Turun yliopisto
University of Turku

CAPTURING THE VALUE OF UNSTRUC-TURED DATA IN PROCUREMENT

Case study on supplier contracts

Master´s Thesis
in Information Systems Science
Tietojärjestelmätieteen
pro gradu -tutkielma

Author (s)/Laatija(t):
Aleksi Mäkinen

Supervisors/Ohjaajat:
Jukka Heikkilä

19.05.2021
Helsinki

Table of contents

List of figures

List of tables

ABSTRACT

Data is becoming more and more important to organizations. Many studies suggest that unstructured data, which cannot be managed traditionally, would be the new frontier of value-adding data for the organizations and contribute over 80% of the data organizations have available in their operations. This alarms us to turn our focus on this phenomenon and study what type of data the unstructured data is, how organizations should manage it, and why they should start investing into it. The research will study how organizations' procurement function, a function responsible for acquiring the services and goods for business operation, should approach unstructured textual data and include this data into their business analytical processes and what is the expected value of unstructured textual data in procurement analytics.

The research presents a thorough literature review on the topic. Literature covers the main principles of unstructured textual data, key factors when assessing the approach and capability to utilize the data, and which type of value procurement should seek from the unstructured textual data. As part of the theoretical review, we will compose a theoretical framework, which is used to assess the main research questions: How should procurement organizations approach unstructured textual data? What is the expected value of the unstructured textual data in procurement? The research also includes a case study with two procurement organizations, which assess their approach and the value of unstructured textual data in their organization. The specific use case of unstructured textual data these organizations evaluate is supplier-related documents, such as supplier contracts. The findings from the case study will be reflected in the theoretical framework of the research, providing a practical point of view to the research results and conclusions.

We consider the research framework and results to provide valuable information for procurement organizations which consider including unstructured textual data from the supplier-related documents into their analytical processes.

1   INTRODUCTION

1.1   Research motivation

Data has been in the focus of research and business field the past years. Though common, the research community and the business organizations focus more on a traditional type of data, what we can call structured data, data we can model, capture, and understand with traditional methods and models. Even though many studies suggest that unstructured data, which cannot be managed traditionally, would be the new frontier of value-adding data for the organizations and contribute over 80% of the data organizations have available in their operations. This alarms us to turn our focus on this phenomenon and study what type of data the unstructured data is, how organizations should manage it, and why they should start investing into it. (Inmon & Nesavich 2007; Accenture 2014.)

Our research will focus on organizations' procurement analytics function and how unstructured data exist there. Procurement is the function in organizations, responsible for acquiring the goods and services needed for all the operations of an organization. The unstructured data exist in the organizations in various places and forms. This scope makes the research valuable. In the current world's economic situation, organizations must have an understanding of all the relevant data and information, structured and unstructured, and the capability to turn this information into value. The procurement function is often overlooked in organizations, even though it has a critical effect on the organizations' profitability and success. (Moretto et al. 2017; Van Weele 2009.)

There is not much prior literature on utilizing unstructured data in procurement and its potential value, which makes the research interesting. This research aims to understand what type of unstructured data procurement could utilize and how, and what is the value proposition of it. We will assess this by presenting a profound literature review, constructing a theoretical framework, and empirical research using a case study methodology. Our empirical research, with the case study approach, will be scoped into supplier contracts analysis in the selected procurement organizations. We will study the unstructured data found from the supplier contracts. This data is typically unstructured textual data. This scoping will make the empirical analysis more interesting and feasible. Still, in our theoretical review, we will assess the phenomenon in a broader view. Our main research questions are:

1. How should procurement organizations approach unstructured textual data?
2. What is the expected value of the unstructured textual data in procurement?

The research questions will be answered through the theoretical framework and the findings generated from the case study interviews. The theoretical framework will use the main research questions as the base, creating a logical presentation of our studied subject. First, assessing the subject, unstructured data, giving the definitions and main principles related to it. Then, assessing how organizations should approach it by addressing the main requirements and capabilities that would need to be in place. And lastly, assessing why organizations should be interested in it, discussing the business problems it would solve and the value it could provide. The research will conduct a case study with two organizations exploring the possibilities with unstructured textual data in their procurement analytics processes. The case study is done in collaboration with Sievo, an analytics software provider. The case study will give an understanding of how the parties could utilize unstructured textual data in their procurement analytical processes with Sievo. The findings from the case study will be reflected and analyzed using the theoretical framework. This will strengthen our theoretical framework with a practical point of view. The empirical research data is gathered by interviews with procurement analytics domain experts in their organizations, which is characteristic for qualitative research. Case organizations are global multi-billion organizations, which currently are in some stage of incorporating unstructured data into their analytical processes within their procurement function. These organizations are in a relationship with a software company Sievo, and the research is done in collaboration with the parties. The research focuses on assessing these case organizations and their efforts with the un-structured procurement data. With these interviews, research will get a more practical understanding of what type of unstructured data organizations could use, what are the requirements and capabilities related to it, and the potential value of unstructured data in procurement. The conclusions of the research will give answers to the presented research questions.

1.2   Research structure

In chapter 2, we will present key definition related to unstructured data in procurement. We will first define the unstructured data format, its source, and the key principles related to it. After this, we will discuss some additional general data-related factors, such as challenges and maturity level, affecting the availability of the defined unstructured data in the organizations. In chapter 3, we will assess the existing literature related to data and procurement, such as strategy, infrastructure, analytics, and performance. This literature review will be used as a base when constructing the theoretical framework. In chapter 4, we will showcase the capability assessment approach using resource-based-view. The discussed theory in these chapters give us a good literature review for our two research questions. In chapter 5, we will construct the theoretical framework of our research based

on the previously presented literature. This framework will then be used in the empirical research phase. Our empirical research methodology is a case study. In chapter 6, the methodology is discussed with key considerations, and then chapter 7 will analyze the results from the case study interviews. And lastly, in chapter 8, we will summarize the findings from the case study, draw conclusions, answer the main research questions, assess the research's trustworthiness, give recommendations for future research, and discuss the research's contribution.

1.3   Research background

This research is done in cooperation with Sievo, a procurement analytics software company. Sievo has solid experience serving large global procurement organizations and turning their procurement data into value. The selected organizations for the empirical research case study are Sievo's clients, with an ongoing project to incorporate unstructured data into their procurement analytics processes. The empirical case study focuses on these two organizations, and how they see unstructured data in their business environment, and how Sievo's and other providers services could help them. The research is independent from Sievo's business, and all the views and findings are research own.

2   UNSTRUCTURED DATA

In this chapter, we will first define the unstructured data in the procurement context, then discuss where it exists and how it should be captured and managed. After which, we will assess the critical factors in terms of the availability of this data, which are the challenges related to it, the data maturity level, and the data processes.

2.1   Definition of the unstructured data

Organizations are facing a new era of available data and business analytics. A common term to describe these is big data and technologies relating to it. Gartner (2021) defines big data as: "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." Existing literature characterizes this data with 5Vs: Volume – increasing amounts of available data; Velocity – information forming rate near time; Variety – various forms of data, structured, semi-structured and unstructured data; Veracity – trustworthiness, origin, availability, and accountability of the data; Value – which is the added value the data will bring into the business processes. (Demchenko et al. 2013; Russom 2011). In this research, we will focus on unstructured data, which possess all the characteristics of the big data mentioned above.

On a high level, the format of all data can be divided into two: structured data, which can be managed in a traditional way in relational data models presented as numbers, tables, rows, and columns, and unstructured data, which could be understood as an opposite of the above definition for structured data.

The unstructured data can be divided into textual and non-textual data. Example sources of unstructured textual data are emails, supplier contracts, written documents, and web pages. The amount of this data is substantial, and the actual unstructured textual data is the data content from these named locations. The data is created when writing an email, creating the contracts, or a new web context is generated from specific events. There exist only a few constructs when the textual data is created. The unstructured data does heterogeneous and has no standard schemas. There exist no format, structure, or guidelines to define what the data is. Unstructured data has many meanings in different contexts. This makes the management of it difficult, and traditional tools are limited to manage this data asset. Organizations are struggling to make the data available and usable. (Inmon & Nesavich 2007; Lomotey & Deters 2016; Syed et al. 2013) The mentioned non-textual unstructured data then refers to images, colors, sounds, and shapes (Inmon & Nesavich 2007).

We define in our research the unstructured data as unstructured textual data. Negash (2008) argues that all data, including unstructured data, has some structure since all data is treated with some structure, making the data more structured, and this holds. However, our definition for the unstructured textual data enables us to discuss and investigate the formatting process as well. In the Table 1 we have listed examples of data formats, their structure, and possible use case and sources in the procurement context.

