Atte Koskimäki

Data quality challenges experienced by specialists in reporting and analytics

School of Technology and Innovations
Master's thesis in Economics and Business Administration
Industrial Management

Vaasa 2023

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Atte Koskimäki
Title of the Thesis: Data quality challenges experienced by experts in reporting and analytics
Degree: Master of Economics and Business Administration
Programme: Industrial Management
Supervisor: Dr. Rayko Toshev
Year: 2023
Pages: 72

ABSTRACT:

Research shows that data quality as crucial element to companies and specialists using analytics tools in their daily work. Issues experienced by experts in the reporting and analytics field and best tools and techniques to counter the bad data in big companies tend raise costs indirectly.

This study aims to aggregate useful framework for specialists in big companies using analytics tools in their daily work. Primary data was collected from 14 respondents with qualitative methods by executing a survey with open-ended questions and analysing data by applying thematic analysis. Answers were reviewed using respondent validation. Sample size was limited using voluntary and purposive sampling method by searching specialists using analytics tools for decision-making and who expressed interest in participating in the study. Search and contact were conducted using LinkedIn. Secondary data was collected by introducing previous similar studies, journals, as well as information from books and articles related to business intelligence, data, decision-making and data quality.

Main findings of the study show the key data quality challenges are related to process ownership, user awareness and wrong data formats. Results were presented in word cloud figures and analyzed in the text. The answers show that optimizing processes, establishing process ownerships, raising user awareness, defining clear procedures and training personnel about data importance and quality as important factors countering bad data quality.

Main contributions of the thesis are the unique experiences of specialists using analytics tools for decision-making, their opinions and suggestions to solve data quality problems in analytics. Potential research questions for future case studies in companies recognizing the issue and practitioners interested in the topic to study it in more specified and quantified context. Broader implications include practical information and tools for managers mainly in the business sector countering data quality issues in reporting and analytics field, using the summarizing Ishikawa diagram in the conclusion.

KEYWORDS: data, data quality, reporting, business intelligence, decision-making, analytics tools, data analysis

Vaasan yliopisto
Tekniikan ja innovaatiojohtamisen yksikkö
Tekijä: Atte Koskimäki
Tutkielman nimi: Asiantuntijoiden kokemat datan laatuhaasteet analytiikassa ja raportoinnissa
Tutkinto: Master of Economics and Business Administration
Oppiaine: Industrial Management
Työn ohjaaja: Rayko Toshev
Valmistumisvuosi: 2023
Sivumäärä: 72

ABSTRAKTI:

Tutkimukset osoittavat, että datan laatu on keskeinen tekijä yrityksille ja asiantuntijoille, jotka käyttävät analytiikkatyökaluja päivittäisessä työssään. Raportoinnin ja analytiikan asiantuntijoiden kokemat ongelmat sekä parhaiden työkalujen etsiminen huonon laadun torjumiseksi nostavat suurien yritysten kustannuksia epäsuorasti.

Tämän tutkimuksen tavoitteena on koota hyödyllinen viitekehys suuryritysten asiantuntijoille, jotka käyttävät analytiikkatyökaluja päivittäisessä työssään. Ensisijaiset tiedot kerättiin 14 vastaajalta laadullisin menetelmin suorittamalla kysely avoimilla kysymyksillä ja analysoimalla aineistoa teema-analyysiä. Vastaukset tarkistettiin käyttämällä vastaajavalidointia. Otoskokoa rajoitettiin vapaaehtoisella ja tarkoituksenmukaisella otantamenetelmällä etsimällä asiantuntijoita, jotka käyttivät analytiikkatyökaluja päätöksenteossa ja jotka ilmaisivat kiinnostuksensa tutkimukseen. Yhteydenotto tehtiin LinkedInin avulla. Toissijaista dataa kerättiin analysoimalla aikaisempia vastaavia tutkimuksia, lehtiä sekä tietoa kirjoista ja artikkeleista, jotka liittyvät dataan, päätöksentekoon ja datan laatuun.

Tutkimuksen tärkeimmät havainnot osoittavat, että keskeiset tiedon laatuhaasteet liittyvät prosessien omistajuuteen, käyttäjien tietoisuuteen ja vääriin tietomuotoihin. Tulokset esitettiin sanapilvikuvioissa ja analysoitiin teoksessa. Vastaukset osoittavat, että prosessien optimointi, prosessien omistajuus, käyttäjien tietoisuuden lisääminen, selkeiden menettelytapojen määritteleminen ja henkilöstön kouluttaminen tiedon tärkeydestä ja laadusta ovat tärkeitä tekijöitä huonon tiedon laadun torjunnassa.

Opinnäytetyön tärkeimmät panokset ovat analytiikkatyökaluja päätöksenteossa käyttävien asiantuntijoiden ainutlaatuiset kokemukset, heidän mielipiteensä ja ehdotuksensa analytiikan tiedonlaatuongelmien ratkaisemiseksi. Työ ehdottaa myös mahdollisia tutkimuskysymyksiä vastaaviin aiheisiin ongelman tunnistavissa yrityksissä ja aiheesta kiinnostuneissa tutkijoissa. Laajemmat vaikutukset sisältävät käytännön tietoa ja työkaluja pääosin yrityssektorin johtajille tiedon laatuongelmien ratkaisemiseksi raportoinnin ja analytiikan alalla käyttäen esimerkiksi työn Ishikawa-kaaviota datahaasteiden lyömiseen.

KEYWORDS: data, datan laatu, raportointi, liiketoimintatiedonteko, päätöksenteko, analytiikkatyökalut, data-analyysi

Contents

1 Introduction 7
1.1 Background 7
1.2 Research purpose and research question 10
1.3 Limitations and delimitations 11
2 Literature review 12
2.1 Data 12
2.2 Data quality 14
2.3 Business Intelligence 21
2.4 Data Quality Assessment Framework 23
2.5 Theoretical summary 26
3 Methodology 28
4 Empirical research 31
4.1 Qualitative survey 31
4.2 Demographics results 32
4.3 Survey questions and results 35
4.3.1 When do you use business intelligence tools for data analysis or decision-making? 35
4.3.2 What kind of data quality issues do, or have you face(d)? 37
4.3.3 What type of tools or techniques can be used to solve data quality issues? 40
4.3.4 What type of tools or techniques have you used at work? 42
4.3.6. How do you feel the data quality effects on daily operational and/or periodical tactical decision-making? 45
4.3.7 What would you define as best practices to counter bad data quality (or to improve data quality?) 47
4.3.8 Additional comments 49
4.4 Results summary 49
5 Discussion 51
5.1 Primary and secondary data 51
5.2 Key findings 52

5.3 Data quality factors 53

5.4 Business intelligence factors 55

5.5 Links to theoretical framework and decision-making 56

6 Conclusions 59

6.1 Summary 59

6.2 Managerial implications 60

6.3 Future research opportunities 62

7 References 63

Appendix 1 – Data collection cover letter 69

Appendix 2 - Qualitative survey questions 71

Figures

Figure 1 Evolution of Business Intelligence (Sharda et al. 2018, p. 43) 8
Figure 2 The data life cycle illustrated by Mahanti (2019, p.15).................................. 13
Figure 3 Dimensions, Measurement Types, and Specific Metrics (Sebastian-Coleman, 2013, p. 48).. 17
Figure 4 Causes of bad quality illustrated by Mahanti (2019, p.15). 18
Figure 5 “Total costs incurred by data quality on the company” (Haug et al. 2011)..... 19
Figure 6 “Four types of costs incurred by poor quality data” (Haug et al., 2011) 20
Figure 7 The analytics process model (Lemahieu et al, 2018) 22
Figure 8 Data Quality Assessment Framework. Sebastian-Coleman (2013, p. 66)........ 24
Figure 9 DQAF: Initial assessment Sebastian-Coleman (2013, p. 99) 25
Figure 10 DQAF: In-line measurement Sebastian-Coleman (2013, pp. 122, 125) 26
Figure 11 Theoretical summary... 27
Figure 12 Distribution of country, work and industries 33
Figure 13 Word cloud of responsibilities at work 34
Figure 14 World cloud: When analytics tools are used at work 37
Figure 15 World cloud: "What type of data quality issues have you faced?"................ 39
Figure 16 Type of tools or techniques can be used to solve data quality issues 41
Figure 17 Type of tools or techniques respondents have used at work....................... 43
Figure 18 Data quality effect on reliability of the reports.. 45
Figure 19 Data quality effects on decision-making 47
Figure 20 Best practices to counter bad data quality or to improve data quality 48
Figure 21 Ishilawa diagram of problem and solutions 50

1 Introduction

1.1 Background

Data quality challenges experienced by experts in reports and analytics tools issues continue to be a problem in various work roles, and quick overview shows that the data quality issues are faced across different industries and companies. Increase in the amount of data in relation to its utilisation is not on par in companies. At the same time, the data used in reports and analytics tools pose data quality challenges and issues in decision-making. For instance, studies conducted in different countries and companies show challenges with data quality in reports and business intelligence solutions (Seguer & Hasnas 2022; Lennerholt, Van Laere & Söderström, 2021; Dragos, 2021; Fakhouri, Al-Aamr & Jabbar; 2021) and different companies and industries need to have good quality data to make good decisions (Bechman 2022; Marshawn 2022; Austin, Carpenter, Christ & Nielson 2021; Botes, Hamer, Jaarsveld & Kleingeld 2019; Fagan 2012; Hughes & Forrest 2012). However, use of different business intelligence solutions, analytic tools and data analysis for decision-making has risen rapidly and we can spot not only data related jobs, but also analyst and specialist job postings that require skills in handling reports, data analysis and different analytics tools as well as their related processes. "Data is the main ingredient for any BI, data science, and business analytics initiative." Sharda, Delen & Turban (2018, p. 83). Figure 1 below illustrates the evolution of business intelligence by showing different aspects and themes that relate to business intelligence and use of analytics tools:

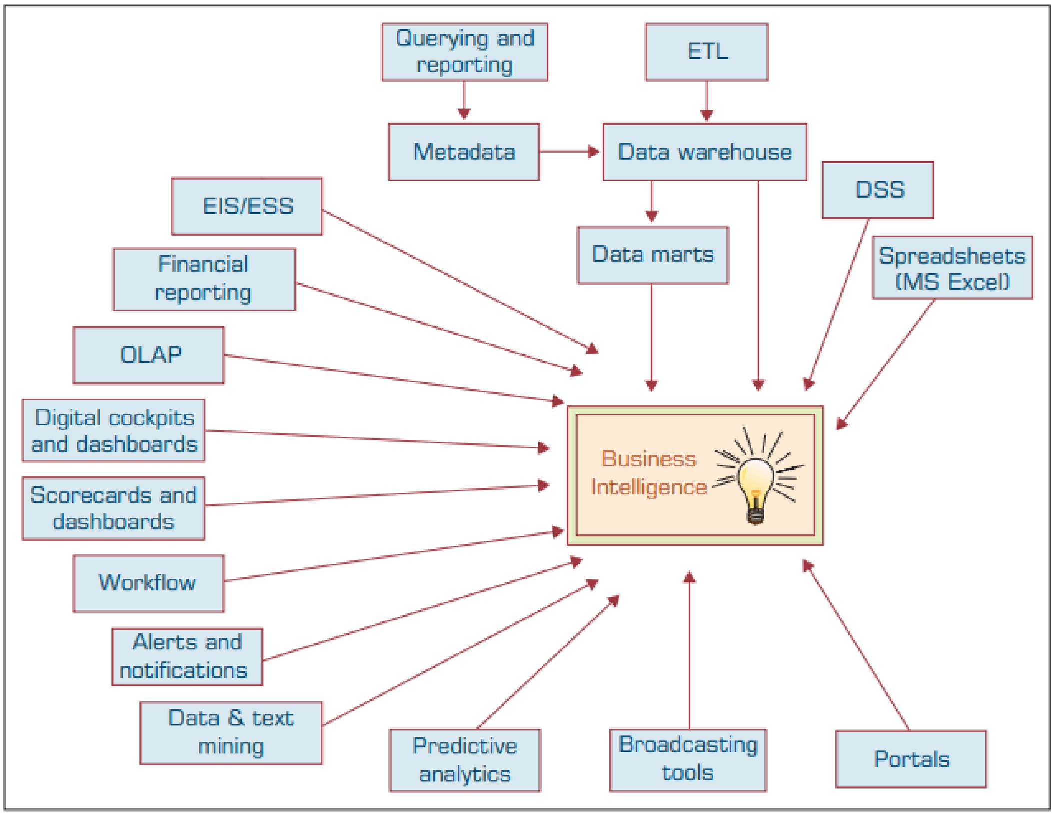

Figure 1 Evolution of Business Intelligence (Sharda et al. 2018, p. 43)

As the Figure 1 above illustrates, business intelligence arises from many different objects and concepts. The main concern the author has is what to do when the data within business intelligence solution is untrustworthy or providing wrong insights? This thesis investigates the impact of data quality in the business intelligence, reporting and analytics field, as well as in the operational decision-making. The research area of this thesis includes experts and specialists from different companies using business intelligence, analytics tools and reports in their work and for decision-making. The author wants to understand different data quality challenges specialists face when using analytics tools for decision-making across different companies and business sectors, as well as learn and form recommendations for best practices in countering these issues. Reports are often byproducts of business intelligence tools and are displayed in the form of a dashboard or a presentation, containing various insights, key performance indicators, visualizations and other metrics.

The author has worked as a system, reporting and analytics specialist and has grown interest and passion in the business intelligence field. Data quality issues are one of the challenges the author is facing even daily. Additionally, some preliminary discussions with colleagues and friends using analytics tools for work suggested that similar issues occur also in various companies and industries. For example, quite often, data quality proves to be an issue when generating reports, improving processes, or providing input for decisions at operational daily work and sometimes even for managers deciding on a strategic level. The field of study for this thesis is mainly focused on experts in business analytics or system-science. However, some earlier study results from health care data quality and responses from other disciplines, such as public administration was accepted for richer comparison. Business sector was not limited to any specific industry, but research problem or related issues were studied interdisciplinary.

1.2 Research purpose and research question

The introduction briefly describes business intelligence and its contemporary challenges, also the interest of the author to study this phenomenon. Research purpose is to analyse the current issues, discover and provide solutions for specialists using analytics tools in their work. Summarizing the research problem at hand and solve issue this a research question of this thesis was formed as “What are the data quality challenges experienced by experts in reporting and analytics?” and sub research questions “How are the data quality challenges tools reflecting in reliability and impact to decision-making?” and “What are the best techniques and analytics tools when encountering bad data quality in decision-making?”. Additionally, this thesis aims to produce and explore meaningful insights for specialists involved in reporting, analytics tools, and decision-making that they can start using in their daily work. What is more, the author aims to understand the problem in a detailed way to gain insight to form hypotheses for future quantitative studies. Therefore, study will be conducted with a qualitative and narrative research approach.

