoui, H. (2015, September 8). Pricing the (European) option to switch between two energy sources: An application to crude oil and natural gas . *Energy Policy*, pp. 270-283.

Granig, P., Hartlieb, E., & Heiden, B. (2018). *Mit Innovationsmanagement zu Industrie 4.0.* Villach: Springer Gabler.

Häckel, B., & Übelhör, J. (2019). Digitale Geschäftsmodelle in der Industrie: Eine Analyse der Auswirkungen und Herausforderungen. In S. Meinhardt, & A. Pflaum, *Digitale Geschäftsmodelle Band 2* (pp. 167-183). Nürnberg: Springer Vieweg.

Haug, A., Zachariassen, F., & van Liempd, D. (2011, January). The costs of poor data quality. *Journal of Industrial Engineering and Management (JIEM)*, pp. 168-193.

Heinrich, B., & Klier, M. (2011). Datenqualitätsmetriken für ökonomisch orientiertes Qualitätsmanagement. In K. Hildebrand, M. Gebauer, H. Hinrichs, & M. Mielke, *Daten und Informationsqualität* (pp. 47-66). Innsbruck: Springer Vieweg.

Heldt, C. (2018, 02 19). *Gabler Wirtschaftslexikon*. Retrieved from https://wirtschaftslexikon.gabler.de/definition/option-42018/version-265373

Henao, A., & et al. (2017, May 01). What is the value of the option to defer an investment in transmission Expansion Planning? An estimation using Real Options. *Energy Economics*, pp. 197-207.

Hillman, T., Zhang, N., & Jin, Z. (2018, May 4). Real-option valuation in a finite-time, incomplete market with jump diffusion and investor-utility inflation. *Risks*, pp. 1-20.

Hinrichs, H. (2002). *Datenqualitätsmanagement in Data Warehouse-Systemen.* Oldenburg: Orldenburger Forschungs- und Entwicklungsinstitut für Informatik-Werkzeuge und -Systeme (OFFIS).

Hofmann, J. (2018). *Arbeit 4.0 - Digitalisierung, IT und Arbeit.* Stuttgart: Springer Vieweg.

Hofmann, J. (2018). Ausgewählte Technologische Grundlagen. In L. Fend, & J. Hofmann, *Digitalisierung in Industrie-, Handels- und Dienstleistungsunternehmen* (pp. 3-28). Ingolstadt: Springer Gabler.

Huber, W. (2018). *Industrie 4.0 kompakt Wie Technologien unsere Wirtschaft und unsere unternehmen verändern.* Haar: Springer Vieweg.

IBM. (2020, 01 28). *IBM Big Data & Analytics Hub*. Retrieved from https://www.ibmbigdatahub.com/infographic/four-vs-big-data

Jun, J. (2008). *Valuation of Governmental.* Texas: Texas A&M.

Kauffmann, R. J., Liu, J., & Ma, D. (2015). Technology investment decision-making under uncertainty. *Inf. Technological Management*, pp. 153-172.

Klingebiel, N. (2019). Shared Services & Digitalisierung. In T. Kümpel, K. Schlenkrich, & T. Heupel, *Controlling und Innovation 2019* (pp. 156-169). Gelsenkirchen: Springer Fachmedien.

Knorre, S., Müller-Peters, H., & Wagner, F. (2020). *Die Big-Data-Debatte.* Wiesbaden: Springer Gabler.

Kümpel, T., Schlenkrich, K., & Heupel, T. (2019). *Controlling & Innovation 2019.* Wiesbaden: Springer Gabler.

Leahy, J. V. (1993, November). Investment in Competitive Equilibrium: The Optimality of Myopic Behavior. *The Quarterly Journal of Economics*, pp. 1106-1133.

Lee, S.-C. (2011, 09 15). Using real option analysis for highly uncertain technology investments: The case of wind energy. *Renewable and Sustainable Energy Reviews*, pp. 4443-4450.

Li, J. (2007). Real Options theory and International Strategy: A critical review. *Advances in Strategic Management*, pp. 67-101.

Maisch, B., & Palacios Valdés, C. (2018). Kundenzentrierte Digitale Geschäftsmodelle. In L. Fend, & J. Hofmann, *Digitalisierung in Industrie-, Handels- und Dienstleistungsunternehmen* (pp. 29-52). Ingolstadt: Springer Gabler.

Majd, S., & Pindyck, R. (1986, June). Time to Build, Option Value and Investment Decisions. *Journal of Financial Economics*, pp. 7-27.

Mertens, P., Barbian, D., & Baier, S. (2017). *Digitalisierung und Industrie 4.0 - eine Relativierung.* Nürnberg: Springer Vieweg.

Morbey, G. (2011). *Datenqualität für Entscheider in Unternehmen.* Wiesbaden: Springer Gabler.