Data source	Data structure	Use case
ERP data	Structured	Spend analysis
Supplier contracts	Textual unstructured	Contract analysis
Written documents	Textual unstructured	Prodcut description analysis
E-mails	Textual unstructured	RFx processes analysis, improved information sharing
External (web) data	Textual unstructured	Supplier risk and market analysis
Images	Non-textual unstructured	Satellite images for detailed location analysis

Table 1 Data source, structure and use case classification

Traditionally procurement analytics has been focusing internal data, which is found from the ERPs. The data are the structured transactions and the master data related to these transactions. This data is used in procurement analytics, for example, in spend analysis and supplier management. These processes are an essential foundation for any successful procurement function. (Van Weele 2009.) However, for organizations to enable additional value to their business intelligence and analytics processes, the focus should be shifted towards unstructured data. This type of data is not currently utilized in organizations, though having prominent value for organizations' analytics processes (Inmon & Nesavich 2007; Russom 2011, 27).

2.2   Management of the unstructured data

For structured data, organizations have strategies and processes in place. However, the unstructured data can provide additional value. The question remains how organizations should manage the unstructured data and combine all the distinct data formats and sources in their data infrastructure and business processes. We should examine how organizations could include unstructured data into their data infrastructure and business processes, thus unlocking the potential value.

There are two major high-level approaches to take. We could manage and analyze the unstructured data in its own unstructured environment or alternatively transform the unstructured data to an existing environment, which was primarily designed for structured data. (Inmon & Nesvich 2007.) We are interested in the latter since this environment exists already for processing and analyzing all the data. This environment has already required extensive resources from organizations, so it would be unwise not to capitalize on those investments. Additionally, when combining the structured and unstructured data into the same environment, organizations can receive additional value to their business analytics processes. (Inmon & Nesvich 2007.) Having all the available data in one common place also ensures better results with organization information and business strategy alignment, which is critical for results. (McAfee et al. 2012; Deloitte 2019; Loshin 2013, 31-36.)

We acknowledge integrating the unstructured data as is, without any processing, into a structured environment would not be feasible. It requires lots of effort and definitions in the integration phase to holistically combine the data environment for value-added business analytics. (Inmon & Nesvich 2007; Kassner et al. 2014, 37.) Traditional data management and analysis systems are based on relational database management systems, and by design, are limited to answer to challenges and requirements of unstructured data (Hu et al. 2014, 653; Inmon & Nesvich 2007).

To solve this problem Blumberg (2003) presents few conceptual things to cover. First, we must locate where the unstructured data exists within our organization, and outside of it, and in which format. After locating where the data exists and in which format, we must make it searchable. This means either formatting the data with some structure, for example, applying metadata to textual documents, transforming the for-mat to more structured, or developing techniques to search relevant information from the unstructured content of those documents. Lastly, we should apply context to unstructured data to ensure its' meaningfulness to business when utilized.

Applying metadata to the unstructured data is one process for making the unstructured data available for further analysis. The goal is to connect the unstructured data into the existing systems in a structured manner. In this process, the unstructured data is given structured descriptions linked to the existing structured data models and assets. For example, contracts are treated with creation date, author, and other business-relevant dimensions. (Baars & Kemper 2008; Negash 2008.)

Existing research has also explored how supply chain records and documents containing unstructured textual data could be stored and managed in the global supply chains. Many organizations could have the motivation to use standards for the data documentation, such as OLB. Having the data standardized would tackle many challenges and meet the requirements discussed previously. (Korpela et al. 2017.) Alternatively, existing research has presented other ways for capturing and managing the data. Kasravi et al.

(2003), and Kruk et al. (2007) introduce a system for managing supplier contracts. These approaches interest our research given the context and approach.
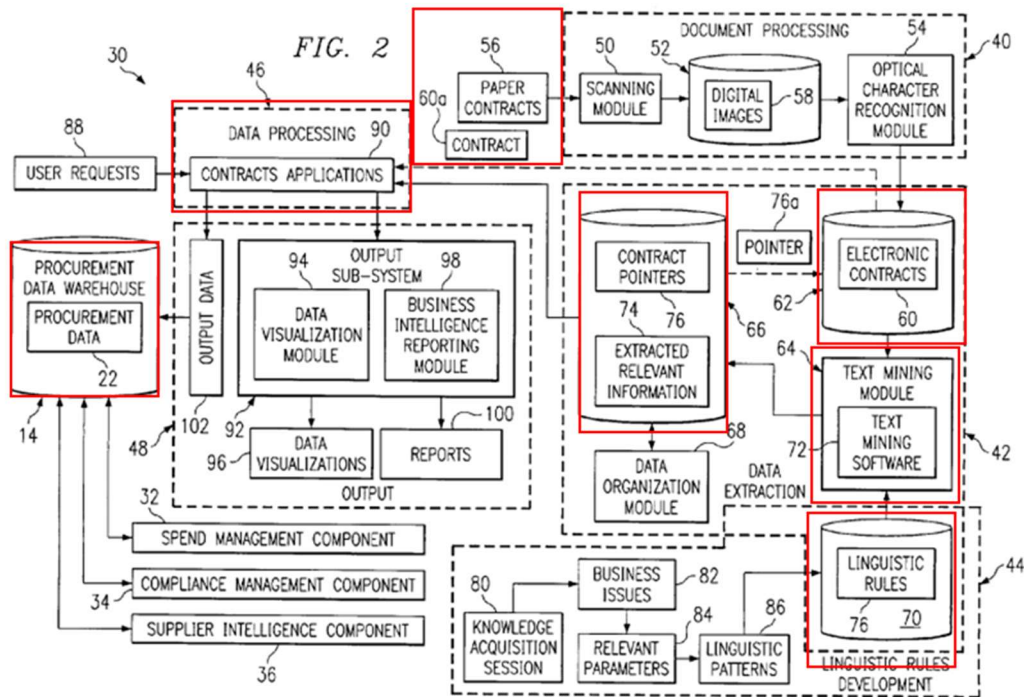
Figure 1 Procurement data management process illustration (Kasravi et al. 2003, Kruk et al. 2007)

In Figure 1, contracts (56), containing unstructured textual data are transformed into electronic contracts (60). After this, electronic contracts are transferred into the data integration sub-module (42). Here, using text mining tools (64), the relevant information is extracted from the contracts. The relevancy is according to the business parameters, which are defined in the linguistic rules database (70). Then this information is linked to other information systems, such as contract applications (46) and procurement data (14) with the contract pointers (76).

A similar type of approach is also proposed in Kassner et al. (2014, 38) for the integration of various distinct data sources into one platform with Figure 2.
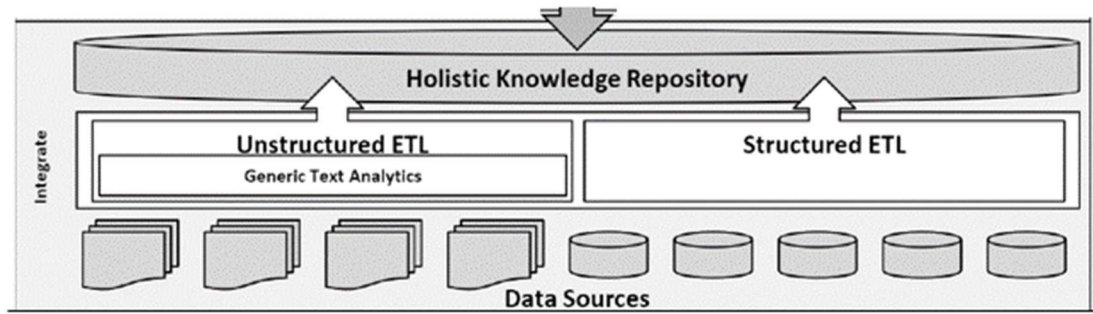
Figure 2 Data integration layer (Kassner et al. 2014, 38)

First, structured and unstructured data are imported into the system as different data sources. Both data types have their separate ETL-processes. The integration of unstructured data to the knowledge repository is done by enriching the data with different types of structured information utilizing generic text analytics processes. These generic text analytics processes can be, for example, part-of-speech-tagging or named-entity recognition. After this process, the whole dataset is stored in the knowledge repository. By implementing this type of process for managing the unstructured textual data, organizations can enable an automatic approach for transforming the contracts to electronic documents, extracting the relevant information with text mining tools based on the linguistic rules set, which can be syntactic and semantic, for large volumes of supplier contracts in a short of period of time. (Kassner et al. 2015, 38.)

2.3 Data challenges

There are numerous challenges organizations face in the management of the data, both structured and unstructured, and in the analytics processes using that data. These are both organizational challenges (Vidgen et al. 2017, 630; Loshin 2013, 30; Kassner et al. 2014; Dhar et al. 2014; Wang et al. 2018, 10; Watson 2014) as well as technical challenges (Inmon & Nesavich 2007; Kassner et al. 2015, 36; Kadadi et al. 2014; Bakshi 2012; Vidgen 2017, 630; Wang et al. 2018, Dhar et al. 2014, Hu et al. 2014).