1.3 Limitations and delimitations

This master thesis studies the impact of data quality to operational or tactical decision-making, when specialists who are using business intelligence tools, reports, and data analysis for their daily work. Furthermore, when discussing reports, author focuses on internally produced reports for decision-making with business intelligence, for example business intelligence tools, analytics related working tasks and dashboards. Therefore, external reports, such as sustainability reports or annual reports produced for public or other stakeholders are excluded from this study. Theory part will focus on exploring the basic data definition, reporting, business intelligence tools, and how they associate with decision-making. Also, some case studies in business intelligence and data quality challenges will be explored to get broader view to the topic. Author wants to understand how different specialists experience the problem in their daily work, therefore the sample size was chosen to be specialists using analytics or business analysis in their work, and also data will be collected and analyzed in qualitative research method to get more detailed understanding from the answers. Author contacted specialists interested in the research problem and topic from various industries and work positions, therefore no limitation to specific industry or business sector was made. However, results are assumed to be more applicable for bigger companies with dedicated analysts positions or analytics teams. Business sector was not limited to any specific industry; however, the focus was mainly aimed towards employees working in big companies and the research problem, and its related issues in different business areas were studied interdisciplinary.

2 Literature review

2.1 Data

Data as it simplest can be considered as raw facts in the form of names and numbers. (Sharda, Delen & Turban 2018, p. 494) or abstraction and a piece of information of person, object, or event (Kelleher & Tienery, 2018, p.40). Same set of data can be used for many different purposes. For instance, product data for sales, inventory, forecasting, marketing, and other supply chain management. Another way to define a data set is for example by the number of rows multiplied with the number of columns, where the rows are entities and columns its attributes. For example, a book could be an entity having attributes, such as: author, title, topic, genre, publisher, date published, and more. Each row in a data set would describe one book through its attributes. Data illustrated in Table below:

Row	Title	Author	Genre	Publisher	Date Published
1	Book A	Author A	Genre A	Publisher A	25/08/2004
2	Book B	Author B	Genre A	Publisher A	24/04/1995
3	Book C	Author B	Genre B	Publisher A	10/05/2014

Data itself will not mean much unless the value can be captured efficiently, therefore management of data and analytics is as important as simply having or obtaining, especially large amounts of data (Kitchin, 2014). However, it can be argued that having too many attributes hinders the quality of data analytics or performance of different algorithms that aim to recognize patterns in the data. Also, choosing the correct attributes proves to be challenging for almost all data science projects (Kelleher & Tienery, 2018, p.41). Data gets interacted in many activities in a data life cycle, including data creation, collection, processing, storage, visualization and analysis. Data life cycle illustrated in the Figure 2 below:

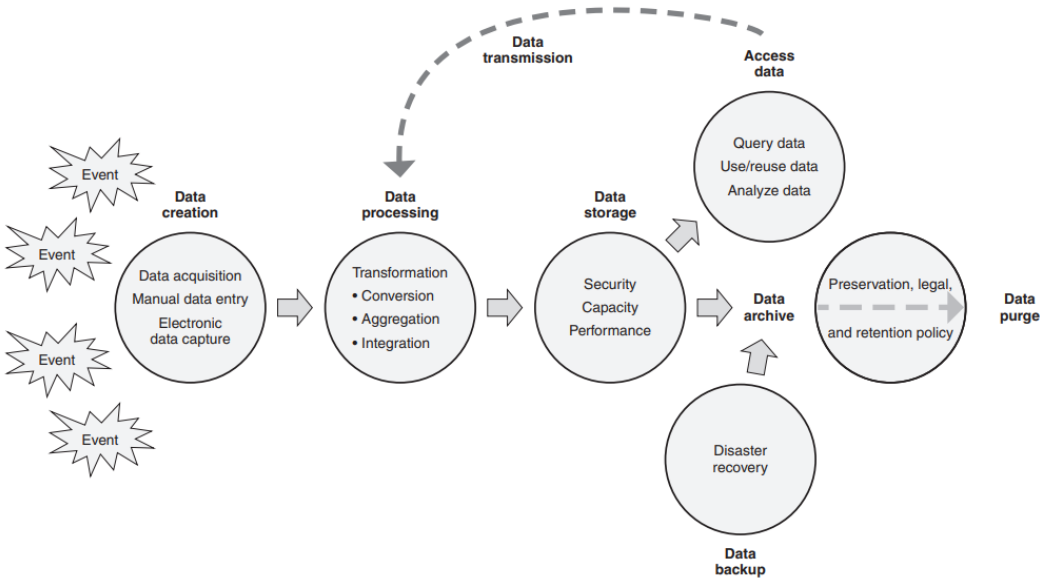

Figure 2 The data life cycle illustrated by Mahanti (2019, p.15)

Data should be available and in standardised format. Data warehouses make access to data handy. Sharda et al. (2018, p. 494) calls data warehouse as "A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format." Simply, data warehouse can be also described as an integrated intangible storage, where collecting data throughout the organisation for analytics is convenient (Kelleher & Tienery, 2018, p.11). According to Aktas et al. (2021, p.300) cloud technology has enabled digital platforms for data storage and availability for analysis across organization and its partners. Some prevailing ways to share, monitor and improve data and information sharing is for instance with strategic collaboration. This collaboration can potentially increase productivity and reduce costs in organizations (Desai, 2018, p.239).

Manual work is needed when consolidating data from different sources. In order to avoid disparity, processes like "extraction, transformation and load" are used to clean the data and to move it between various origins (Kelleher & Tiener 2018, p. 74). Sharda et al. (2014, p.144, 239) claim that data cleansing is an essential part of data warehousing, however it might prove to be challenging, when data has confusing definitions and formats that require changes organization-wide and at the executive level. Simply, data cleansing includes imputing missing values, reducing data errors and discrepancies. Furthermore, Botes et al (2019) found that data availability, analytics and visualization all have various levels that needs to be keep in mind when evaluating data quality in reporting.

Risks of sharing data include organisations exposing their internal challenges related to their operations and performance. On one hand, this might lead to shutting oneself in and not sharing data with external stakeholders to preserve organizational legitimacy. On the other hand, Desai (2018, pp.221-222) suggests that organizations might share the data and it as an opportunity to solve the problems related to their legitimacy. Studies also show that external stakeholders might use the opportunity to detect organizations' accountability to its stakeholders and the ability to meet their expectations. When sharing incomplete data, there is a risk for misinterpretation and effect on the complete picture. Additionally, Austin et al. (2021) studied auditor's perspective when evaluating companies' analytics and the possibility of influencing data analytics rules.

2.2 Data quality

Data quality can be described as "the holistic quality of data, including their accuracy, precision, completeness, and relevance" (Sharda et al. 2014, p. 666). Mahanti (2019, p. 9) presents data quality as "the capability of data to satisfy the stated business, system, and technical requirements of an enterprise". According to Cai & Zhu (2015) "data quality depends not only on its own features but also on the business environment using the data, including business processes and business users.". Sebastian-Coleman (2013, p. xxxi) defines data quality in two related factors: the conformity for the data consumers and the ability to represent the objects, events, and concepts it was created to represent. Similarly, it can be stated that the data quality is met when the business is able to use the data and it is complete, relevant, and timely. Also, "with the amount of disruptive technology-based changes now taking place, data quality within digital transformation has become a central component of major successful organizations' strategies in the twenty-first century." (Oakland & Oakland, 2018, p.17; Whiting 2006). Furthermore, the quality of data determines how much a person can trust conclusions from an analysis (Kitchin, 2014). Nonetheless, Mahanti (2019, p.10) and Kahn et al (2015) claim that it is important to note that data quality can never be perfect, because it depends on its applicability to meet requirements in different contexts.

Focusing on good data quality, Towse, Ellis & Towse (2021), this was also studied by Nash (2009) with insights of business intelligence system being crucial to help make decisions in reducing expenses. Additionally, in transportation industry stable good data has led to significantly better safety decisions (Hughes & Council 2012). Suppiah & Arumugam (2023) studied impact of data analytics and concluded analytics with good data is vital for supporting businesses to grow.

According to study of data quality issues in production planning and control by Lindström, Persson, Viswanathan & Rajendran (2023), data quality issues in production planning occur due to inaccurate data entries, data production error, multiple sources of data, task variation, resource deficiency and distributed system. DeSimone & Harms (2018) studied techniques of detecting low quality data in surveys and determined data screening as one effective method. Data screening when includes in the survey design indicators, instructions, measuring response time, detection of discrepancies based on answers and questions, and response variability. However, results were applied in quantitative research context and are not valid in all studies. Also, researchers recommend analysing results both before and after applying the screening options. Similar methods could be tried to detect low data quality in different contexts.

However, based on Sebastian-Coleman (2013, p. 57) data quality usually is understood as reference to the dimensions of data quality described by experts or specific data quality problems people have encountered in their daily work. According to Cai & Zhu

(2015) “quality dimension needs different measurement tools, techniques, and processes, which leads to differences in assessment times, costs, and human resources.”. Lennorholt, Van Laere & Söderström (2023) studied the impact of user-related challenges of self-service business intelligence and identified data quality as one of the key challenges in using analytics tools.

Data quality can be assessed for instance in a simple qualitative or in a more quantitative measured manner. Assessment can be based on the knowledge, principles or standards and it can be assessed from all the way from the generic macro level to specific values at the micro level. Assessment should be made to understand the capability and condition of the data to meet the expectations and its purpose. Data Quality Assessment Framework created by Sebastian-Coleman, describes data assessment as “a set of processes that are directed evaluating the condition and the value of data within an organisation”. (Sebastian-Coleman, 2013, pp. 46-47)

Data quality plays a key role in providing good customer service, efficient operations, accurate decision-making and strategic planning. Data quality can be examined from an assessment perspective, which divides it into two aspects: intrinsic data quality and contextual data quality. In this view, the intrinsic data is the unchangeable data elements such as accuracy and accessibility. In turn, contextual data quality is more of observations and reciprocity with the intrinsic data. Desired characteristics for data quality can be seen in the Table below (Mahanti, 2019, pp. 2, 9-10):

Free of defects	**Desired features**
Correct	Contextual
Complete	Pertinent
Valid	Comprehensive
Reliable	Easy to read
Consistent	Unambiguous

Unique	Easy to understand
Current	Right level of detail

Similarly, Sebastian-Coleman (2013, p.48) lists data quality dimensions in five different categories of completeness, which enables comparison of summarised data, timeliness, which allows to check actual time of data delivery to planned one, validity, which compares incoming data to defined valid values, consistency, which compares count of records to past records populated in the same field(s) and lastly, integrity, which confirms the level of the record and its referential integrity and relationship between tables to identify parent/child relationship. According to study conducted to employees working in production planning, similar desired characteristics were important (Lindström et al., 2023). Dimensions, measurement types, specific metrics illustrated in Figure 3 below.

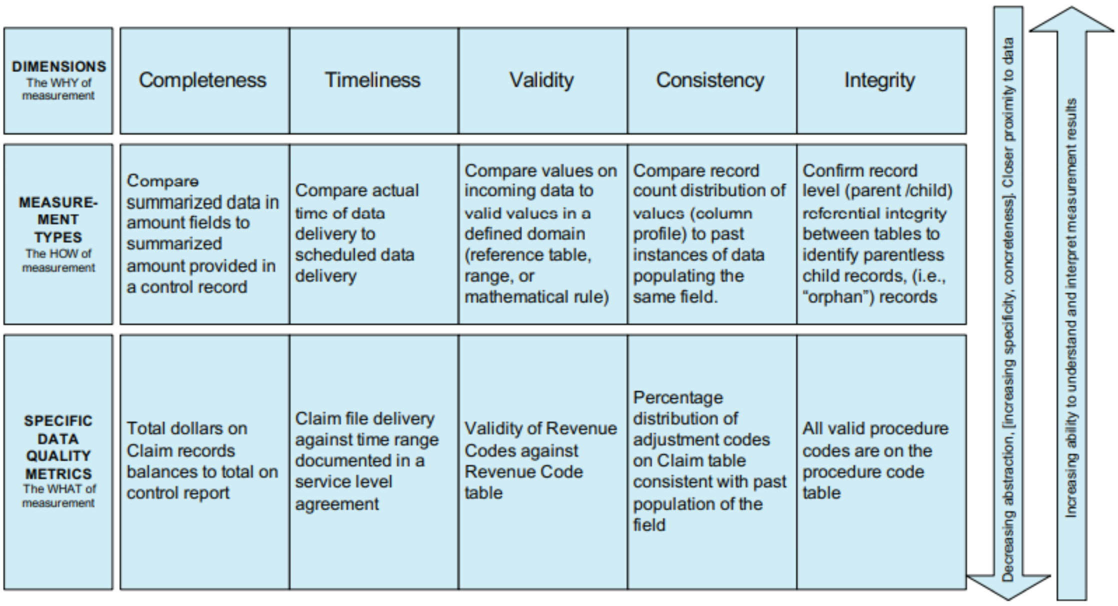

Figure 3 Dimensions, Measurement Types, and Specific Metrics (Sebastian-Coleman, 2013, p. 48)

Aktas et al. (2021, p.300) recognizes data quality issues as one of the several challenges in smart and sustainable supply chain management in Industry 4.0, data analysis and decision-making. Some of the methods suggested to tackle the problem are optimizing database management and collection techniques, especially statistical collection methods. According to Mahanti (2019, p.12) causes for bad data quality derive from all levels of information sharing and data life cycle. Some of the areas are illustrated in Figure 4 below:

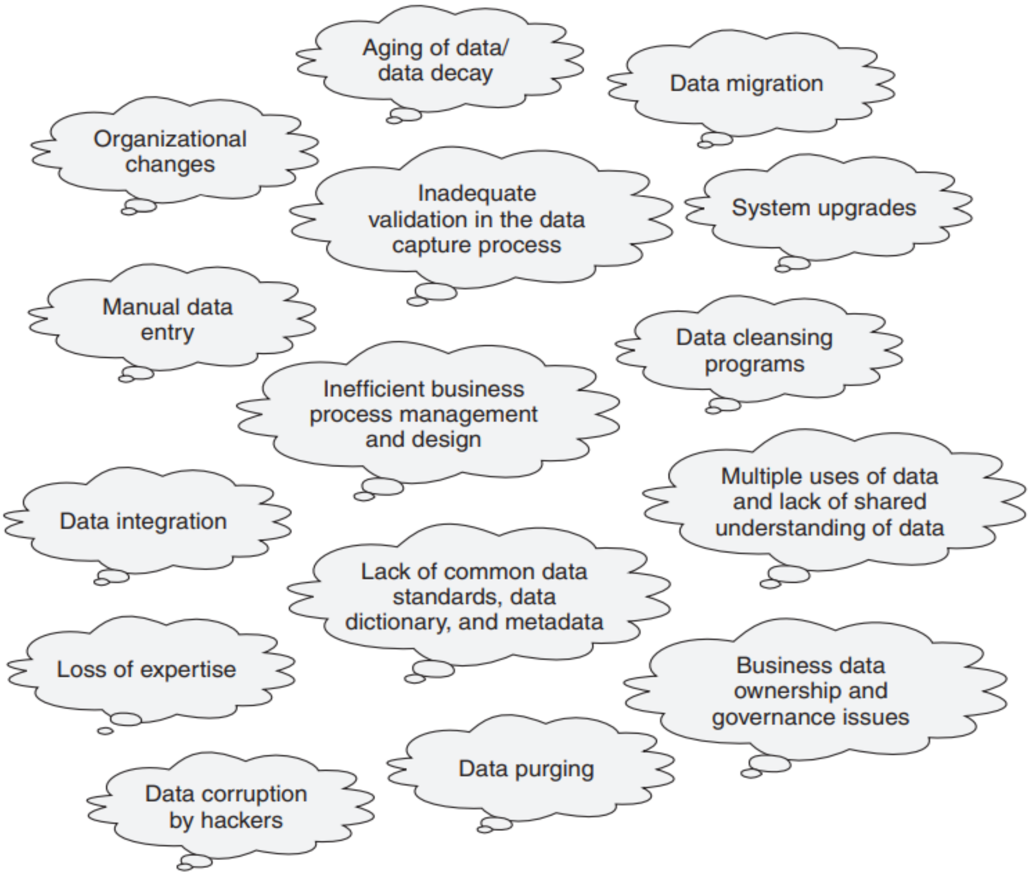

Figure 4 Causes of bad quality illustrated by Mahanti (2019, p.15).