Moriarty, J., & Palczewski, J. (2019, 04 11). Imbalance Market Real Options and the Valuation of Strorage in Future Energy Systems . *Risks*, pp. 7-39.

Myers, S. C. (1977, 09). Determinants of Corporate Borrowing. *Journal of Financial Economics*, pp. 147-175.

Obermaier, R. (2019). Industrie 4.0 und Digitale Transformation als unternehmerische Gestaltungsaufbabe. In R. Obermaier, *Handbuch Industrie 4.0 und Digitale Transformation* (pp. 3-46). Passau: Springer Gabler.

Ortner, B., Papp, S., & Meir-Huber, M. (2019). Einleitung. In S. Papp, W. Weidinger, M. Meir-Huber, B. Ortner, G. Langs, & R. Wazir, *Handbuch Data Science Mit Daenanalyse und Machine Lerning Wert aus Daten generieren* (pp. 1-14). München: Carl Hanser Verlag.

Otto, B., & Österle, H. (2016). *Corporate Data Quality.* Dortmund: Springer Gabler.

Papp, S. (2019). Grundlagen Datenplattformen. In S. Papp, W. Weidinger, Meir-Huber, B. Ortner, B. Langs, & R. Wazir, *Handbuch Data Science Mit Datenanalysen und Machine Learning Wert aus Daten gegenrieren* (pp. 15-102). München: Carl Hanser Verlag.

Rambaud, S., & Pérez, A. M. (2016, January 17). Assessing the Option to Abandon an Investment Project by the Binomial Options Pricing Model. *Creative Commony Attribution License*, pp. 1-12.

Rohweder, J. P., & et al. (2018). Informationsqulalität - Definitionen, Domensionen und Begriffe. In K. Hildebrand, & e. al., *Daten- und Informationsqualität auf dem Weg zur Information Excellence* (pp. 23-46). Mannheim: Springer Fachmedien.

Rohweder, J., & et. al. (2015). Informationsqualität - Definitionen, Dimensionen und Begriffe. In K. Hildebrand, M. Gebauer, H. Hinrichs, & M. Mielke, *Daten- und Informationsqualität auf dem Weg zur Information Excellence* (pp. 25-48). Mannheim: Springer Vieweg.

Schiel, C. (2019). *Real option based apprasal on environmental investments: An assessment of NOx emission control techniques in large combustion plants.* Karlsruhe: KIT Scientific Publishing.

Schildhauer, T., & et al. (2019). Schlüsselfaktoren der Digitalisierung - Entwicklungen auf dem Weg in die digitale Zukunft. In M. Stumpf, *Digitalisierung und Kommunikation* (pp. 13-34). Frankfurt am Main: Springer VS.

Schlotmann, R. (2018). *Digitalisierung auf mittelständisch.* Frankfurt: Springer Vieweg.

Sendler, U. (2018). *Das Gespinst der Digitalisierung.* Wiesbaden: Springer Fachmedien.

Stumpf, M. (2019). Europäische Kulturen in der Wirtschaftskommunikation. In N. Janich, C. Schmidt, & D. Neuendorff. Frankfurt: Springer.

Tallon, P. (2013, June). Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost. *IEEE Computer Society*, pp. 32-38.

Tauer, L. W. (2006). *When to Get In and Out of Dairy Farming: A Real Option Analysis.* New York: Agricultural and Resource Economics Review.

Tomanek, D., & Schröder, J. (2018). *Value Addad Heat Map.* Ingolstadt: Springer Gabler.

Tong, T., & Reuer, J. (2007, March). Real Options in Strategic Management. *Advances in Strategic Management*, pp. 3-28.

Villani, G. (2009, 07). A strategic R&D investment with flexible development time in real option game analysis. Munich: Economic Studies and Ifo Institute (CESifo).

Wagner, R. M. (2018). *Industrie 4.0 für die Praxis.* Langenneufnach: Springer Gabler.

Wiegand, B. (2018). *Der Weg aus der Digitalisierungsfalle.* Muelheim an der Ruhr: Springer Fachmedien.

Wirtschaft, I. a. (2019, 02 14). Datemanagement: Es hapert an der Bewertung. *Datemanagement: Es hapert an der Bewertung*. Köln, Nordrhein-Westfalen, Deutschland: Wirtschaft, Informationen aus dem Institut der deutschen.

Wolff, D., & Göbel, R. (2018). *Digitalisierung Segen oder Fluch?* Hof: Springer Gabler.

VII Declaration of originality

I hereby declare that the submitted thesis and the work reported herein was composed by and originated entirely by me without further assistance. Appropriate credit has been given where reference has been made to the work of others. The thesis was not examined before, nor has it been published. The submitted electronic version of the thesis matches the printed version.

[redacted]