For organizations to achieve the full value from the data they have available, they must ensure the alignment of data and information strategy and the overall business strategy. There should be a defined way how the data and analytics are used in their decision-making process. These two topics can be considered as key challenges from an organizational perspective. (Vidgen et al. 2017, 630; Loshin 2013, 30; Deloitte 2019.) To ensure organization data strategy is aligned with the overall business, proper data governance is needed with clear definitions and performance goals (Wang et al. 2018, 10). Organiza-

tion's data management and responsibilities can be scattered and siloed across the different functions and departments and not shared. Data is collected, processed, and stored solely for compliance purposes but not utilized, for example, in analytics to provide valuable insights. (Kassner et al. 2015, 37; Dhar et al. 2014; Wang et al. 2018, 10.) When businesses cannot use the vast amount of data available for value, it can be referred as dark data (Gartner). This will result to increased costs and risks throughout the data lifecycle and an inefficient decision-making process. (Kassner et al. 2014, 36; Wang et al. 2018, 10.) Sharing the information within the organization is a classic problem, and digitalization adds complexity to it, which needs to be addressed (Carlsson 2018, 426). Many organizations, especially in the procurement domain, still lack knowledge of what big data and business analytics mean to their business (Davenport 2006; Watson 2014).

In terms of technical challenges organizations face with unstructured data, the most common issue is the integration of the existing data environments and business analytics processes. (Kassner et al. 2014, 36; Kadadi et al. 2014; Bakshi 2012; Vidgen et al. 2017, 630; Wang et al. 2018, 10; Deloitte 2019, Hu et al. 2014.) Even if the architecture and design would differ between the various data systems, there are common goals: massive scaling-out support, the systems' elasticity, and high availability of the systems (Bakshi 2012). The integration between structured and unstructured data is a challenging task for organizations. The linkage between the structured and unstructured data needs to be defined. It might be costly and challenging for organizations, given the context-sensitivity, uncertainty, and complexity of unstructured data. This tends to lead to cost- and resource-intensive case-approach implementation and management of data environments. And not to a general, holistic and heterogeneous approach for integrating the various data environments and sources for valuable information in the analytics platform. (Kassner et al. 2014, 37, Deloitte 2019.) Table 2 lists the challenges organization are due to face with some additions to previously discussed. The table has the nature of the challenge, its definition, reference to existing literature, and example description of the challenge.

Nature	Challenge	Research	Example description
Organizational	Alignment with business strategy and clear understanding of the value proposal	Vidgen et al. 2017, 630 – Loshin 2013, 30 – Kassner et al. 2015, 37 – Wang et al. 2018, 10 – Deloitte 2019	Procurement, IT, and finance are not aligned the data and business strategy. There are different vision, goals, processes, and measures in use
Organizational	Defined data–driven decision–making process	Vidgen et al. 2017, 630 – Loshin 2013, 30 – Wang et al. 2018, 10	Decisions are base on feelings rather than data and analysis, e.g. in supplier selections. There are no defined processes for the data.
Organizational	Distribution of information	Carlsson 2018, 426, 436 – Kassner et al. 2015, 37 – Dhar et al. 2014 – Wang et al. 2018, 10	Data management processes and the responsibilities are scattered and siloed across the organization resulting to inefficiency, risks and costs
Techincal	Integration of the data	Kassner et al. 2015, 37 – Kadadi et al. 2014 – Bakoshi 2012 – Vidgen et al. 2017, 630 – Wang et al. 2018, 10 – Deloitte 2019 – Hu et al. 2014, 653	Nature of unstructured data: context–sensitivity, uncertainty, and complexity present the challenge
Techincal	Interoperability	Kadadi et al. 2014 – Dhar et al. 2014	Current information systems are not necessarily designed to support unstructured data
Techincal	Variety and volume of the data	Inmon & Nesavich 2007 – Dhar et al. 2014 – Hu et al. 2014, 653	The massive volume and variety of textual data makes it hard to manage and exploit the value
Technical	Accessibility of the data	Inmon & Nesavich 2007	Unstructured data is stored in different formats thus presenting challenge on how organization receive the value from all the available data
Techincal	Cost– and resource–intensity	Inmon & Nesavich 2007 – Kassner et al. 2015, 37 – Dhar et al. 2014 – Deloitte 2019	Traditional environments are not designed to manage unstructured data resulting to high operating costs
Techincal	Searchability and terminology of the data	Inmon & Nesavich 2007	Users calling the same thing different thus disabling the direct search within the data
Techincal	Security and privacy of the data	Dhar et al. 2014 – Vidgen et al. 2017, 630 – Inmon & Nesavich 2007	Compliance issues, unintentional data loss, exposure to data leakages
Techincal	Scalability, development, and maintenance of data environments	Dhar et al. 2014 – Inmon & Nesavich 2007 – Deloitte 2019	IT infrastructure not capable to scale, maintain or develop data ecosystem with reasonable costs due to unstructured data

Table 2 Data management challenges

From Table 2, we learn that organizations face challenges of different nature. We focus on analyzing the organizational and technical challenges in this research. This suits our research questions of answering what type of data and how the data should be embedded into the analytics processes. Brandon-Jones & Kauppi (2018, 36) presents interesting research on human nature challenges organizations might face with user acceptance. Our research will not include this subject into scope. However, the theme is touched when discussing organizational challenges related to data-driven decision-making.

2.4 Data maturity and processes

Maturity models are great tools for addressing the current issues, capabilities, and requirements of an organization in a particular domain. By assessing the maturity level, the organization receives an understanding of the current situation and what is needed to improve the capabilities and performance for the future. (Becker et al. 2009.) This is beneficial for our research to assess the capability to manage and utilize the unstructured textual data in procurement.

To analyze the data management capability and maturity, organizations can utilize the classic framework of Capability Maturity Model (Figure 3) for software development organization by Paulk (1993).
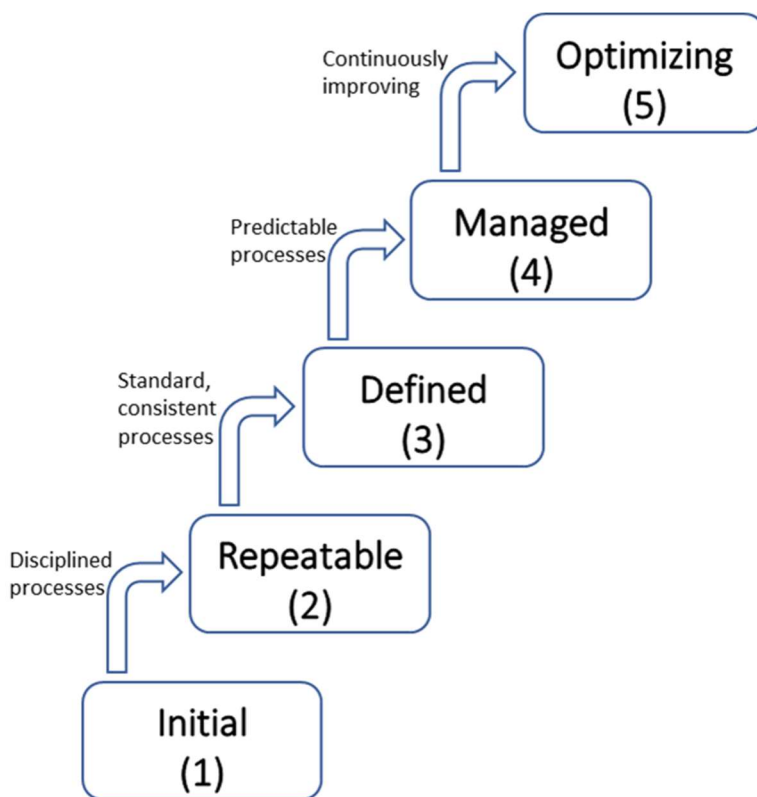
Figure 3 Data management capability maturity level (Paulk 1993, p. 8)

- Level 1: Initial, the data management processes are ad hoc, and only a few are defined. The success of projects depends on individual efforts.
- Level 2: Repeatable, the data management processes are established. Planning and managing data management processes and projects is based on the experience of similar projects. Object to achieve Level 2 is to institutionalize the data management processes.
- Level 3: Defined, the data management processes are documented, standardized, and integrated into a universal organization data management process.

- Level 4: Quantitatively managed, the data management processes are predictable since they are measured and controlled within the measured limits.
- Level 5: Optimizing, the data management processes are continuously improved with feedback and piloting of new innovative ideas, processes, and technologies.