Mahanti (2019, p. 41) lists the costs of poor quality in business operations into four categories:

1. Financial, for instance increased operating costs.
2. Confidence and satisfaction, for instance organizational trust and customer dissatisfaction
3. Productivity, for instance bigger workload
4. Risk and compliance, for instance investments and competition

Similar findings were made by Cai & Zhu (2015) & Choi & Luo (2019), for instance one major consequence of poor data quality is wrongfully conducted decision-making.

Haug, A., Zachariassen, F., & Liempd, D. V. (2011) studied the cost of poor data quality. Furthermore, they defined the optimal data maintenance effort and total costs incurred by data quality on the company. Their description of the Figure 5 below is that the "connection between costs inflicted by poor quality data and costs of ensuring high data quality can be logically categorized as a trade-off, which is a situation involving the loss of one quality in return for gaining another quality.". This means that organization needs to consider and find optimum they can bare to reach data quality.

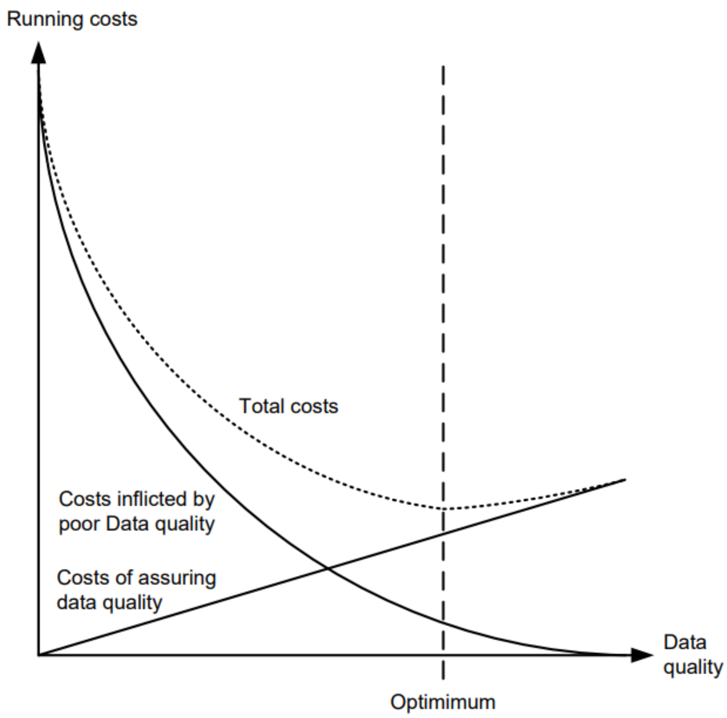

Figure 5 "Total costs incurred by data quality on the company" (Haug et al. 2011)

Also, Haug et al. (2011) illustrated four types of costs incurred by poor quality data as seen in Figure 6 below. Mainly, hidden costs occur for example due to longer lead times and duplicate entries of same data, which increases manual work and employee dissatisfaction. Further, on strategic level, hidden costs can lead to focusing on wrong segments, inefficient production planning and more. When looking at direct costs, they will occur due to manufacturing errors, deliveries that are not on time or has wrong items and payment errors. On strategic level direct costs can include for instance penalties of contracts due to low on-time delivery numbers, lower sales figures and low efficiency number. Conversely, good data quality helps departments to understand different dimensions for cutting expenses and creating savings (Nash 2009).

	Effects of poor quality data on operational tasks	**Effects of poor quality data on strategic decisions**
Hidden costs	E.g. long lead times, data being registered multiple times, employee dissatisfaction, etc.	E.g. focus on wrong customer segments, poor overall production planning, poor price policies, etc.
Direct costs	E.g. manufacturing errors, wrong deliveries, payment errors, etc.	E.g. few sales, low efficiency, problems in keeping delivery times, etc.

Figure 6 "Four types of costs incurred by poor quality data" (Haug et al., 2011)

Some contemporary studies of data quality show the effects of poor data quality in different industries. For example, Choi & Luo (2019) measured data quality of supply chain's and found out that poor data quality directly lowers social welfare and profit through supply chain. They claim that data quality, although critical, is still major problem in emerging markets. Ozmen-Ertekin & Ozbay (2012) claim that organizations seek for latest technologies and tools for managing data quality and they used process mapping to modernize the data management model for transportation council of New York. Dakka et al. (2021) presented untrainable data cleansing for helping reporting by allowing AI to detect poor quality data. In their study, the technique helped to improve data quality in healthcare industry. Callahan et al. (2017) study showed that lack of human resources can cause poor data quality and data quality might invalidate other work. Bhandari, Kumar & Sangal (2022) conducted extensive literature review for data quality issues with software fault prediction data sets and found that high data dimensionality as key issue.

2.3 Business Intelligence

Definition of Business Intelligence according to Kelleher & Tierney (2018, pp.72-73) is a decision-support system that is designed to produce data assembling, integration of several systems, as well as analytics and reporting possibilities. Kohtamäki (2017, pp.23-24) sees business intelligence as an information system that guides decision-makers at all levels of organization. Furthermore, it provides, stores and analyzes information, which enables applying the data in strategic planning as well. One of the future trends of enterprise resource planning systems is to offer solutions in analytics, for instance by offering users a platform to access their operation data, develop key performance indicators and create reports. Indeed, data visualization is one of the key components of business intelligence, visualization is defined by Sharda et al. (2018, p.494) as: “A graphical, animation, or video presentation of data and the results of data analysis.”. Additionally, according to Kumar (2017, pp. 32, 53) data visualization and regularly published reports enables top management and the stakeholders to easily gain awareness and be able to communicate different aspects of business. What is more, nowadays many reports, especially ones that use direct queries and real-time data are quickly accessible for the decision makers even on smartphones, which allows quick and proactive reaction (Kumar 2017, p. 32). Lemahieu (2018, p. 1424) argues that the key issue in data analytics is to “to find the unknown but interesting and actionable patterns that can provide new insights into your data”.

According to Aktas et al. (2021, p.183) data analytics and reporting have become more data-driven and therefore top priority for companies. However, according to Kohtamäki (2017, p. 46) many companies continue to waste a big chunk of the data they collect and fail to utilize it in decision-making. Similarly, Fagan (2012) claimed that analytics offers too much data, which either requires or limits taking action. Many scholars have reported significant correlation between business performance and data analytics.

Some of the major companies, like Amazon, Apple, Google, and Facebook use considerably data analytics in their product development and operations. Companies used to make decisions based on opinions instead of analysing their data, which often led to faulty conclusions and were harmful for businesses. However, in recent years data-driven decision making has become more prominent and often a primary reason for high-class performance. This performance has been measured for instance in positive company profit and share development. One can say that the two big picture objectives of data analytics are problem solving and decision making (Kumar, 2017, pp. 2-4). The key driver for efficient business intelligence is to develop data-driven strategy and decision-making (Kohtamäki, 2017, p.24). Lemahieu, Broucke & Baesens (2018) presented the analytics process model, where the pre-processing phase includes identifying business problems, data sources, selection, cleaning and transformation of the data for business use. Pre-processing phase is usually done by a data scientist, which is considered a top position for its unique requirement for quantitative, programming, communication, as well as visualization skills. Combination of these skills is crucial when presenting data models and preparing user-friendly platforms to analyze and monitor data. The analytics process model is illustrated in the Figure 7 below.

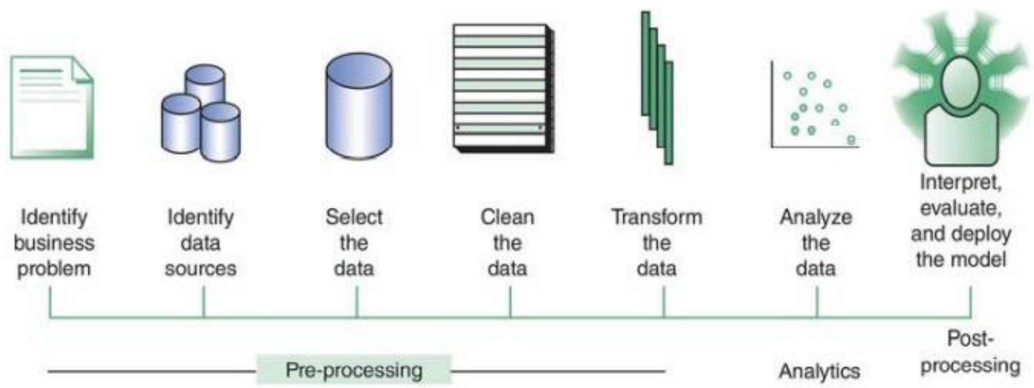

Figure 7 The analytics process model (Lemahieu et al, 2018)

According to Lemahieu et al. (2018, pp.552-553) companies have different levels of decision-making. Decision-making can be categorised as operational, tactical and strategic decision-making. Operational is more related to daily business decisions, typically in real-time and/or in a short time span. For example, real time transactions made during a day in the company. Tactical decision making is made by middle management in a more regular cycle, for instance monthly, quarterly or annually. For example, monthly sales and production forecasts to make purchasing decisions for raw material. Finally, the strategic decision-making is made by the senior management with longer goals and objectives. For example, the decision to investigate a new area, its competition and whether to enter the new market or not. For tactical and strategic level, the information systems are usually called decision support systems due to their nature of providing information for decision-making in the medium or long term.

2.4 Data Quality Assessment Framework

Sebastian-Coleman introduces a framework called data quality assessment framework (DQAF), which enables a structured measurement method for data quality. The framework was originally created to help improve data quality with a systematic approach that works across data storage systems and was understandable by the users from various systems. The DQAF requires defining metadata to understand the data chain. Specifically, business concepts represented by the data, processes that create the data, technical or business processes that maintain, update, or delete the data, the data model where it is measured, and the data processing rules for the target system. Nonetheless, many companies might not necessarily have documented their data model, despite having data stored in applications and different data storages. Sebastian-Coleman (2013, pp. 57, 64-65)

The DQAF approach and context has been summarized in the Figure 8 below. The top row goal defines data management goals of the assessment categories. The assessment categories describe the frequency and type of the assessment. For example, control, measurement, and assessment either one-time, ongoing or in periodic frequency. Third row represents assessment activities related to their measurement types, and whether they focus on process or the content. The foundation for the framework is found at the bottom in support processes and skills, because in order to be successful, competent personnel need to be able to assess and evaluate the findings. Sebastian-Coleman (2013, p. 66)

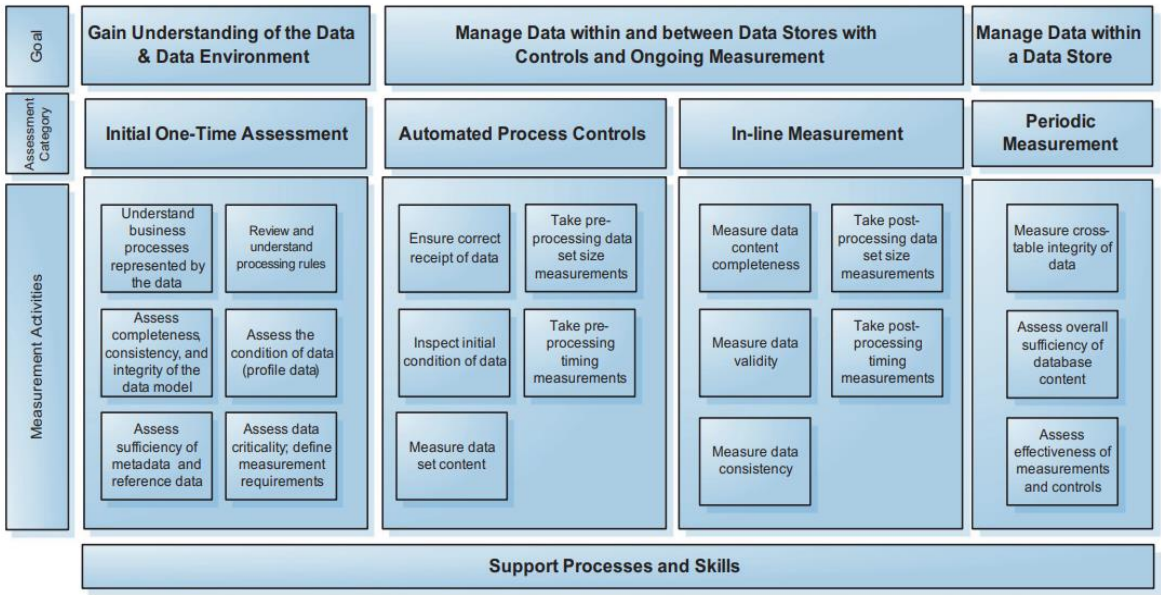

Figure 8 Data Quality Assessment Framework. Sebastian-Coleman (2013, p. 66)

Quite often one will not know much about the data first at hand, which is why during an initial one-time assessment one can try to get a first grasp at the content and structure of existing data stores or even potential data to be used in a project. One should gain understanding and explore variables that the data should be about, such as documented processes represented by that data, data model itself, processing in the target systems, producer as well as consumer expectations and assumptions. One of the critical components is business criticality perceived by the actual business users, which is needed for assessing potential projects and in-line measurements. Initial assessment ultimately forms metadata and therefore foundation for upcoming assessments (Sebastian-Coleman, 2013, pp. 97- 98, 111). Figure 9 below presents aforementioned as well as measurement activities, assessment categories and their goals in summary boxes.