Table 3 below presents data management processes organizations could have as developed and mature data organizations (Ball 2012; Crowston & Qin 2011).

Data life cycle and key process area	Goal	Proceses
Data acquisition, processing, and quality assurance	Reliably capture and describe scientific data in a way that facilitates preservation and reuse	Capture, process and prepare data for storage analysis and distribution with auditable quality
Data description and representation	Create quality metadata for data discovery, preservation, and provenance functions.	Develop and apply metadata specifications and schemas
		Contextualize, describe and document data
		Document data, software, sensors and mission
		Create descriptive and semantic metadata for datasets
		Design mechanisms to link datasets with publications
		Ensure interoperability with data and metadata standards
		Ensure compliance to standards
Data dissemination	Design and implement interfaces for users to obtain and interact with data	Identify and manage data products
		Encourage sharing
		Distribute data
		Provide access (e.g., by creating and piloting service models)
Repository services/preservation	Preserve collected data for long-term use	Store, backup and secure data
		Manage schedules for archive generation, validation and delivery
		Curate data
		Perform data migration
		Build digital preservation network
		Validate data archives
		Package and deliver data archives

Table 3 Data life cycle and processes (Crowston & Qin 2011)

The information and processes in Table 3 can be used as a basis when defining the different maturity levels of organization data management. The maturity of organization

data management can be described with the five levels in Figure 3, ranging from the ad-hoc type of data management to well-planned and -managed processes.

3 PROCUREMENT ANALYTICS STRATEGY

In this chapter, we will discuss the key elements and factors in organizations' data strategy and data infrastructure. Then, we will discuss on organizations' business analytics and procurement processes.

3.1 Developing the data strategy

Organizations have data available more than ever. Still, studies show that organizations utilize less than half of the available structured data on average and less than 1% of the available unstructured data in their decision-making processes. (DalleMule & Dav-enport 2017.) Studies also show that organizations characterizing themselves as data-driven, i.e., having strategy for data utilization and analytics in place, have better performance in financial and operational measures. Organizations which were in the top third of their industry, in terms of data-driven decision-making, were 6% more profitable and 5% more productive than their peers. (McAfee et al. 2012.) The requirement of having a defined data strategy is even more critical, going into the future with the presence of data coming even more substantial (Marr 2017, Deloitte 2019).

Without a defined data strategy, organizations are limited to capture the full value of the data and are liable for poor business results. Still, studies show organizations are lacking in the data strategy (Baldwin 2018). Two essential questions exist for organizations to answer when defining the data strategy relating to new technology, such as utilizing unstructured textual data: What are the processes for determining the strategy's feasibility and business value? What must happen in an organization to implement the new technology into its data environment? (Loshin 2013, 21). When developing the data strategy, we must define the potential benefits and challenges as in any decision-making process. When assessing the potential value and benefits of implementing a new data strategy for unstructured data, or any new technology, we can analyze the potential value in terms of increased profitability, reduced organizational risk, and enhanced customer experience. These value propositions should be compared to the costs and challenges which arise when introducing the new technology. (Loshin 2013, 29.) The management of unstructured data has been researched to be one of the significant challenges in organization data management and business intelligence for a long time now (Blumberg 2003).

Often organizations still rush into implementing new technologies without a proper process for defining the strategy. Even if organizations would understand the potential value of the new disruptive technology, it may be that the strategy for utilizing it is not

aligned with the organization's overall strategy, or we might encounter unexpected challenges during the process. Therefore organizations are unable to unlock the potential value of the new technology. (Loshin 2013, 30.)

A thorough strategic plan is needed to enable the utilization of new technology, such as unstructured textual data, in any organization's business analytics domain. The strategic plan should balance the new agile, innovative processes with the continued operations in the existing environment. It should align the adoption of new technology with the organization's vision and governance and consider the viability and feasibility of the new technology. (Loshin 2013, 30.)

Listed are the key topics that should be assesses when developing the data strategy for new disruptive technology, with description in the case of implementing unstructured data in organization data strategy (Loshin 2013, 31-36):

- Ensure there are processes to involve all interested stakeholders
  - Regardless of who is running the projects, IT, or business, it is essential to involve business early on in the process. With the early involvement of the business in strategy planning, organizations can consider the expected value of the new strategy implementation, manage those expectations throughout the process, and link the business value to business utilization of unstructured data.
- Clear specification of evaluation criteria for acceptability and adoption
  - It is essential to precisely define a set of measures to determine if the strategy implementation is successful or not, i.e., success criteria of the project and utilization of unstructured data. This reduces risks to commit unnecessary projects, avoids killing projects without defined end-result, and improves internal decision-making and project management with clear goals and action plans. Business users can evaluate the project's result, and for the technologist, it is an audit and proof that provided benefit leads to business value.
- Prepare the data environment for massive scalability
  - The characteristic of unstructured data differs from structured data. Therefore, it is critical to assess the combability and feasibility of the implementation on a large scale. Small-size tests might hide performance issues in the existing environment. Organizations must adjust the current data environment model when implementing new technology.
- Promote data reusage
  - Organizations must enable value creation through the collection, integration, and analysis of many large disparate datasets. Data should be integrated and shared across the organization with trust.
- Establish data governance and oversight policies

- o When transitioning for mainstream processes in organization data strategy, it is critical to institute governance and oversight policies. Without these, the newly integrated data cannot be monitored and controlled accordingly to corporate policies. This limits the potential and usability of the data across the organization. When incorporating the governance into the data strategy, the organization receives visibility and awareness of the potential value of unstructured data.
- Provide a governed process for mainstreaming technology
    - o When the organization has conducted pilot projects where the potential business value has been proved and success criteria have been met, it is vital to define a technology adoption plan. This plan will guide the organization from an experimental phase to a production design and development phase. Organizations should define needed resources and requirements, such as staffing and technical requirements.

McAfee et al. (2012) identify five areas in organizations which should be covered when defining data strategy and transforming to a data-driven organization: leadership, talent management, technology, decision-making, and company culture. We learn that defining the data strategy is not only a technical matter, but organizations must also be able to manage the change that comes with the adoption of the new technology. Not only is enabling data enough for organizations to be successful, but leadership must also be capable to define proper goals, define the vision of the organization, how success should be achieved, and ask the right questions. It is not though only the leaders that are needed, as the data becomes more available for organizations, it is becoming more important how organizations' internal resources, the people, are utilizing it. People must have the technical expertise to handle the data and provide answers to business challenges. Of course, for this process, new types of tools and technologies are needed to handle the volume, velocity, and variety of new kinds of data existing in organizations, though it is not sufficient solely. In effective organizations, information is shared across the organization, and decision-making utilizes all the organization's available information. A shift in organizational culture is needed, from feeling-based decision-making to data-driven decision-making. This can be more challenging than expected and should be given attention.

In the annual Deloitte's CPO study (2019), three key guidelines were composed for organizations to successfully manage their data in a global digitalized world. Organizations must have courage in their vision but base the plans on a solid foundation, which means that the basic steps need to be defined before accessing the full value of new technologies. The procurement's digital strategy must also be aligned with the overall strategy of the organization. This was highlighted previously already, as was the last guide: the data strategy and implementation of new technologies must be supported with strong and

clear leadership acknowledging the industry's potentials and executing the implemented strategy with discipline.

From the above, key requirements for successfully implementing the data strategy for adoption of new technology can be consolidated into alignment of business strategy with data strategy, involvement of the whole organization, governance, and support of the mature data processes, and the clear vision and goals of the data strategy (McAfee et al. 2012; Deloitte 2019; Loshin 2013, 31-36).

3.2   Business analytics infrastructure

The data infrastructure should fulfill certain requirements. The infrastructure must cover the whole lifecycle of data and the data management processes. It must be heterogeneous and flexible to include various data sources and analytics tools. It must integrate the structured and unstructured data into a common system. It must be comprehensive data management and analytics framework. Lastly, it must offer the possibility to analyze both structured and unstructured data in an efficient matter in one platform and provide additional value to the organization. (Kassner 2014, 37.) The infrastructure should also include clear definitions for the data access, usage, transfer, and reproduction rights. These rights are vital to include into the infrastructure design, since it is plausible the infrastructure includes multiple sources for the data and tools to manage the data. In this type of network ecosystem these rights need to be aligned and promoted across the parties to enable the full value of the data to the parties which have the rights to access the data. (Munoz-Arcentales et al. 2019.)

We present end-to-end data infrastructure, following the frameworks by Baars & Kemper (2008), Land et al. (2009), and Kassner et al. (2014):

| Analytics layer | Value added analytics | | |
| | Applications | Applications | Applications |

| Data Management layer | Common data repository | |
| | Text analytics, meta-info, business relevancy | Structured database |

| Integration layer | Unstructured data ETL | Structured data ETL |
| | Data sources | |
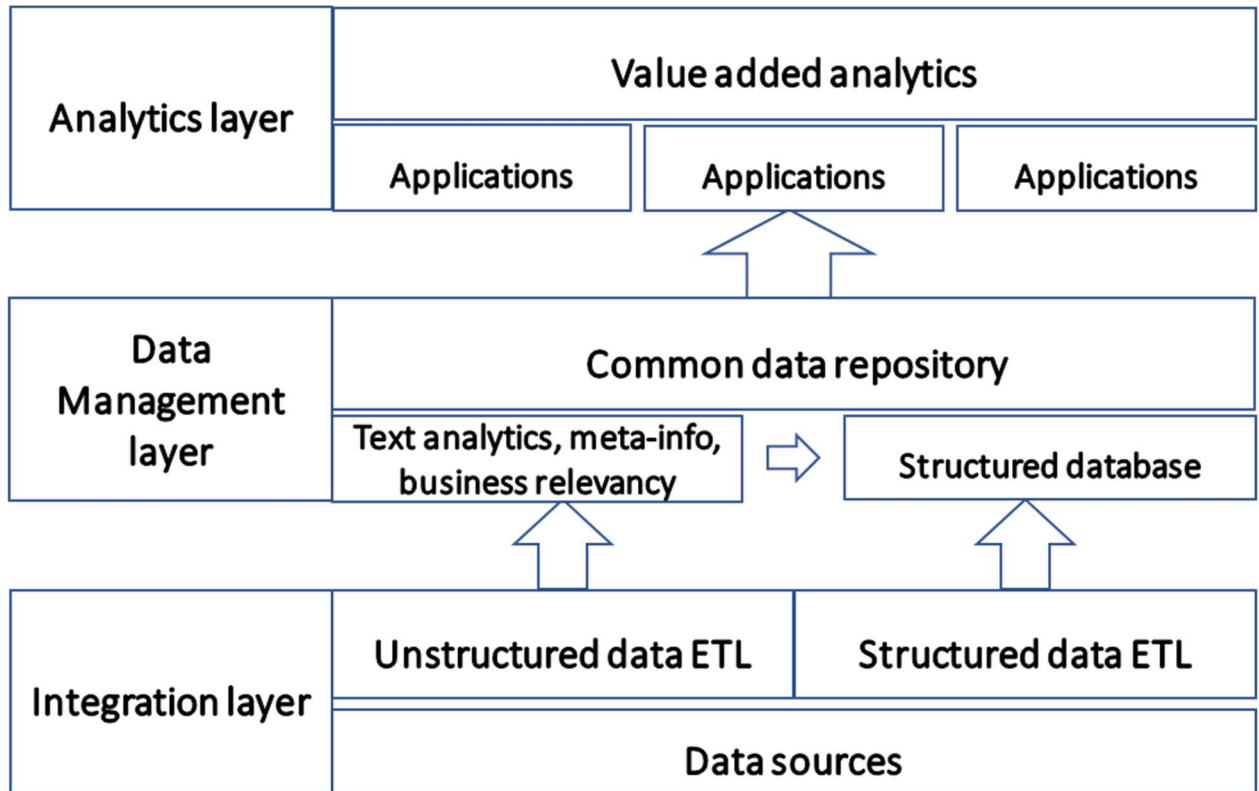
Figure 4 End-to-end data infrastructure

The first layer is the integration layer. Structured data should follow the standard ETL-process. The important thing is the extraction and integration of the unstructured data. We have discussed the challenges previously. Proposed solution follows Kassner (2014, 38) and Kasravi et al. (2003). The unstructured data should be processed with defined unstructured data ETL-process, such as text analytics processes, to obtain the relevant information. This could be, for example, product groups from the supplier contracts with defined conditions or contract owners with expiry dates. These processes provide us structured information, which we can use to integrate the unstructured data into the structured environment.

The second layer is the data management layer. In the layer, data is stored in a structured way from both source environments of data, structured and unstructured. This allows us to manage, manipulate, store, and share the data in our organization. In this layer, all the information is stored in the common data repository, which acts as a source for the next layer, the analytical layer.

The third layer is the analytical layer. The data from the common data repository is fed into the analytics platform, where it can be used for analytics. With the processes before data is fed into the analytics layer, the unstructured data can be combined with the structured data in a common analytics layer for value-adding analytics.

Baars & Kemper (2008), Land et al. (2009), and Kassner et al. (2014) frameworks follow the same principles as the framework we have now demonstrated. In the Baars &

Kemper (2008), and Land et al. (2009) frameworks, the unstructured data is processed with a similar ETL-process as the structured data. For example, Baars & Kemper (2009) propose to search keywords from the already extracted, transformed, and loaded unstructured data content to spot errors in the automotive records. So, all though we fulfill the key requirement of having the unstructured data searchable, we must still know what analytics we ought to perform to the unstructured data with this approach. Thus, we might limit our potential value since some data might be left out of analysis with this non-optimized and fixed approach. Therefore, it is important to have a unique ETL-process for the unstructured data to ensure all information is captured and make the framework heterogeneous and adaptable for various data sources and cases. (Kassner et al. 2014.)

3.3 Business analytics

In the above chapters, we have discussed how organizations should ensure the unstructured data's availability and quality. However, even though the organization would have proper processes to acquire the data, integrate it, and process it to its availability, it might be that the expected insight and value is not received; something is missing. Once the technical layer and data are in place for analytics, the value would need to be captured. Efficient business analytics continues, from capturing and storing the data to accessing and using it and creating value. For this, significant domain competencies are needed, as well as organizational level enablement. (Mikalef et al. 2020, 2,12)

The underlying idea of turning data into information, and thus with efficient decision-making, into a value is a well-researched area in IS and in practice. In the 1950s general intelligence was introduced, in the 1990s business intelligence, and in the 2000s business analytics, and later big data analytics. So even though the terms of big data, unstructured data, advanced analytics, text analytics, and big data analytics might be relatively new, the idea has remained the same. (Sharma et al. 2014, 434; Chen et al. 2012, 1166.) Already in 1985, Zuboff presented the role of IT to be also informative. Information technology and systems constantly generated information on the processes of an organization. This information would be used to gain a different, comprehensive, and insightful understanding of the business, resulting in improvement and innovation, strengthening the organization's competitive situation. Recent definitions for business analytics emphasize the extensive usage of analytical methods to turn all data available in all situations into insightful information, to be used in data-driven decision-making and actions (Davenport & Harris 2017, 7; Inform 2020). Analytics can be divided into three distinctive types descriptive, predictive, or prescriptive. Descriptive analytics gives information from the past, predictive analytics gives insights on what is to come in the future, and prescriptive analytics helps with the decision-making providing actionable insights. (Inform 2020.)

Competitive analytics organizations are designed to find a competitive edge from business analytics. They search the value from all available data sources, the structured internal purchasing data, for example, in spend analysis, and then the unstructured textual data from supplier contracts for value-adding analysis. In these organizations, analytics functions have the best possible processes and tools to make the best decisions. Also, sufficient resources are allocated to analytics processes, and the organization culture across embrace the analytics processes, starting from the management rooting to the individual worker level. (Davenport 2006.)

3.4 Procurement in organization

Procurement includes all the activities required from the organization to deliver the goods or services to the final destination. Procurement aims to achieve this with the best possible terms and conditions for the organization. It guides purchasing function in the strategical and tactical sourcing decisions. Procurement covers the organization's whole supply chain and substantially affects its performance, both in qualitative and quantitative ways. (Van Weele 2009; Mena et al. 2014.)