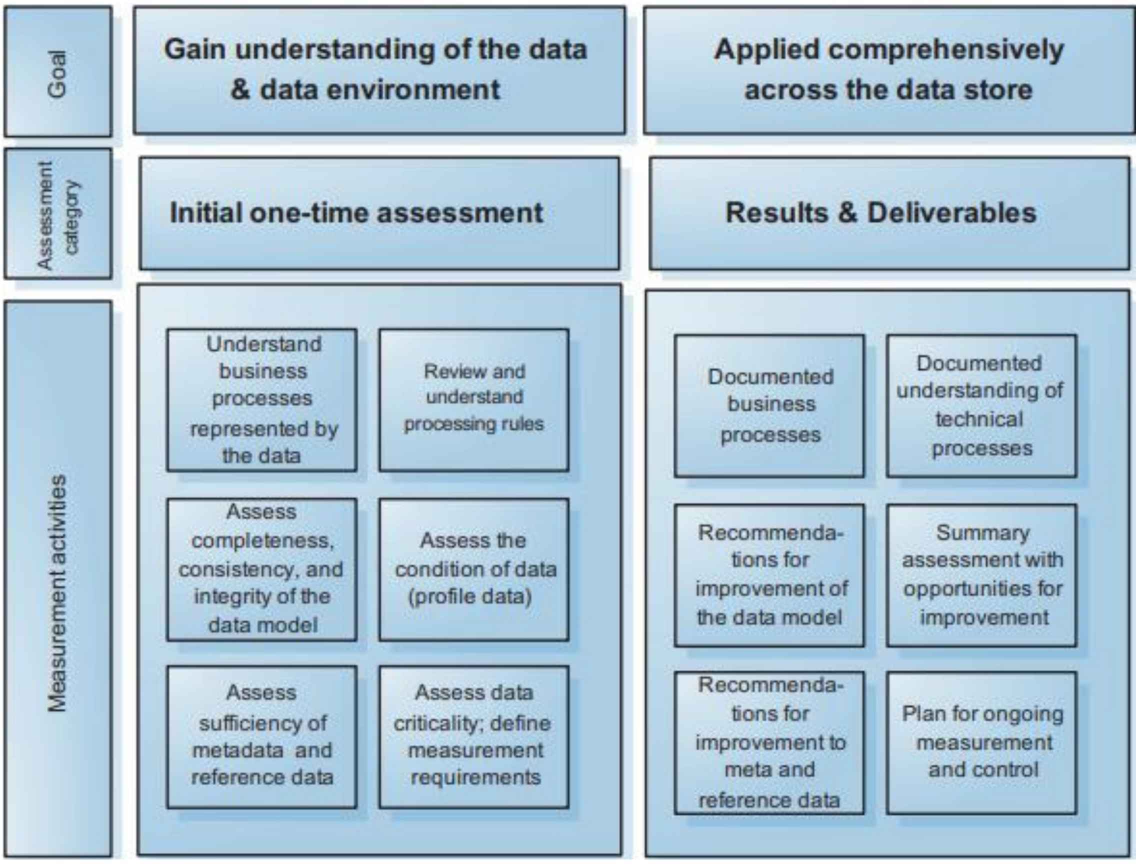

Figure 9 DQAF: Initial assessment Sebastian-Coleman (2013, p. 99)

After initial one-time assessment, the automated process controls are defined. This includes correct receipt of data, initial inspection of quality, and other pre-processing measures. These inputs include "documented understanding of business process that produce data, documented understanding of technical processes in source and target systems, data definitions, data lineage, data elements, and documented limitations in metadata and reference data." (Sebastian-Coleman, 2013, pp. 121). Furthermore, the In-line measurement as shown in Figure 9 below, characterize what the decision-maker or analyst should evaluate within the in-line measurement, it should indicate which data is supposed to be measured, which controls to use and if data should be measured periodically. Results & deliverables for this part is to collaborate with data consumers and data producers, as the collaboration will help to understand data management choices and their impact. (Sebastian-Coleman, 2013, pp. 121-122)

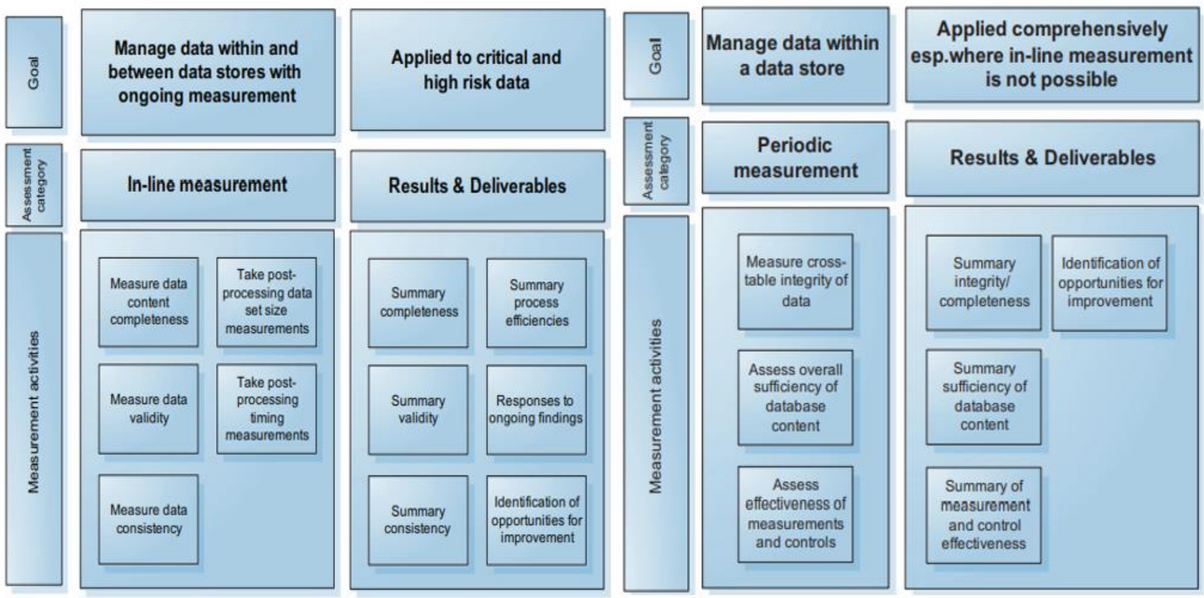

Figure 10 DQAF: In-line measurement Sebastian-Coleman (2013, pp. 122, 125)

2.5 Theoretical summary

Reliable and carefully studied theory is backbone for any research. For this study, the author wanted to find and summarize information supporting the research question, therefore understanding the nature of data and its related concepts, such as data quality, business intelligence and decision-making was important. Key findings included for example data quality assessment framework by Sebastian-Coleman (2013), business intelligence process models by Lemahieu et al. (2018), data life cycle by Mahanti (2018) and basic data concepts by Haug et al. (2011) that help to understand the overall problem. Furthermore, data and its instant connection to decision-making was handled and different costs occurring due to poor data quality was discovered.

Author wanted to limit the study to be conducted for specialists handling reports and data of different qualities. Therefore, differentiating and understanding levels for the decision-making, such as of strategic, tactical and operational decision-making was necessary and explained with help of database management book by Lemahieu et al. (2018). Data comes in different shapes and forms and is collected and maintained in different sources which already increases risks with decision-making. Countermeasures discovered in the theory clear identification of business need, data sources and periodic measurements of specified data. These discoveries already help to partially reveal solution to the research question and form useful insights for specialists struggling with data quality and decision-making, but also understand the powerful impact it has on company's operations and financials. Finally, author created summarized findings of theory, illustration can be found in Figure 10 below.

Some previous research on similar topic were found and summarized. For example, Bhandari et al. (2022) study in data quality issues in software fault predictions, Cai & Zhu (2015) with a study of the challenges of data quality and data quality assessment in the big data era, Choi & Luo (2019) with a study in data quality challenges in supply chain operations, Lindström et al. (2023) with study on data quality issues in production planning , Seguer & Hasna (2022) business intelligence as a challenge for the managerial function, Dragos (2021) with a study in project management and implementing business intelligcnce approach, Lennerholt et. al (2021) with a study in user-related challenges of self-service business intelligence, and Dakka et al. (2021) with a study in automated detection of poor-quality data in health care. Figure 11 illustrates theoretical findings in form of fishbone diagram, with contributing factors and possible solution proposals:

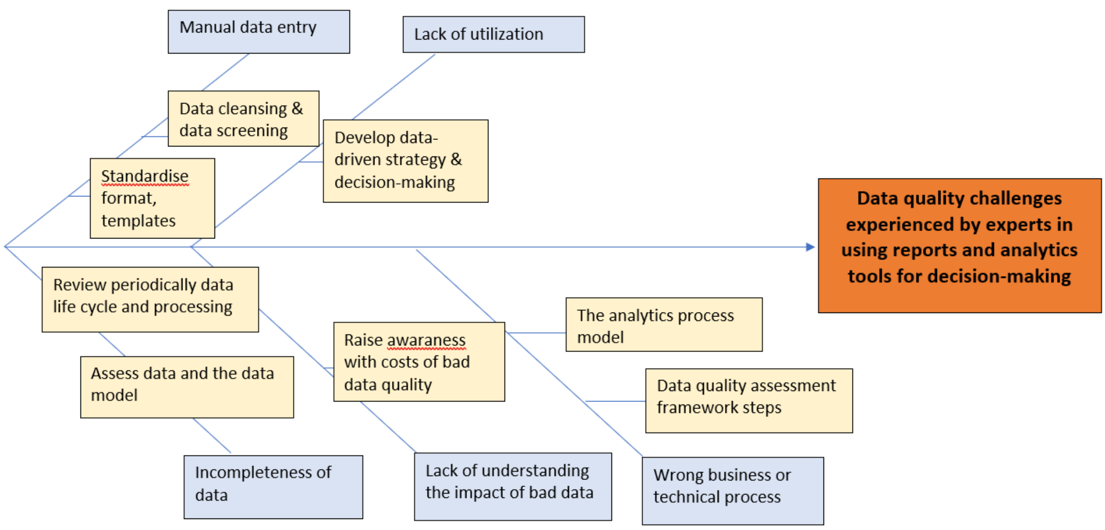

Figure 11 Theoretical summary

3 Methodology

Author wanted to understand the topic from a qualitative point of view to gain more insight, by gaining information on how specialists in different industries or similar positions have faced and experienced data quality issues in their daily work. Also, the author wanted to understand and study the phenomenon in detailed manner, gaining insights and experiences of data quality in daily operations or decision-making for specialists working with analytics tools and reporting. To gain definite answers of the topic qualitative research was chosen method for this study. Qualitative survey was planned to collect responses from professionals and employees handling reports of different quality and directly with experience and knowledge in working with data driven reports and operational decision-making or supporting decision-making. To meet the delimitations set for the thesis sample size was targeted at specialists working in bigger organizations and using analytics in their daily work.

Mix of purposive and voluntary response sampling method was chosen to find the respondents with LinkedIn search. The author purposefully contacted specialists that author estimated to be interested in the study and possibly facing the similar issue in their work role. Sample size was chosen using LinkedIn and studying job descriptions of specialists potentially using reports, analytics tools and business intelligence in their work. Total of 20 candidates were contacted with a cover letter explaining the research problem, premade qualitative survey with open-ended questions and inquired whether the candidate is interested in participating in the study. Response rate for the sample size was 70%, which was a result of the voluntary sampling method. Questionnaire structure, cover letter and content can be seen in Appendix 1. Further, the qualitative results will be compared to the theoretical findings to form meaningful recommendations to answer the research question.

Author chose to utilise Applied Thematic Analysis short (ATA) and SAGE Handbook of Qualitative Data Analysis to help process and analyse the data. “ATA is a type of inductive analysis of qualitative data that can involve multiple analytic techniques”

(Guest, MacQueen & Namey 2012, p.2). ATA is a methodological framework that consists of parts from different qualitative methods, like grounded theory, positivism, interpretivism and phenomenology. ATA was built to support applied research context by combining the most useful techniques by being credible and transparent, therefore it is valid help for the researcher to study and break down collected textual data. ATA creators do not claim it to be a novel approach, but state that similar approaches have been conducted over the course of year by many researchers, nonetheless ATA is structed guide that can be used by researcher to analyze qualitative data (Guest et al. 2014, pp. 13-14). Author chose ATA approach after researching best practices to analyze qualitative data, for instance Sage Publishing describes ATA book as follows: "After collecting qualitative data from in-depth interviews, focus groups, or field observations, students and researchers often struggle to make sense of it. This step-by-step guide draws on the authors' many years of experience carrying out qualitative research and conducting trainings on the subject." (Applied Thematic Analysis 2023).

Thematic analysis requires careful look in identifying themes in the collected qualitative data, especially which help to answer the research question (Flick et al. 2013, p. 148). Additionally, different phases of coding might be conducted: initial coding, focused coding and theoretical coding. Initial coding helps the researcher to answer questions, such as "What is this data a study of?" or "What category does this incident indicate?". Thinking different type of analytical questions with help of coding helps to summarize and create segments of data that can be evaluated further. "The researcher reads and analyses the data word by word, line by line, paragraph by paragraph, or incident by incident, and might use more than one of these strategies." (Flick et al. 2013, pp. 156-157). Author will establish and utilize coding to help breakdown and critically review the gathered data. Furthermore, the codes could be either keywords or concepts recognised in the answers, which the author planned to summarize in world cloud visualisation per survey question and answers. Discovered keywords will be represented in word cloud figures in the following chapter.

Guest et al. (2014. pp. 79-82) has found studies that contradict whether reliability & validity should have any value when analysis qualitative data. Nevertheless, Guest et al. claims that validity is the most important aspect in any research, and researchers are aiming to collect "the most meaningful and truthful data possible". Guest et. lists techniques that should be taken into consideration in enhancing validity and reliability of qualitative research. This includes using multiple methods and data sources when possible, involving research teams and pretesting sample group, training field team in collection techniques, adjusting data collection structures, monitoring data in real-time, transcribing, creating translation expectations before collecting data, using pre-developed codebook, using multiple coders, using peer review of coding, creating audit trails, triangulating data sources, using negative case analysis and supporting themes and interpretations with quotes from the respondents. For this master thesis, the author has used aforementioned techniques, when possible, to validate the data collection and analysis. For instance, pretesting of the group was not conducted due to time strain, but the respondent was given chance to ask elaborative questions through LinkedIn chat or e-mail, for instance if some questions or context was unclear to them. Translation expectations were not documented prior to the research, but full comparison to the theoretical framework was conducted to draw meaningful insights that link empirical research with theory.

4 Empirical research

4.1 Qualitative survey

Qualitative survey that was chosen as the primary data collection method was created with carefully prepared questions that measure the respondents' interest, knowledge and experiences regarding data quality and decision-making. Particularly the focus was to collect unique data regarding the challenges experienced by specialists using the analytics tools as mentioned with the purposive and voluntary response sampling method. Candidates were search using LinkedIn search with a cover letter and the qualitative survey was planned and distributed to candidates that expressed interest in participating in the study. Set of qualitative survey questions were prepared in neutral and open-ended. Reason for qualitative survey was to ensure availability and flexibility for the respondent, as well as author's time. Qualitative survey was built using Google Forms and the respondent had option to answer either in Finnish or in English. According to Guest et al. (2012, pp. 97-100) "translating a data collection event from one language to another adds an additional layer of complexity that can affect both the validity and reliability of the data with which a researcher works.". Therefore, translation process can affect the interpretation of the data, keeping this in mind the author will make effort mitigating researcher bias when doing the translation from Finnish to English. Cover letter and structure of the survey can be seen in Appendix 1.

Author aimed to reach people who are involved in operational and tactical decision-making discussed in the theory part of the thesis. Therefore, the focus group was created by finding specialists handling reports and analytics tools in their daily work. Focus group was not limited for instance based on gender, working country or age group. Additionally, to narrow down candidates from the professional network the author created a list of candidates based on work title and/or work description. The ideal respondent profiles for the research were searched in LinkedIn and contacted with cover letter introducing research topic and objectives. Overall, 20 people expressed interest in participating in the study, however within given time frame the qualitative survey received overall 14 responses. This made the response rate fairly high at 70%.

“Quotes containing concrete examples of a theme are especially informative.” Sage et al. (2014, p. 267). Therefore, author decided to use direct quote from unique respondent that author felt described the problem per each qualitative question. However, quotes will not be taken only from respondent that could strengthen research bias, but with careful consideration of different respondent to each question. Also, in case a negative answer or a reply that did not meet expectations or align with the rest of the answers, the comment will be also brought up when analysing the results to diversify the analysis. Additionally, the demographic of that respondent will be represented within formed groups based. Given that age and gender were not part of the study, demographics will be grouped based on the country of work, work experience, industry and/or working title for a better context.

4.2 Demographics results

According to Guest et al. (2012, pp. 97-98) key demographics can give advantage in giving context to the reader and for instance different factors in the background of the respondent helps to understand the answer differently. Results from the qualitative survey showed primarily respondents from Finland with 72% out of all respondents. Other two representative countries were Poland and Germany with 14% each. Work titles of the respondents offered diversity and specialists of different fields, such as Product Owner, Development Engineer, ERP Consultant, Product Owners, Strategic Purchaser, Supply Chain Analyst, Data Engineer, Expert in Public Sector, HR Specialists, Project Consultant, Archiving Expert, Provincial Expert, Quality Analyst. Work titles were coded into following categories: Engineer, Consultant, Analyst & Operational Specialist. Additionally work industries were labelled and harmonized into three different categories: Services, Technology and Public. Country of work can be labelled based on the three unique countries or Finland with ten representatives and Central Europe with four representatives. Keeping these coded categories in my mind, they will be used to analyse if there is significance based on the backgrounds. Figure 12 shows a quick look at the demographics per grouping.

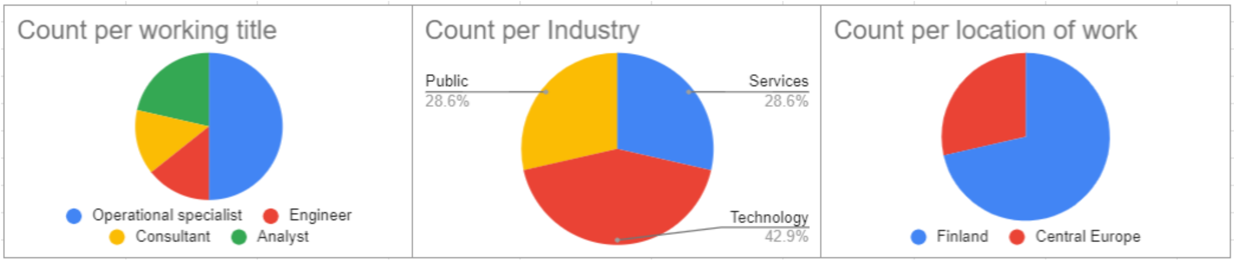

Figure 12 Distribution of country, work and industries

Operational specialist and Analyst respondents could be found in both locations of work, Finland and Central Europe. Engineers were only represented in Finland, whereas Consultant in Central Europe. Although work titles varied between respondents, when asked to freely describe work responsibilities the target group already showed that they have similar duties in their line work. For instance, for the key words in the answers to the question 'describe briefly your job positions and responsibilities at work', word cloud in Figure 13 was generated.

Figure 13 Word cloud of responsibilities at work

Figure 12 shows repetition of certain responsible words, such as documentation, analytics, supporting, reports, management and development. All in all, all the respondents have work experience for several years, with two respondents over ten years. Furthermore, when describing work responsibilities, we can already interpret that specialists are using data and analytics tools for their work and decision-making in their responsibility area. Analytics and reports were commonly mentioned, and often used and interpreted in the work itself to conduct daily work as well as operational decision-making. Also, few respondents mentioned that they do not use reports too often, but more periodically, for instance on quarter basis.

When analysing from the harmonized work title point of view, the most common keywords mentioned by operational specialists were reports, analytics and some type of operational work, such as purchasing, recruiting and documentation. For Engineers, the most common key words were development, training and analytics. Analysts mentioned analytics, reports and documentation. Finally, consultants mentioned consulting, supporting, reports and development. While this gives some context on the responsibilities of the working title, similar keywords can be found in all categories, and this validates the respondent understanding of the issue when answering the survey.

Looking at the harmonized industry and keywords, we can see significance in words “development” and “training” for industries grouped in Technology. Those grouped in Public, most commonly used words were “analytics”, “documentation” and “management”. Finally, those grouped in services, the key words were “reports”, “analytics”, “consulting” and “supporting”. While there were differences, this observation is not necessarily externally valid to these industries but gives the reader more different context of the ideology when checking certain group’s answer. For instance, technology grouped people tend to develop reports and train people for using reports or dashboards as analytics tools. Quote from the most suitable reply from this group by a data engineer: “Developing an enterprise analytics platform in an agile team and helping end-users to adopt it.”. Public group might use the tools for analytical managerial decisions and documentation, for instance as one respondent would describe it as: “we analyze different supplier options based on price, responsibility, capacity, delivery times and more…”. The services group diversely for analytics, consulting and collaboration between teams. One key intake from services group was described as “…communicate insights and implications to key stakeholders.”, which is at very root of the research question as the persons using analytics tools might have impactful influence not only their own work, but also stakeholders around them.

All in all, the respondents pool in the survey is diverse and valuable to the research as there are many representatives from various industries and from three different countries. Key words in the demographics study already imply high interaction with reports and analytics tools, which validates the study as we are received rich data from relevant audience. Following the demographics breakdown, the author decides to open the answers question by question with some additional patterns and codes found in the answers.

4.3 Survey questions and results

4.3.1 When do you use business intelligence tools for data analysis or decision-making?

“I use reporting when my team conceptualizes the thing to be developed. We also create new reports so that we can monitor how the developed thing behaves and refine the process based on that.” Question 4.3.1 of the actual qualitative questions regarding the respondents’ experience towards research problem was formed to understand whether respondents’ use analytics tool or other method for daily work and decision-making. As the opening quote suggests, one prime situation a specialist uses analytics is a planning phase, such as innovating a product, modelling new service design or a process improvement that requires. Monitoring development of that phase with available data or carefully planned and standardized metrics can provide a huge benefit. Similar

answers and concepts were recognized from other respondents as well, and word reports and reporting development plays pivotal role as word report can be used interchangeably with business intelligence with the chosen focus group, since business intelligence outputs are from simplest excel reports to interactive and smart dashboards, which in spoken office language can be simply called “reports”.

Author wanted to collect more contextual understanding of data analytics in their work area. Additionally, question can be understood at least in two ways: in which situations the respondent uses the tool in their work and/or how often they use it, and both answers were accepted and taken into consideration. Additionally, from the results can be interpreted that majority of the respondents use analytics tools daily or weekly, whereas minority more rarely – monthly or quarterly basis. For instance, every respondent grouped to public industry uses analytic tools more rarely than respondents belonging to Services or Technology group. This might indicate that business intelligence tools are in less frequent use in public use but generalizing needs further study when comparing public sector and other businesses.

Author created another word cloud and measurement of how often the tool is used based on the answers. World cloud describing situations when tools are used for data analysis or decision-making can be seen the Figure 14 below. The identified and formed concepts were operational decisions, conceptualizing, configuration of the tool, communication, business intelligence and reporting development. Out of these concepts, communication stood out the most from the survey entries and was mentioned in almost every answer. This validates the fact that data analytics and reports are used and are connected to large audience in organizations, since for example even for the respondents using tools more rarely, they have mentioned that when it is used, they have to communicate or handle data for large number of people. Second most significant theme was that analytics tool is utilized in making operational decisions, no matter the industry or work title. Furthermore, some answers confirm job descriptions in action, such as reporting development, business intelligence, configuration of tool and conceptualizing prior projects.

Figure 14 World cloud: When analytics tools are used at work

Results from this question prove that respondents' interaction with business intelligence tools or different "reports" is evident. The level and frequency of the interaction changes, but not very significantly between the respondents. Answers from Finland and Central Europe had similar characteristics, and only with grouped industries Services, Technology and Public, the Public was noticeably different from rest of the answers. Key themes based on working title had no big variation and tools are used in similar contexts when grouping is made to logical concepts. Nevertheless, variation could grow if larger and more specified groups were made, but such assumption would need larger respondent group and possibly a quantitative study. Keeping in mind the extent of analytics tool or report usage, we can state that answers for the follow up question in the survey are adept.

4.3.2 What kind of data quality issues do, or have you face(d)?

"Lack of data caused by humans not using the system. If the system needs frequent manual inputs from operators, it cannot be trusted. If there are data, it is not telling time wise correct assembly times, since operators usually use the system only at the start and end of their shift.". Question 4.3.2 was created to understand whether respondents' face data quality issues in the context and how they describe them. As seen in the opening quote's experience, the data is not always fully automated and handled solely relying on generation by intelligent systems in the background, but also require human

intervention. Additionally, we can notice that the data might be produced by different stakeholders, such as operators or external parties, who might have different understanding of importance and accuracy of the entered data, as well as its impact to the rest of business operations. What is more, some factors related to person's motivation and energy level can be taken into consideration, for example in some manual data inputs the timing might not be correct and data entry might be based on short-term memory of a worker, who is at the end of their shift, presumably tired and thinking of ending their working day. With this example "UserAwaraness" and "HumanErrors" groups were formed.

When creating a world cloud summary of data quality issues, the respondents have faced, the words and concept that were described most were: "Wrong format, using external data source, human errors, wrong process, user awareness, lack of data, no documentation". Breaking down the "wrong format" theme, the respondents shared similarities in either in data provided from various data sources, which also relates to concept "wrong process" and "external data source". This group was formed, due to fact that researcher recognized in the theoretical study, that wrong format is one key issue with data and its quality. Respondents described wrong format with key words, such as inadequate data collection template, which indicates wrong process in place in related to data collection or data processing Additionally, lack of data was provided as one of the key problems. Few answers claimed the issue is due to missing compulsory fields in the data collection template or process has been set up without considering all important aspects related to the particular data sets or operation. This might mean incorrect baseline or inconsistent measurements. What is more, some specialists are depending on data that are handled or input by another team and sometimes the responsibilities regarding timing has not been clear to all parties. Also, in some cases documentation for data collection does not exists, is expired or inaccurate. In these cases, also the responsible party might be unknown within the team or organization.

Another key data quality issue was source of information, which included situations when multiple one sources needed to be used, either internally or externally. Data might be found in systems, where you do not have access to and are expecting people to extract and provide the report. Another issue from same context was described not having data in the same format when extracted from different sources and the unifying of the data is manual work, that could cause human errors. The data format might be differently defined when using multiple reporting formats that are combined into one and this might cause mismatch in the data quality. The same issue was further linked to the company's culture in using analytic tools, in example:" In my opinion, the main issue is data appearing in various forms that decision makers cannot utilize the data if they are not strong in analytics culture. Since the natural of the work is to collaborate different departments, the data can be drawn from many sources and often not collected in one central place.". Summarizing the data quality issues, the generated word cloud can be found in the Figure 15 below.

Figure 15 World cloud: "What type of data quality issues have you faced?"

When looking at the demographics defined in 4.2. chapter and the keywords formed in this section, the most common themes described by respondents in technology as well as public the words were human errors, user awareness, wrong process, wrong format and external data source. Services section had only three key themes in common - lack of data, wrong format and wrong process. Based on working title or working country no significant observations were made.

4.3.3 What type of tools or techniques can be used to solve data quality issues?

"Proper up-to-date documentation, Data keys must be defined and added to the data, Missing data reduction by making data fields mandatory to fill up, Data fields include selection possibilities from a list to avoid invalid or wrong data." & *"Direct integration to information systems or internal reporting tools. Personnel guidance on consistent records.".* As with the previous question we got understanding of different data quality issues specialists face, the Question 4.3.3 was created to measure whether respondents understand how data quality issues are countered and/or how they feel data quality issues should be solved. The opening quote opens good techniques in countering bad data quality. Documentation can help the specialist to define and store measurements or indicators in use, but also how they are built, what they are measuring, what is the source and logic behind them. Documentation as concept integrates to further themes, like user awareness, communication and process optimization. The communication helps to raise user awareness, whereas process optimization ensures that correct baseline and measures are in use in the analytic tool. Overall, various themes were identified, such as "User awareness, Training, Communication, Automatisation, Human Intervention, Process Optimization, Business Intelligence Tools".

Further process optimization techniques described by the respondents were data cleansing operations, which relates to pre-processing of the data as discovered in the theory part. In the data cleansing the data is transformed into a useful format that can be used as final source for the analytic tool and its user interface. However, as mentioned by one respondent, sometimes it takes experience and tacit knowledge to recognize data quality issue in the ready report. Additionally, sometimes when the data quality is known issue, it is the user's responsibility to determine the reliability of the report. User awareness and communication are the other techniques helping the business user to determine whether they can trust the data presented in the analytic tool or not. For instance, there might be user guide or tooltips that advise the user on the dynamics behind the measures in the report.