Van Weele (2009, 33) presents the extended purchasing process model with different procurement activities. This model was used as a base by van Raaij (2016, 14), categorizing these activities into strategic, tactical, and operational activities. Strategic activities include spend & demand analysis, supply market analysis, sourcing strategy development. Tactical activities include tendering, supplier selection, and contracting & implementation. Operational activities include searching for specific products, preparing the purchase orders, submitting them to suppliers, securing the delivery, checking the invoices, and paying the suppliers with the negotiated terms. The strategic and tactical activities, in the outer ring of the Figure 5, result into sourcing strategies and contracts with the suppliers. This contract is then carry-out to invoices, which are used as an input to spend analysis. The inner ring specifies which supplier management activity is ongoing which are strategic.
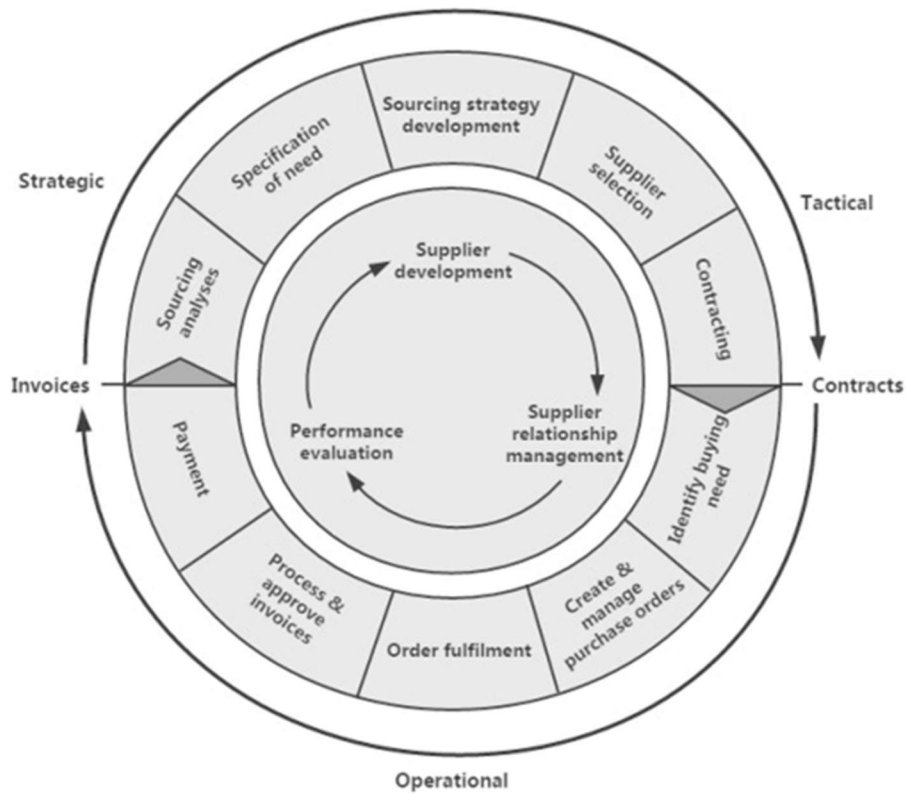
Figure 5 Procurement process (van Raaij 2016 adopted from van Weele 2009)

In our research, we will scope the procurement activities to strategic and tactical activities and analyze how unstructured data may affect those activities when used in procurement analytics. We can see that the contracts are, for the most part, managed in these activities. When discussing procurement, the Kraljic matrix is worth involving. The Kraljic matrix helps purchasing and procurement function to manage and optimize their purchasing processes more efficiently and with better results. The purpose of the matrix is to classify purchased materials and related suppliers into four distinct categories, based on their financial impact on the company and their risk to the organization and its processes, ranging from low to high (Kraljic 1983).

3.5 Procurement analytics

As discussed before, organizations using data analytics in their decision-making outperform the competitors (McAfee et al. 2012). Procurement is in a key position for organizations to outperform with data analytics due to its impact on profitability. In the latest Deloitte's (2019) CPO survey, analytics tools, such as spend and contract analytics, were cited for having the most significant impact from all the digital procurement applications. Organizations' performance is highly dependent on procurement performance. Data-

driven decision-making processes and analytics are needed for improved performance. Decisions are complex processes based on a massive amount of data coming in from various sources and forms. Procurement decisions affect the organization's bottom line, general operationality, and the availability of goods and services, which is the core of business in most cases. (Tan et al. 2015; Kassner et al 2014; Kasravi et al. 2003; Chae & Olson 2013, 10; Handfield et al. 2019; Wang et al. 2016, Batenburg & Versendaal 2008).

Traditionally, procurement analytics has been focusing on internal data and descriptive analytics methods. The focus has been on spend management, cost reduction and analysis, and supplier management domain. (Monczka et al. 2016; Van Weele 2009). Insights and information generated from procurement analytics are critical for procurement activities and generate value in many forms. In addition to improved visibility and control of organizations' spend and supplier base, analytics provide information on procurement patterns and enable more efficient sourcing processes for key stakeholders. (Tan et al. 2015).

The new frontier of unstructured data and advanced analytics provides new value to procurement. This is estimated to be one of the most disruptive forces in the field. (Handfield et al. 2019.) However, procurement needs to possess the technical capability to manage all the various data and organizational capability to make the data-driven decisions efficiently, to achieve the full value. (Handfield et al. 2019; Ransbotham et al. 2016). These capabilities complement each other (Chae & Olson 2013, 20).

Contracts are critical for procurement. They define the terms and conditions for suppliers to deliver the product and service. Notably, contracts can have various formats, are typically scattered across organizations, and the amount of contracts can be substantial in large organizations. Current contract analytics applications focus on efficiency and compliance analysis. For example, analyzing does the supplier have a contract in place and when this contract will expire. But where the organizations could find the most value is the unstructured textual data found within those contracts. (Blumberg 2003; Inmon & Nesavich 2007; Negash et al. 2008; Kassner et al. 2014; Bakshi 2012.) When this data from the contracts is extracted and analyzed properly, it can provide valuable information on their procurement decisions. (Von der Gracht et al. 2016; Kasravi et al. 2003.) These benefits are related to cost savings using agreed discounts, increasing the margins with rebates and refunds, and controlling the risks by ensuring compliance requirements with their suppliers. Organizations can also find opportunities when comparing the agreed terms of purchasing to actuals, for example, agreed payment terms to actual payment terms on paid invoices. The sharing of relevant information from the contracts across the organization can also be seen as a key advantage. The same information could be used in different business decisions in different places. (Handfield et al. 2019; Kasravi et al. 2003, Denton et al. 2013.)

3.6   Procurement maturity

Rozemeijer et al. (2003) describe purchasing maturity as "the level of professionalism in the purchasing function." The model helps assess the organization's procurement function role, availability of information systems, quality of processes and people involved, and integration with external parties and knowledge. Many development models of procurement maturity have been presented in the research. These maturity models present the stages organization are due to go through to achieve the highest level of professionalism. (Rozemeijer et al. 2003; Schiele 2007). Rozemeijer et al. (2003) present a six-stage model for purchasing development: Transactional orientation; Commercial orientation; Purchasing co-ordination; Internal integration; External integration; Value chain integration.
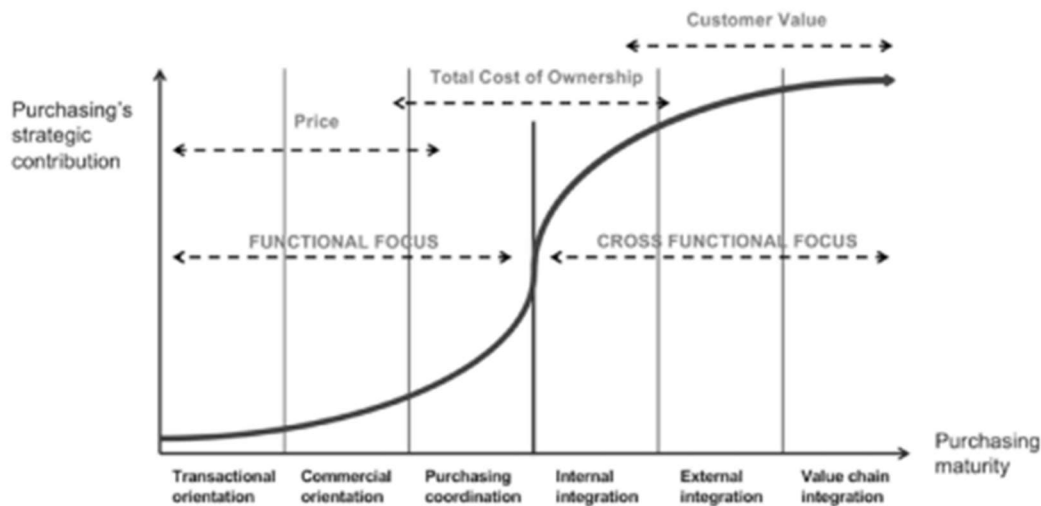
Figure 6 Procurement maturity development by Rozemeijer et al. 2003

In the first stage, procurement's primary task is to ensure the availability of the goods and services. No defined purchasing strategies are in place, and the purchasing is operational and administrative. The performance of procurement is measured by the availability of the right materials and goods for internal stakeholders.

In the second stage, procurement has a defined purchasing strategy in place to source and negotiate the goods and services with the lowest possible price from suppliers. Performance is monitored primarily based on the price and delivery performance of the suppliers. The effectiveness of procurement is measured through cost savings.

In the third stage, procurement is adapting to uniform the purchasing policies and systems across the organization. Procurement has developed cross-organizational sourcing strategies, aiming for internal coordination and synergy. The performance is measured, in addition to price and cost, by the quality level of the products and service. Also, non-

production-related purchasing is important for the procurement organization. Supplier management is a key topic at this stage in the organization, and supplier strategies are developed, for example, portfolio analysis. Information systems are also in place but still not fully integrated across the organization.