Training was mentioned multiple times amongst the respondents, and it can contribute significantly to all other themes. For example, when data collection is relying on different stakeholders, all must be aligned with the data collection template and data format of different values entered into the template. This can be achieved with not only communication, but also efficient training, where the experts teach and explain about the data quality issues and its impact on other parts of the business processes and operations. However, to mitigate continuous human errors, final theme to be addressed was automatization. With automatisation certain standardized processes, where the data is clean, usable and trustworthy from the start is ideally possible. Also, direct integration between systems enables automated flow of data from one point to another. For instance, data collection point to data storage to database to front end report and analytics tool. Qualitative results emphasize also automated parameters and restrictions that can be set to ensure right data is in correct infrastructure and collected in the right format. Summarizing the tools and techniques, the generated word cloud can be found in the Figure 16 below.

Figure 16 Type of tools or techniques can be used to solve data quality issues

When looking at the demographics defined in 4.2. chapter and the keywords formed in this section, the public group described words automatization, process optimization, training and user awareness as most important techniques. Services group top themes were user awareness, training, communication, process optimization and human intervention. Respondents in technology group mentioned more business intelligence tool capabilities, automatization and human intervention to counter the issue, as well as user training and awareness. In the group division the techniques were also overlapping a lot, meaning similar practices are being used in different working fields. However, the technology group's mentioning of automatization and business intelligence tools could be expanded or tried with the other industries to draw more detailed conclusions.

4.3.4 What type of tools or techniques have you used at work?

"Manuel matching, using knowledge graph tools and Semantics, Educating supplier and customers, implementing a more automated data infrastructure which organizes data.". As the opening quote shows, battling data quality issues can be multi oriented task for any specialist. Most common method to counter data quality issue at hand is by the person's manual intervention, as per based on the author and respondents' experience. Compared to previous question, Question 4.3.4 was created to understand whether respondents are using same or different tools or practices in countering data quality issues at work compared to what they know is available. Some overlapping between the two question answers were found, which shows that same tools and techniques that are known are already being used in the respondents' work. However, some differences were also discovered, which shows us that respondent wanted to bring specific ideas that are currently in use and could be improved.

Business intelligence tools were one key take from this survey question and almost every respondent mentioned tool and application names which functionalities they are using at work. Some common and powerful tools include Power BI, Flowcharts, SSMS, and Excel. With business intelligence tools, some data quality issues might be preprocessed in the import query and hard coded to be cleaned by default. Nevertheless, based on responses the tools in the year of writing this research have not been smart enough to completely handle completely perfect extract, load and transformation phase without manual intervention by human.

Manual intervention was another popular theme recognized in the answers. Lot of respondents mentioned that they need and continue to double check report data and key performance indicators manually, when necessary. For instance, they might recgonize based on experience that something might be off on what the analytic tool or report is showing. Manual intervention was however supplemented with two key themes, process improvement and communication. Process improvements could be clear documentation set by the specialist and standardization of data collection procedure, whereas communication meaning for example the collaboration and presentation of the findings to the audience using tools and reports. Indeed, recognizing data problem and starting initiatives based on them with mentioned techniques is a key factor to counter the issue. Additionally, some minor themes also recognized were recognized that could be added in previous themes, such as ERP and error queue.

Summarizing the tools and techniques in use at respondents' work, the generated word cloud was created and is shown in the Figure 17 below. Excel was mentioned as separate entity from business intelligence tool, since the respondents' emphasized often that too much of data handling is conducted manually in Excel and should be processed in more advanced analytics tools.

Figure 17 Type of tools or techniques respondents have used at work

When looking at the demographics defined in 4.2. chapter and the keywords formed in this section, tools were among most popular theme for the technology group. For public, process improvement, manual intervention and business intelligence tools. For services, business intelligence tools, communication and manual intervention. Again, the demographics did not seem to formulate significant differences in tools used to counter the issue, however manual intervention was mentioned fewer times amongst the respondents grouped in the technology industry. This does not exclusively mean that manual intervention is not used by the target group, but it was less considered option when thinking about the survey question. When looking at the tools and techniques used at work, innovative ways to counter the issue can be identified and solutions to the problems help to understand the challenges presented in the research question as well.

4.3.5 How do you feel the data quality effects on reliability of the reports?

"The quality of information is everything. If the information is not of high quality, it is not reliable, and the reporting is not correct.". As mentioned by one of the respondent, quality is everything, and the challenges it provides is significant as even in reporting, the audience and stakeholders are in close touch with analysing data. In this context reports and reporting are also used as one form of analytic tool. Another candidate brought up interesting comment, where in some situations one comprehensive qualitative answer in a report could bring more value than dozens of quantitative answers visualized without proper context or elaboration to the decision maker. Truly, this could be truthful and valid in situation when the respondent is known and reliable, like certified expert of the field, experienced user of certain service or a creator of certain product.

Question 4.3.5 was created to understand how respondents evaluate the importance and impact of data quality in reports (that specialists are using in their work to make decisions). Most common themes recognized in this answer were "highly", "reliability", "relationship" and "employee engagement". As the answers show, the specialist find the relationship between data quality and reliability of reports very high. In some cases, the reliability and availability of correct data might have direct effect on employee engagement and morale, as it was pointed out by some respondents that poor reports demotivate stakeholders in their job and interest in company. Additionally, relationship theme was mentioned, as in addition to engagement, the relationship between a company and its suppliers or external stakeholders might be affected by unreliable report. This question had least variety between respondents, and mutual deduction could be made that good data quality is basis for any report. Summarizing the data quality's effect on reliability of the reports, the generated word cloud was created and is shown in the Figure 18 below.

Figure 18 Data quality effect on reliability of the reports

4.3.6. How do you feel the data quality effects on daily operational and/or periodical tactical decision-making?

"The poor data quality can affect the decision-maker's ability to judge the real situation. For example, mistakes in financial analysis can give the impression that the quarterly results were better than in reality.". As opening quote suggests, poor data quality clearly poses challenges and high risks for specialists using analytics tools and reports for decision-making. Poor data quality can falsify can give wrong picture to the business users. Especially in daily decision-making, decisions need to be made within same day and any concern risen during the decision-making process and data quality might hinder, slowdown or even stop a person from making a decision. This furthermore damages the productivity of the company as many business operations are depending on decision made by specialist, team leaders, process owners or other persons in charge.

Additionally, falsely prepared data for instance in financial indicators can have huge impact and give a distorted impression to all stakeholders.

Question 4.3.6 was created to understand how respondents evaluate the importance and impact of data quality in purely decision-making, focus on micro-level decisions. Identified themes were: “highly”, “productivity”, “quality”, “tacit and silent information”, where two latter ones were negative case analysis compared to other answers. Common conclusion was that the relationship to decision-making is high, and that the productivity is in jeopardy, when analytic tool is providing information with poor data quality. Another quote to emphasize the issue, with the word worried in this context: *“Daily work is suffering from poor data, it brings uncertainty and extra workload. We always have to be worried.”*.

One answer that the author found interesting and important was the impact of tacit knowledge and orally transferred knowledge when it comes to decision-making. Direct translated quote from Finnish this respondent had was as follows: “*There is no need for comprehensive reports in everyday work. Still, information must be correct, but it can also be oral or tacit knowledge.”*. Indeed, the author with his almost ten years of work experience on different fields agrees so called silent information is highly powerful within different work environments. This point was presented, because it was exceptional mention compared to other answers. However, as part of study no comprehensive deduction can be made, and comparison of tacit knowledge and use of analytical tools should be studied further. Summarizing the data quality’s effects on decision-making, the generated word cloud was created and shown in the Figure 19 below.

Figure 19 Data quality effects on decision-making

4.3.7 What would you define as best practices to counter bad data quality (or to improve data quality?)

"Clear responsibility of the data quality and ownership. KPIs and communication of the issue." & *"Ensure that all involved parties have the same understanding of data and measurement. Cleanse the data before doing the actual analysis. Cross-check the outcomes."*. Two quotes chosen summarizes key findings from the empirical part. Question 4.3.7 was created to understand and summarize thoughts on best practices to counter bad data quality per the respondent's opinion and experience. This question also helped us to gather deeper understanding of detailed ways to improve data quality and most importantly, help to form worthwhile answer to the research question. Summarizing the best practices to counter bad data quality or to improve data quality, the generated word cloud was created and shown in the Figure 20 below.

Figure 20 Best practices to counter bad data quality or to improve data quality

ProcessOwnership tag was formed to indicate that there needs to be clear responsible specialist or team for ensuring the data quality. Additionally, it was clear according to the respondents that some form of ownership needs to be formed around the process where data is flowing from one party to another. Author finds that the ownership is created in form of working title or responsible team within the organization. With established process ownership, continuous improvement or change initiatives can be followed through much more easier as there is one specialist who can lead the change. What is more, it can help in maintaining system and automation of information retrieval, as well as the quality guidelines and the quality control with the process.

ProcessOptimization tag means the process itself need to be clearly optimized with defined documentation, guidelines and more. Some of the keywords included to the optimization tag were transparency in reporting determination, data cleansing, quality controls, quality guidelines and instructions of data handling. Additionally, after implementing ownership, recognizing which data is useful to which business analytic tool, and establish suitable indicator to solve specific business problem. Integrating different data sets to analytic tool with clearly defined data model.

Communication includes some aspects that are part of process ownership, such as making sure the employees are aware of teams responsible for providing and handling different data sets. Additionally, communication holds key in telling employees the importance the data plays when data is being entered manually to the system. For instance, in situations where operators have direct access to data tables can modify or enter data. One keyword, “Training”, correlates with communication as many respondents mentioned training the users as one of the key techniques in countering the research problem. Monitoring was mentioned once as a keyword, however it indirectly relates to the other themes, as monitoring is sometimes key reason to create a KPI report or analytical tool. Additionally, one respondent emphasized that information should be based on scientific research, when possible, which is interesting intake and truthfully valid characteristic for any data entry.

4.3.8 Additional comments

Question 4.3.8 was created to give an opportunity for the respondent to add anything that might come to their mind not handled in the earlier questions. Only two additional comments related to research question were given and author decided to present them as quotes: *“Overall, without data quality, decision-making is similar to making predictions without proper evidence. Moreover, tools and technologies built not considering data quality are a waste of resources.” “…overemphasis on tools and technology rather than proper training and awareness on how data quality can make impact in analytics.”*.

4.4 Results summary

Ishigawa diagram of the problem and solutions presented in the Figure 21 was formed as a result of the answers. Conclusions from the diagram show that main contributors to bad data quality are lack of responsibility, wrongly configures business or technical process, lack of communication, incompleteness of data, user awareness and manual data entries. Empirical solutions indicate transparent definitions, talent management with analytics tools and improving communication. Theoretical findings recommend assessing the data and data model periodically, educating employees about costs of bad quality, developing data-driven strategy and data quality assessment framework. Also, the most important factors were links with theoretical as well as empirical findings, which are establishing process with clear roles, user training, standardized data collection models, data cleansing processes and automatization of manual processes.

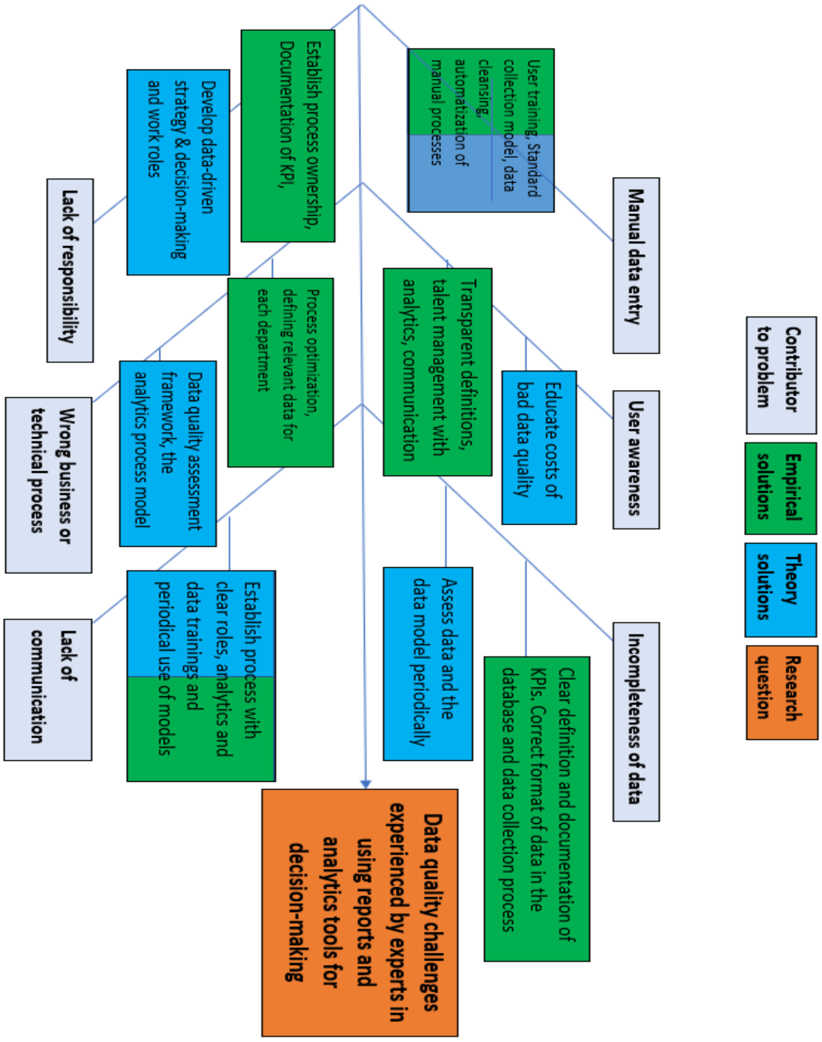

Figure 21 Ishilawa diagram of problem and solutions

5 Discussion

5.1 Primary and secondary data

The primary data for this thesis included qualitative survey conducted with form and seven open questions breaking down the research problem in unique manner from different perspectives. The secondary data included previous studies published in journals, articles and books related to data, data quality, business intelligence and decision-making. "Verification is the process of checking, confirming, making sure, and being certain." (Morse et al. 2002). Further verification of the survey results was conducted by applying respondent validation, which Guest et al. (2014, pp. 79-106) & Flick (2013, pp.511-512) claim is conducted by asking participants to review summarized data and the reflection between author's interpretation of the answers. Author used the form of respondent validation before and after the survey was finished. For instance, before the survey instructions and description of the problem were given in cover letter, and concepts in the research context were elaborated personally when needed.