In the fourth step, procurement has become strategically important in the organization. It has an influence on organization strategical issues, such as core / non-core activity questions and make-or-buy decisions. Its objective is to reduce the total cost of the systems, not just the unit costs of the products. Procurement is integrated cross-functionally across the organization with process orientation. Procurement performance is measured with internal customer satisfaction surveys and benchmarking. Information systems are integrated across the organization, but not yet with external parties.

In the fifth step, procurement has defined outsourcing strategies in place. Collaboration with suppliers is emphasized. The total effects of the supply chain resources of the organization are analyzed. Procurement has a key role in non-production purchasing. It has established processes and systems for the organization to source goods and services with the best efficiency possible, for example, EDI-systems and system contracting. Procurement invests heavily in supplier relationships, including integration of information systems to external sources and parties. The responsibility of purchasing is cross-functional. Management is result-driven, but continuous learning and participation in decision-making are emphasized.

In the last step, the procurement strategy is based on the value provided to the end-customer, which is also the performance criteria. The supply chain is challenged to perform with the best possible efficiency, and procurement actively participates in the processes of the whole value chain. The procurement strategy is aligned with the business strategy, and the visions of all functions are integrated together as one. (Rozemeijer et al. 2003.)

Existing research presents that there is a positive relationship between the high-maturity level of procurement and organization performance. With a high maturity level, procurement organizations are more successful in implementing new external knowledge sources, for example, sourcing strategies in this research context. Organizations with low-maturity levels might fail in innovative performance improvement processes, even though one could assume low-maturity organizations would have more savings and performance improvement potential in their procurement processes. (Schiele 2007.) Both Schiele (2007) and Rozemeijer et al. (2003) find that the more mature the procurement function, the more capability it has to perform, especially related to new and advanced technologies and methods.

3.7   Procurement performance

Once organizations have defined the processes, comes the time to assess the performance. Previously we discussed increased profitability, reduced organizational risk, and enhanced customer experience to be key value propositions for organizations when introducing new technology or process (Loshin 2013, 29). In the procurement domain, Handfield et al. (2019) research identifies reduced costs and increased quality in the procurement core processes, such as spend management and sourcing activities, as the key benefits of emerging technologies in procurement analytics. Example benefits are efficiency improvement in processes when time can be switched to value-adding activities, enhanced contract visibility, reduced working capital, and better management of supply chain disruption with the new technologies in procurement analytics. Deloitte's (2019) CPO survey highlights cost reduction still to be primary goal for procurement, but adds risk reduction as other major driver in procurement.

To measure the performance of a procurement organization, we conclude that cost savings is a key indicator. It is tangible and can be calculated from the organization's financial statements. Another critical indicator is risk reduction as the emerging theme in the domain. In addition to these two, overall quality improvement related to organization processes and systems can be held as an important indicator. We are also motivated to add overall value provided to all interested stakeholders as one dimension based on the literature of Schiele (2007), Rozemeijer et al. (2003), Loshin (2013), which all highlighted the overall value as one key indicator for measuring the performance of an organization with new and advanced technologies. We will also recommend including the Kraljix (1983) matrix assessment as one part of the performance evaluation. This will help the organizations focus on the most important activities that possess the most prominent value proposal.

Cost

• Tangible savings

Risk

• Reduction and management

Quality

• Overall quality improvement related to processes and systems of organization

Value

• Overall value promise to all interested stakeholders
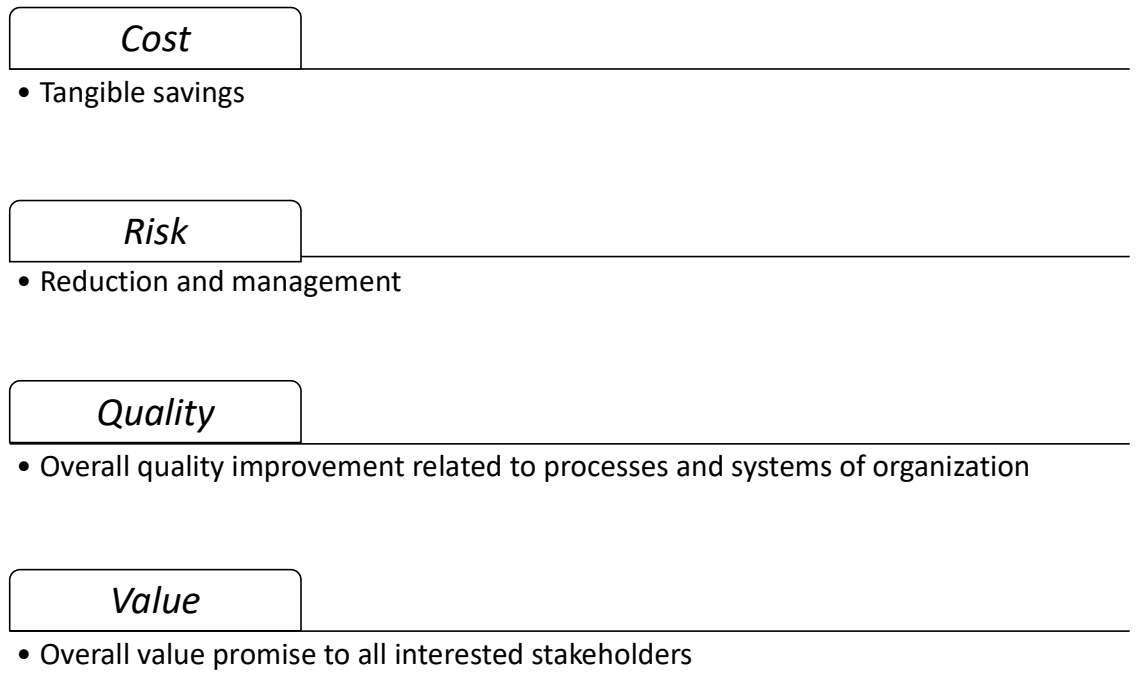
Figure 7 Performance indicators

These performance indicators in Figure 7 will be used as a base in the case study interviews. Part of our research was to study why procurement organizations should start utilizing unstructured textual data. Furthermore, to study the motivation, we want to present a framework for assessing the potential value proposals. These indicators serve us in those discussions when constructing the view on why the unstructured data should be utilized.

4   RESOURCE-BASED-VIEW ON ORGANIZATION PRO-CUREMENT ANALYTICS

In this chapter, we will present a resource-based-view on organization capability. The resource-based-view will be used as base in our theoretical framework for assessing the capability and requirements organizations should have when approaching the unstructured textual data in their procurement processes.

4.1   Resource-based-view theory

The resource-based-view is practical for analyzing the capabilities and influences of each resource organizations have (Wade & Hulland 2004, 131). Resource-based-view also allows us to assess the capabilities from a cross-organizational perspective, which is critical to ensure alignment across the organization. The literature proposes the competitiveness derives from the combination of various resources which are valuable, rare, inimitable, or non-substitutable. These are important features of each resource, but we will focus on our research solely on the actual capabilities related to the resource and not include the consideration of the resource attribute. (Barney 1991.) The theory suggests that all the resources must collaborate to receive the full value (Vidgen et al. 2017, 25.) Resource-based-view has been applied in information system research as well (Peppard & Ward 2004, 174). Organizations' resources in procurement analytics can be divided into three sub-categories. First are the tangible resources: the data, the technology infrastructure, and the basic resources, time and money, within organizations. The second resources are human skills: how equipped people are with technical, managerial, and professional skills. Lastly, organizations have intangible resources: data-driven culture, organizational learning intensity, and actionability. (Gupta & George 2016, 1062; Mikalef et al. 2020, 12; Vidgen et al. 2017, 630; Janssen et al. 2017, 344.)

Organization resource classification		
Tangible	**Human**	**Intangible**
• Data • Technology • Basic resource	• Technical skills • Managerial skills • Procurement skills	• Data-driven culture • Organization learning intensity • Actionability

Figure 8 Organization resource classification following Gupta & George (2016)

Next, we will analyze how organizations could approach and assess their capability towards unstructured data in their procurement analytics using the presented resource-based-view. First, we will discuss the tangible resources and capabilities related to them. We use the previous discussion from chapters 2.1-2.4 and 3.1-3.2. Then we will discuss the intangible resources and capabilities related to them. We use the previous discussion on the technical and managerial skills from chapters 3.1-3.3, and on procurement skills, we use chapters 3.4-3.6. For intangible resources, which relate to the overall data strategy and capability of the organizations, we use chapter 3.1.