Few participants used the opportunity to get more elaboration to the questions and thus got clearer understanding that their answers are valid and important for the study. Afterwards, few participants of the qualitative survey were contacted again with follow-up discussion and by introducing interpretation of their answer and discussion of generalized summaries of the findings. Verification process did not end in disagreement, therefore the discussion with the experts who had time for follow up discussion gave more confidence in the researcher's analysis and interpretation. However not all respondents were reached, and further validation could have been done with second round of questions or by interviewing all candidates with more in-depth questions. Nevertheless, the verification with experts gives more notion to the qualitative answers.

5.2 Key findings

Key findings are similar in both theory and empirical study, such as issues with data quality, data characteristics, and data-driven decision-making and their impact on reliability of the reports. When discussing the first theoretical subject, data, empirical findings found similarities in data creation process. As presented by Kelleher & Tiener (2018, p. 74) and Mahanti (2019) in Figure 1, data creation can be initiated from a manual data entry, data acquisition or electronic data capture. Additionally, past studies show same problem to be one of the root causes for bad data quality, for instance especially Lennerholt et al. (2023) study showed that data reports made by inexperienced users can be even more problematic, since additional users might try to make decisions or insights based on "erroneous reports". Similarly, the respondents in the qualitative survey claimed that many bad data quality issues arise from manual data entries or even wrong procedures in automated data collection models. According to study by Lindström et al. (2023), same main causes are found in production planning industries. For instance, data format and human errors may occur due to inaccurate data entries and more. Furthermore, Mahanti (2019) presented data processing and access points for users, which were discussed by the respondents in form of unoptimized processing of data. Sharda et al. (2018, p.494) also claimed that data should be readily available in standardised format, which was one of the key findings in the empirical study.

Data storages presented in the theory, for instance by Kelleher & Tienery (2018) as well as Aktas et al. (2019) claimed that for example cloud technology enables analytics across organization and stakeholders. Data storage as concept was not discussed in the empirical findings, but similar link could be created with the cloud technology, as two respondents mentioned cloud technology as powerful tool for data cleansing process, and thus as one of important techniques to improve data quality for reports and analytics tools. According to Desai (2018), strategic collaboration is usual way to improve productivity, communication and data sharing. Strategic collaboration was another concept not recognized with undirect question from the respondents, however communication, training and process optimization were recognized as key themes and could be linked to the claim. Communication was continuously emphasized by the respondents as prevalent key in solving the challenges when using analytic tools.

5.3 Data quality factors

Speaking of data quality, the theoretical framework identified sources of data quality as “the holistic quality of data, including their accuracy, precision, completeness, and relevance” (Sharda et al. 2014, p. 666), whereas Mahanti (2019, p. 9) presents data quality as “the capability of data to satisfy the stated business, system, and technical requirements of an enterprise”. Bad data quality could therefore consider the opposite of these definitions in theoretical point of view, inaccurate, not precise, incomplete, irrelevant, unsatisfying to business needs or technical requirements of the organisation. Additionally, Sebastian-Coleman (2013, p.48) presented data quality dimensions in five different categories that could be utilized in understanding data quality sources. In the empirical part we could recognize similar patterns when asked to describe bad data quality. For example, as illustrated in Figure 14, the key themes formed included lack of data, wrong process, human errors, external data sources, lack of documentation. Also, as stated by one of the respondent: *“The data is not precise enough (doesn't tell the whole truth, could be wrongly interpreted). Data is not consistently gathered, results are falsified by wrong parameters. Old data is not updated”*. When discussing desired features of good data quality, the theory offered adjectives such as correct, complete, valid, reliable, consistent, unique, current, contextual, pertinent, comprehensive, easy to read, unambiguous, easy to understand and right level of detail. In empirical study these characteristics could have been part of process optimization, automatisation and training.

Causes of bad quality was illustrated by Mahanti (2019, p.15) & in Figure 3 of this thesis. Similarities to the responses from the qualitative survey were found, for instance data decay, manual data entries, inadequate validation in data capture process, inefficient business process, loss of expertise, business data ownership, lack of common data standards, data cleansing programs and lack of shared understanding of data amongst many things. Some causes that were not discovered during the empirical research were: data migration, system upgrades, governance issues, data corruption by hackers and organizational changes. Although, these concepts might be familiar to the respondents' they were not directly mentioned in any of the questions. As there is clear link, at least the first part of the key causes of bad quality in both theoretical and empirical study can be validated as valid sources to bad quality.

Costs of bad data quality were more generic in empirical research compared to theory. More perspective could have been studied with more specified question related to the costs of bad quality, however despite being important effect, not complete scope. Further, the costs of bad quality could be studied in another research with more quantitative methods. Nevertheless, Mahanti (2019, p. 41) & Haug et al. (2011) claimed that bad data quality leads into increased costs, organizational and customer satisfaction, productivity issues, risk and compliance. Additionally, according to study conducted by Choi & Luo (2019), direct link between profitability of supply chain and poor data quality was found. Aforementioned points were also recognized in some themes presented in the Chapter 4, such as productivity and relationships. These findings prove significant correlation between bad data quality and productivity, which includes all the costs related to it. Additionally, cost of poor quality can be loss of relationship or reliability between trade partners or even internal stakeholders.

Conversely, it was stated that good data quality tends to significantly improve decision-making and lead to more cost improving activities in different companies (Bechman 2022; Marshawn 2022; Fagan 2012; Hughes & Forrest 2012). Additionally, respondents' answers highly indicate that the trust to data to be good data quality matters when handling analytical tools and evaluating daily decision-making. Like Scott (2017) emphasized good data quality on financial reports, which many of respondents noted during their answers.

5.4 Business intelligence factors

Business Intelligence was widely discussed in theoretical part and the qualitative part. In the theoretical framework, Kelleher & Tierney (2018, pp.72-73) and Kohtamäki (2017, pp.23-24) claimed business intelligence as decision-support system designed to produce data modelling, integration of systems, and to provide analytics and reporting possibilities for decision-making. Respondents of the survey highlighted similar findings and we can argue that business intelligence integration within business processes and decision-making is crucial to efficient and productive work. However, data quality flowing through the business intelligence system must be validated and up to standards as per findings in the research. Whiting (2006) mentioned data quality systems and for instance using Lean Six Sigma in analyzing data collection.

When discussing about visualization opportunities, no major findings were conducted during the empirical research. However, it is important to point out the findings in the theoretical part, for instance mentioned by Sharda (2018, p.494) and Kumar (2017, pp. 32, 53), where data visualization plays one of the key part in analytics and analysis of data. Only link to the empirical find can be argued is the representation of bad data in visuals can lead to wrong conclusion for the user and decision-maker. Nevertheless, author believes also visualization plays pivotal role in data analytics and analysis tools.

Risks of sharing data was not fully identified in the empirical research part, as the qualitative questions did not ask it directly. However, as it was addressed in theoretical framework by Desai (2018, pp.221-222), where Desai claimed that stakeholders might take advantage of data sharing also in bad scenarios. Nonetheless, in the empirical noteworthy comments were taken from the respondents when talking about bad data quality that could be linked to this claim was that supplier relationship could be damaged by poor data quality reports or decision made with poor data quality. What is more, in answered by one of the repsondents, in some cases bad data quality can lead to poor employee engagement or employee turnover. High standards of employee involvement can be achieved through total quality management framework, which provides tools for cultural change and team empowerement (Krajewski, Malhotra & Ritzman, 2021, p. 128). When applying the known data quality strategies in total quality management system, author believes a strong Process Ownership and continuous training can be achieved to ensure data quality in business intelligence tools. Additionally, one of the respondents brought up that the data flow could not be automatised or corrected, because the data handling process was completely in their supplier's hands, making data sharing role as more receiving one.

5.5 Links to theoretical framework and decision-making

Although not mentioned in the empirical study, the data quality assessment framework introduced by Sebastian-Coleman (2013) could be linked to the ideas brought up by respondent group in the qualitative research. For instance, we can see familiarities and characteristics in the ideas from process ownership and process optimization. Sebastian-Coleman presents the solution in more systemized and detailed manner, whereas the specialists view is practical and generalized view. Nevertheless, as illustrated in Figure 7, initial one-time assessment, automated process controls and periodic measurement types were also brought up in the empirical research part. For instance, one respondent stated with optimal technique to counter the bad quality as: *"Proper systems, Master data in order, automation of information retrieval. Without functioning systems and high-quality master data, it is very difficult to get reliable information out, no matter how you automate information retrieval."*, which could very well be achieved with Sebastian-Coleman's framework. Additionally, study by Dragos (2021) proposes process owners and specialists to have back-up for daily activities preventing errors and mitigating challenges.

Supporting questions that were used in the qualitative survey and helped answering the research question were "What are the data quality challenges experienced by experts in reporting and analytics?" and sub research questions "How are the data quality challenges tools reflecting in reliability and impact to decision-making?" and "What are the best techniques and analytics tools when encountering bad data quality in decision-making?". With these questions we also got to understand how experts are dealing with the problem as well as think what best techniques are in countering the issue, which brings valuable data to all specialists struggling with the issue. Both supportive questions helped to form qualitative survey questions and answers provided by the respondents were rich and analyzed according to qualitative analysis method.

When discussing decision-making, differentiation was conducted in the theory according to Lemahieu et. al (2018, pp.552-553) and focus to operational and tactical decision-making was made. This point of view was acknowledged by the respondents when introduced in the qualitative survey questions and answers were given accordingly. Reflection to decision-making was recognized to be very high in both empirical and theoretical findings. Additionally, study from Austin et al. (2021) suggested that when considering analytics and different regulations in audit situations the auditors, although with mixed results, shift in influencing data analytics through different channels to modify data analytics rules.

Author has aimed to use in addition to his own work experience as reporting, system and analytics specialist. Validity and reliability were ensured by using qualitative handbook and the methods validated in them when building the survey and analysing the answers. Additionally, the respondent pool was diverse, qualitative data was collected from 14 different people, from specialist from several industries, with working titles and three different countries were represented. What is more, all of the respondents replied that they are using analytics tools at work. This gave rich context to the research problem, as individuals from these different backgrounds were able to provide different perspectives. Also, communicative cover letter was created and introduced in the qualitative survey as well as when presenting the research problem to potential interviewee candidates.

Qualitative questions were neutral and open-ended, without character limitation for the respondent, meaning they could write and type as long answer they felt like. What is more, negative case analysis was conducted and outlier answers in relation to rest of the data were presented in the study. Reliability could have been improved if qualitative survey was conducted in form of interviews as the answers could have been discussed more deeply, furthermore the respondent could have had chance to elaborate, or the interviewee (author) have change to add follow-up questions that might have risen during the discussion. Nevertheless, the respondents were given contact detail of the author and offered to stay in touch and ask for clarification in case some questions or concepts were unclear.

6 Conclusions

6.1 Summary

This study collected primary data with qualitative method from specialists from fourteen different workplaces and organizations, as well as three different countries with some in managerial positions, but majority at least directly producing dashboards, analysing data and providing insights for their own and managers' decision-making. Data was analyzed by applying thematic analysis and by forming word clouds of most recognized themes and concepts in the answers. Secondary data included previous studies with similar topic, journals and books regarding data, data quality and business intelligence. Research question: *"What are the data quality challenges experienced by experts in reporting and analytics?"* can be answered by respondents' unique experiences in relation to research problem was carefully analysed and verified with respondent validation. Countering the data quality issue can be achieved by optimizing processes, establishing process ownerships, raising user awareness, defining clear procedures and training personnel about data importance and quality. Furthermore, additional findings in theory findings in costs for bad data quality and its significant impact in productivity, customer satisfaction and increased costs in organizations. Conversely, the importance of good data quality to decision-making and its positive impact in different sectors could not be overstated. Furthermore, secondary data from previous studies, journals, articles and books provided similar results. Therefore, main recommendation author can make is establishing process ownership and applying the data quality assessment framework in the company's context as best tool to counter the issue.

Sub questions: *"How are the data quality challenges tools reflecting in reliability and impact to decision-making?" and "What are the best techniques and analytics tools when encountering bad data quality in decision-making?".* can be answered that the reflection is very high and impactful. Data quality challenges of analytics tools reflect in decision-making highly, and often can lead to productivity issues as well other drastic consequences. This sub question was to assess the significance of the research problem and its impacts, but also to highlight the importance and the need for managerial implications struggling with the issue. For the second sub research question answer was formed along with Ishigawa diagram in Figure 21. Main points from the diagram highlighted that main contributors to data quality issues through are lack of responsibility, wrongly configures business or technical process, lack of communication, incompleteness of data, user awareness and manual data entries. Also, the significant factors to overcome the issues were using analytics tools, such as SAP Analytics Cloud, Salesforce, Tableau and Power BI, creating internal data process mapping with clear roles, conducting user education in data quality impacts, and creating standardized data collection models.

Thesis results can be used to assess data quality challenges for companies having tools with manual entries as well. Afterall, according to empirical findings and studies, for example Lindström et al. (2023), Sharda et al. (2018, p. 130) causes of bad data quality arise often from manual entries and when using multiple sources of data, which creates conflicts for decision-makers to trust the report. Tools to counter the data quality issues by forming data template standardization, clear documentation of indicators, training of the employees using and inserting data, and establishing data assessment ownership and framework in the company. Additionally, summary of good quality data files by Towse, Ellis & Towse (2021) strengthens points discovered in the research. For example, data file must be available, complete, labelled and elaborated, as well as having proper meta data.

6.2 Managerial implications

The broad empirical findings and experiences of specialists in facing data quality challenges with reports and analytical tools gives already information to employers as well as awareness and understanding of this contemporary issue. As the results cannot be formed into one explicit solution, to illustrate and help the managers to assess the research problem, “data quality challenges experienced by experts in using reports and analytics tools for decision-making”, author created an Ishikawa diagram combining findings from both, empirical and theoretical part. Figure 21 represented earlier, and the findings can be used as a tool to initiate the improvements in case organizations.

Conclusion from Ishikawa diagram shows practical steps for manager per contributor to the research problem. Identified data quality issues due to manual data entries, theory and empirical strongly suggests user training, which the company can do by holding trainings of the data collection process in Teams and workshops. Additionally, data collection model for user inputs should be standardized with correct data format and compulsory fields. For instance, Also, data cleansing and automatization of manual processes should be conducted when possible. This can be achieved already at basic level using Microsoft services, such as Power Query, Power BI and Power Automate. These or similar tools should be merged into the business process collecting the data.

Lack of responsibility is another key contributor to the problem. Manager should be informed to have and develop data-driven strategy and applying in the decision-making as well as work roles. Work roles should clearly define process ownership including the data handling phase, clear documentation of performance indicators. Additionally, user awareness regarding the data quality should be raised by trainings of impacts of bad data quality in form of numbers. As the study findings show, there is direct impact on report reliability, operations productivity, decision-making and operating figures of the company when the challenges are not addressed.

When process ownership is established, it should be optimized to answer to the business department to avoid gathering irrelevant data. This can be achieved by setting standard data and key performance indicators meaningful to that business process. The relevancy should be documented into process flow that should be reviewed periodically and different models, such as data quality assessment framework and analytics process model can be utilized to form it. This will help to address rest of the contributors: lack of communication, incompleteness of data as well.

6.3 Future research opportunities

Business intelligence is contemporary topic that can be studied from various angles. Future research opportunities were identified by the author that could form great research topic for example as a master thesis, for people interested in the subject. As the qualitative study brought up good analysis on word level, some study should be conducted by quantifying the same problem or by doing a case study. Also, it would be very interesting study if the survey or interviews were directed directly to a certain group of specialists in same industry, or for instance studying and comparing results from big sample of data engineers, business analysts or team leaders. What is more, topic could be research with focus only in small or startup companies. Additionally, the topic could be conducted as a case study for a company struggling with data quality and planning to implement new data quality management or business intelligence process or a system. For instance, the challenges specialist face in using analytics tools, data quality issues, techniques used would in Author's opinion work in qualitative survey. Some hypothetical research questions based on this study's results could be:

- *"How does Company X use business intelligence tools for decision-making?"*
- *"Implementation of practices against bad data quality in Company X"*
- *"Impact of good and bad data quality for small companies and startups."*
- *"What are the best data cleansing practices in manufacturing industry X in Finland?"*
- *"Measuring costs of bad data quality in Company X"*

7 References

Aktas, E., Bourlakis, M., Minis, I., & Zeimpekis, V. (2021). Supply Chain 4.0: Improving Supply Chains with Analytics and Industry 4.0 Technologies (1st ed.). Kogan Page.

Andriof, Jörg, et al. (2003) Unfolding Stakeholder Thinking 2: Relationships, Communication, Reporting and Performance, Taylor & Francis Group, ProQuest Ebook Central

Andrea S. Towse, David A. Ellis & John N. Towse (2021). Making data meaningful: guidelines for good quality open data, The Journal of Social Psychology, 161:4, 395-402, DOI: 10.1080/00224545.2021.1938811

Applied Thematic Analysis. (2023, March 2). Sage Publications Inc. https://us.sagepub.com/en-us/nam/applied-thematic-analysis/book233379

Austin, A. A., Carpenter, T., Christ, M. H., & Nielson, C. (2021). The Data Analytics Journey: interactions among auditors, managers, regulation, and technology. Contemporary Accounting Research, 38(3), 1888–1924. https://doi.org/10.1111/1911-3846.12680

Bechman, Tom J (2022). Data-driven decisions must start with good data. Corn and Soybean Digest; Overland Park. Trade Journal.

Bhandari, K., Kumar, K., & Sangal, A. L. (2022). Data quality issues in software fault prediction: a systematic literature review. Artificial Intelligence Review. https://doi.org/10.1007/s10462-022-10371-6

Bourne, L. (2009). Stakeholder Relationship Management: A Maturity Model for Organisational Implementation (1st ed.). Routledge.

Botes, L.-A. A., Hamer, W., van Jaarsveld, S., & Kleingeld, M. (2019). FINDING THE FOUR QUALITIES OF INTELLIGENT INDUSTRIAL REPORTING. The South African Journal of Industrial Engineering, 30(3), 262–276. https://doi.org/10.7166/30-3-2235

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, 14(0), 2. https://doi.org/10.5334/dsj-2015-002

Callahan, T. J., Barnard, J., Helmkamp, L., Maertens, J. A., & Kahn, M. G. (2017). Reporting Data Quality Assessment Results: Identifying Individual and Organizational Barriers and Solutions. EGEMS, 5(1), 16.

Carroll, A. B., & Buchholtz, A. K. (2008). Business and Society: Ethics and Stakeholder Management (7th ed.). South-Western College Pub

Choi, T., & Luo, S. (2019). Data quality challenges for sustainable fashion supply chain operations in emerging markets: Roles of blockchain, government sponsors and environment taxes. Transportation Research Part E-logistics and Transportation Review, 131, 139–152. https://doi.org/10.1016/j.tre.2019.09.019

Dakka, M. A., Nguyen, T. V., Hall, J. M. M., Diakiw, S. M., VerMilyea, M., Linke, R., Perugini, M., & Perugini, D. (2021). Automated detection of poor-quality data: case studies in healthcare. Scientific Reports, 11(1), 18005. https://doi.org/10.1038/s41598-021-97341-0

Desai, V. M. (2018). Collaborative Stakeholder Engagement: An Integration Between Theories Of Organizational Legitimacy And Learning. Academy of Management Journal, 220-244. Eskerod, P., & Jepsen, L. A. (2013). Project Stakeholder Management (Fundamentals of Project Management) (1st ed.). Routledge.

DeSimone, J. A., & Harms, P. D. (2018). Dirty Data: The Effects of Screening Respondents Who Provide Low-Quality Data in Survey Research. Journal of Business and Psychology, 33(5), 559–577. https://doi.org/10.1007/s10869-017-9514-9

Dragos, D. (2021). Project Management Challenges in Implementing Business Intelligence Approach inside Organizations. The Romanian Economic Journal, 79, 66–74. https://doi.org/10.24818/rej/2021/79/05

Fakhouri, H. N., Al-Aamr, R., Jabbar, S. A., & Fakhouri, H. N. (2021). Business intelligence in academic libraries in Jordan: Opportunities and challenges. IFLA Journal, 47(1), 37–50. https://doi.org/10.1177/0340035220931882

Flick, U. (2013). The SAGE Handbook of Qualitative Data Analysis. SAGE Publications Limited.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). Applied Thematic Analysis. SAGE.

Haug, A., Zachariassen, F., & Liempd, D. V. (2011). The costs of poor data quality. Journal of Industrial Engineering and Management, 4(2). https://doi.org/10.3926/jiem.2011.v4n2.p168-193

Hughes, Jonathan, PE; Council, Forrest M, PHD (2012). Institute of Transportation Engineers. ITE Journal; Washington Vol. 82, Iss. 4: 14-18.

Iyer, L. S., & Power, D. J. (2014). Reshaping Society through Analytics, Collaboration, and Decision Support: Role of Business Intelligence and Social Media (Annals of Information Systems, 18) (2015th ed.). Springer.

Jody Condit Fagan (2012) Editorial: Connecting the Dots Between
Good Data and Good Decisions, Journal of Web Librarianship, 6:2, 87-93, DOI:

10.1080/19322909.2012.680333

Kahn, M. G., Brown, J. S., Chun, A. T., Davidson, B. N., Meeker, D., Ryan, P. B., Schilling, L. M., Weiskopf, N. G., Williams, A. E., & Zozus, M. N. (2015). Transparent reporting of data quality in distributed data networks. EGEMS, 3(1), 7. https://doi.org/10.13063/2327-9214.1052

Kelleher, J. D., & Tierney, B. (2018). Data Science (The MIT Press Essential Knowledge series) (Illustrated ed.). The MIT Press.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures & their consequences. SAGE Publications Ltd https://www-doi-org.proxy.uwasa.fi/10.4135/9781473909472

Kohtamäki, M. (2017). Real-time Strategy and Business Intelligence: Digitizing Practices and Systems (1st ed. 2017 ed.). Palgrave Macmillan.

Krajewski, L., Malhotra, N., & Ritzman, L. (2021). Operations management: Processes and supply chains, eBook, [GLOBAL EDITION] (13th ed.). Pearson Education.

Kumar, D. U. (2017). Business Analytics: The Science Of Data - Driven Decision Making. Wiley India.

Laudon, J. P., & Laudon, K. C. (2017). Management information systems: Managing the digital firm, global edition (15th ed.). Pearson Education.

Lemahieu, W., Broucke, V. S., & Baesens, B. (2018). Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Big and Small Data (1st ed.). Cambridge University Press.

Lennerholt, C., Van Laere, J., & Söderström, E. (2021). User-Related Challenges of Self-Service Business Intelligence. Information Systems Management, 38(4), 309–323. https://doi.org/10.1080/10580530.2020.1814458

Lindström, V., Persson, F., Viswanathan, A. P. C., & Rajendran, M. (2023). Data quality issues in production planning and control – Linkages to smart PPC. Computers in Industry, 147, 103871. https://doi.org/10.1016/j.compind.2023.103871

Mahanti, Rupa (2019). Data Quality : Dimensions, Measurement, Strategy, Management, and Governance, Quality Press. ProQuest Ebook Central

Marshawn, W. (2022). Indianapolis Business Journal; Indianapolis Vol. 43, Iss. 7,: 5C.

Morse, J. M., Barrett, M. P., Mayan, M., Olson, K., & Spiers, J. (2002). Verification Strategies for Establishing Reliability and Validity in Qualitative Research. International Journal of Qualitative Methods, 1(2), 13–22.
https://doi.org/10.1177/160940690200100202

Nash, Kim S. (2009) CIO. Make a Difference Before the Cuts Come: Good
data from business intelligence systems is essential to deciding where to make reduction. Framingham Vol. 22, Iss. 12.

Oakland, J., & Oakland, R. J. (2018). Statistical Process Control (7th ed.). Routledge.

Ozmen-Ertekin, D., & Ozbay, K. (2012). Dynamic data maintenance for quality data, quality research. International Journal of Information Management, 32(3), 282–293. https://doi.org/10.1016/j.ijinfomgt.2011.11.003

Samitsch, C. (2014). Data Quality and its Impacts on Decision-Making: How Managers can benefit from Good Data (BestMasters) (2015th ed.). Springer Gabler.

Sebastian-Coleman, L. (2013). Measuring data quality for ongoing improvement: A Data Quality Assessment Framework. Morgan Kaufmann Publishers.

Robertson, Scott. (2017). Use good data to drive your business. Supply House Times Preview publication details; Troy Vol. 59, Iss. 12, (Feb 2017): 44.

Sharda, R., Delen, D. & Turban E. (2018). Business intelligence, Analaytics, And Data Science: A Managerial Perspective, Global edition (4th ed.). Pearson Education.

Sharda, R., Turban, E., Delen, D., Efraim Turban, Dursun Delen, Aronson, J. E., Liang, T. P., & King, D. (2014). Business Intelligence and Analytics. Pearson.

Seguer, Z. S., & Hasna, A. M. (2022). Business Intelligence as a Challenge for the Managerial Function: Case Study on Managerial Decision Making in Algerian Companies. Business Ethics and Leadership, 6(3), 35–46. https://doi.org/10.21272/bel.6(3).35-46.2022

Stobierski, T. (2021). 8 Steps in the Data Life Cycle | HBS Online. Business Insights Blog. https://online.hbs.edu/blog/post/data-life-cycle

Suppiah, K., & Arumugam, D. (2023). Impact of data analytics on reporting quality of forensic audit: a study focus in Malaysian auditors. E3S Web of Conferences, 389, 09033. https://doi.org/10.1051/e3sconf/202338909033

Whiting, R. (2006). When good data is the bottom line. InformationWeek; Manhasset Iss. 1088, (May 8, 2006): 44.

Winston, W.L. (2021). Analytics stories: Using data to make good things happen. Indianapolis, IN: John Wiley & Sons, Inc.

Appendix 1 – Data collection cover letter

Greetings! I am a master's student in economics at the University of Vaasa. I'm doing a degree about data quality and decision-making. For research, I need answers from experts in the field who use analytics tools and data in decision-making in their work. This research work focuses on operational and tactical decision-making, which refers to daily decisions made by experts based on reports and data, among other things. I selected you as a potential respondent based on your LinkedIn job description and I believe you have valuable information for my research. The answers are given in an electronic Google Form and the answers are qualitative. Even a short answer is enough, and even a one-sentence answer is important. It takes approximately 15 minutes to give the answers. I treat the answers confidentially and do not pass them on to outsiders. Your answer will be anonymous, and I only use it for research analysis, so your identity cannot be revealed. Please reply if you have any additional questions, I'll be happy to tell you more.

Tervehdys! Olen kauppatieteiden maisteriopiskelija Vaasan yliopistossa. Olen tekemässä gradua aiheesta datalaadun vaikutus päätöksenteossa (Data quality and decision-making). Tarvitsen tutkimusta varten vastauksia alan asiantuntijoilta, jotka käyttävät työssään analytiikkatyökaluja ja dataa päätöksenteossa. Tässä tutkimustyössä keskitytään operatiiviseen ja taktiseen päätöksentekoon, joilla tarkoitetaan muun muassa asiantuntijoiden tekemiä päivittäisiä päätöksiä raporttien ja datan perusteella. Valitsin sinut potentiaaliseksi vastaajaksi LinkedIn työkuvauksen perusteella ja uskon, että sinulla on arvokasta tietoa tutkimukseeni. Vastaukset annetaan sähköiseen Google Form lomakkeeseen ja vastaukset ovat laadullisia. Lyhytkin vastaus riittää ja yhden lauseen vastauskin on tärkeää. Vastausten antamiseen menee arviolta noin 15 minuuttia. Käsittelen vastaukset luottamuksellisesti, enkä luovuta niitä ulkopuolisille. Vastauksenne on nimetön ja käytän niitä ainoastaan tutkimuksen analyysia varten, joten henkilöllisyytenne ei voi tulla ilmi. Vastaathan jos sinulla tulee mieleen lisäkysymyksiä, kerron mielelläni lisää.

Appendix 2 - Qualitative survey questions

- What is your country of work? / Missä maassa työskentelet?
- Work experience in years / Työkokemuksesi?
- Describe briefly your job position and responsibilities at work? / Kuvaile lyhyesti sinun toimeenkuva ja vastuualueesi töissä?
- When do you use business intelligence tools for data analysis or decision-making? / Missä tilanteissa käytät data-analyystyökaluja tai raportteja päätöksentekoon?
- What kind of data quality issues do or have you face(d)? / Millaisia tietojen tai datan laatuongelmia sinulla on tai olet kohdannut?
- What type of tools or techniques can be used to solve data quality issues? / Millaisia työkaluja tai ratkaisuja voidaan käyttää tietojen laatuongelmien ratkaisemiseen?
- What type of tools or techniques have you used at work? / Millaisia työkaluja tai ratkaisuja olet käyttänyt työssäsi?
- How do you feel the data quality effects on reliability of the reports ? / Miten koet tiedon laadun vaikutuksen raporttien luotettavuuteen?
- How do you feel the data quality effects on daily operational and/or periodical tactical decision-making? / Miten koet

tiedon tai datalaadun vaikutukset päivittäiseen operatiiviseen ja/tai säännölliseen operatiiviseen päätöksentekoon?

- What would you define as best practices to counter bad data quality (or to improve data quality?) / Mitä määrittelisit parhaiksi käytännöiksi huonon tiedon tai datalaadun torjumiseksi tai datan laadun parantamiseksi?
- Anything more to add? / Lisättävää?