4.2 Tangible resources

Tangible measures are defined as resources related to data, technology, and basic resources of an organization. Our end-to-end data infrastructure presents a foundation for the procurement data environment utilizing structured and unstructured data, see Figure 4. The presented infrastructure should solve the technical challenges discussed, see Table 2. Prior IT research also presents that the available resources for organization information system should be allocated across the organization, and not only to be managed by the organization's information system or technology department. Having processes and alignment on the available information system resource across the organization is required. (Peppard & Ward 2004, 176.)

4.3 Human resources

Procurement analytics requires competencies from the organization in business analytics and procurement context. Organizations need to emphasize the importance of analytics in their strategy. Domain skills in procurement are needed for full value and improved performance. (Davenport 2006; Chen et al. 2012, 1667.)

Competitive business analytics enables us to share the information needed for decision-making across the organization (Davenport & Harris 2017, 7). According to Deloitte's global CPO survey (2019), digital skills development is key for procurement organizations in the future. Especially, developing the skills in data visualizations and predictive analytics are highlighted in the survey. The study states that organizations who can become masters in managing the external and internal complexity via digital solution outperform in all procurement aspects the peers (Deloitte 2019). Vidgen et al. (2017, 24) emphasize the growing importance of analytical skills in the field.

Involving all stakeholders into procurement strategy is crucial. The change needs to be implemented across the organization. Parties should be involved early phase and supported from the top of the organization. (Rozemeijer et al. 2003.) Rozemeijer et al. (2003) presented us with the maturity model of organization procurement. With determining the maturity of procurement, we can assess the level of professionalism and capability. The more informative and integrated the procurement organization, the more mature the procurement organization is. Moreover, the maturity would also yield better performance (Schiele 2007). The procurement maturity and its importance were also discussed by Marr (2017).

We can define that it is required for the organization to have the capability to provide the information with business analytics in the organization in a manner it can be utilized in the procurement domain. These two entities, digital capability and procurement capability, are interlinked and are reliant on others. The skills can be classified into technical skills related to the digitalization domain, managerial skills related to implementing the strategy and the support, and procurement skills related to the procurement function's professionalism.

4.4 Intangible resources

Henderson & Clark (1990) presented a case where R&D function could not utilize relevant information, which was available, for their emerging technologies in product development. There was no cross-functional or organizational strategy for individual team members to understand the information's insightfulness to the organization's overall strategy. Thus it was not capitalized for the benefit of the organization.

The tangible and human resources could be acquired from external sources, but the intangible resources, data-driven culture, organizational learning, and actionability are embedded into the organization code. It is important to consider them when developing the organizations' capabilities and assessing the investments. (Mikalef et al. 2020, 12; Davenport 2006.) Data-driven culture has been noted as a critical component for organizations working and utilizing new technologies (LaValle et al. 2011). Furthermore, as these new technologies are emerging with high cycle, organizations must have the capacity to continuously learn when managing and implementing these new technologies across the organization's value chain (Vidgen et al. 2017; Rejeb et al. 2018, 84, Peppard & Ward 2014).

Even if an organization would emphasize a sophisticated analytical process to turn data into insightful information with a data-driven decision-making mindset, it might still lack the act of action from that insight. To solve this gap, organizations need to understand that information is just one component of business analytics, the other compoment is the response. For that, organizations need flexibility, structuring of the organization's capabilities, and removing all barriers of turning insight into action. (Mikalef et al. 2020, 12.)

5   FRAMEWORK OF THE RESEARCH

The main research questions of this research were:

- How should procurement organizations approach unstructured textual data?
- What is the expected value of the unstructured textual data in procurement?

To answer these questions, we present theoretical framework derived from the existing literature. This framework can be used as base for reflecting the findings from our empirical research phase. First, we will review the literature question by question, and lastly present the framework.

5.1   How should procurement organizations approach unstructured textual data?

The first research question was: How should procurement organizations approach unstructured textual data? This question is answered by first defining the unstructured data which interests procurement, what is the source, and how it could be managed. We discuss the factors affecting the availability which are challenges related to the data management, and organizations data management maturity and processes.

For the definition of unstructured data, we concluded that the data format must be defined since, unstructured data exist in many formats. In this research, the unstructured data was defined to be unstructured textual data, see Table 1. The most sources of this data were supplier contracts procurement documents, such as contracts and web data. This definition was driven by the interest the data format and source had in procurement. Then, we discussed the general principles and processes related to management of the unstructured textual data (Blumberg 2003, Inmon & Nesavich 2007; Kassner et al. 2014). We also discussed the challenges organization are due to face with data. Challenges were both organizational and technical, see Table 2. Data maturity of organization described the professionalism in data management, see Figure 3. General data processes were presented in Table 3. The maturity evaluation was done to assess the capability of organization to manage the unstructured textual data. We concluded as the main principle to be that the unstructured textual data should be integrated into structured environment and made available for usage (Blumberg 2003, Inmon & Nesavich 2007), and this requires certain processes to be in place, such as standardization of the documents (Korpela et al. 2017) or sophisticated techniques for integrating the data, see Figure 2 and 3.

After we have defined and discussed the unstructured textual data, we can assess the key factors, capabilities, and requirements which should be considered once approaching

the data. We discussed the importance of data strategy, and key requirements of the strategy. Moving from this to more practical level, assessing the data infrastructure requirements. And lastly, presenting the capability framework from resource-based-view with tangible, human, and intangible resources.

Literature presented the importance of having data strategy in place with the key requirements of alignment between business strategy and data strategy, the involvement of whole interested organization, the governance and support of the mature data processes, and the clear vision and goals of the data strategy (McAfee et al. 2012; Deloitte 2019; Loshin 2013, 31-36; Vidgen et al. 2017, 24). The infrastructure for managing the unstructured data was presented in Figure 4 with the critical requirements. The layers of the data infrastructure are integration layer, data management layer, and analytics layer. (Baars & Kemper 2008; Land et al. 2009; Kassner et al. 2014; Figure 5.) The capability for successful procurement analytics with new technology can be assessed from the resource-based-view. This is practical for analysing the capabilities and influences of each resource of organization have (Wade & Hulland, 2004, 131; Peppard & Ward 2004). Tangible resources were related to the data, technology, and basic resource. Human resources were technical, managerial, and procurement skills. Intangible resources were data-driven culture, organizational learning and actionability.

5.2   What is the expected value of the unstructured textual data in procurement?

The third research question was: What is the expected value of the unstructured textual data in procurement? Here we used following performance indicators: cost, risk, quality and value which we consider to interesting for procurement organization. These indicators help us to identify which type of potential lies with the unstructured textual data, and what kind of impact it would have to procurement. In the empirical research, we will discuss together with the organization what type of value they expect to receive from the utilization of the unstructured textual data. These findings are reflected to the mentioned indicators of cost, risk, quality, and value.

5.3   Research framework

RQ1: How should procurement organizations approach unstructured textual data?

Definition of the data

Unstructured data
- Definition and format
- Source and capture
- Availability and management

Data Factors:

- Data maturity and processes
- Data challenges

Approach

Data strategy
- Alignment of business strategy with data strategy
- Involvement of whole interested organization
- Governance and support of the data processes
- Clear vision and goals of the data strategy

Data infrastructure requirements
- Key requirements
  - Whole lifecycle of data and mature data processes
  - Heterogeneous integration of the data and design approach
  - Comprehensive value adding platform
- Layers
  - Analytics, Data management, & Integration layer

- Tangible resources
  - Data
  - Technology
  - Basic resource
- Human resources
  - Technical skills
  - Managerial skills
  - Procurement skills

- Intangible resources
  - Data-driven culture
  - Organization learning intensity
  - Actionability

RQ2: What is the expected value of the unstructured textual data in procurement

Indicators

Cost	Risk	Quality	Value

Figure 9 Research theoretical framework

6 EMPRICAL RESEARCH

6.1 Methodology

When defining the research methodology and approach, it is essential to understand the research motives and nature. The main research questions help to answer what type of research is in question. All the methods when answering these questions are part of the research methodology. The methodology is the structure of how all the parts are linked together in the research and how the received input is transferred into a research output. (Eriksson & Kovalainen 2015.)

The research is a qualitative case study. The nature of this research suits qualitative methods. We are studying new phenomena, exploring the ways and possibilities, and assessing the expected results from various perspectives. These all are suited for qualitative research. If we were to select a quantitative approach and study the results with numeric values, we might not fully understand the research's objects. (Berg & Lune 2012.)

The research focuses on studying the unstructured data in the procurement context, providing a literature framework, and studying how this is reflected in selected procurement organizations. For this purpose, a qualitative case study is recommended. It is characteristic for a case study to focus on a specific case, or cases, on which the research approach, question, and analysis base on (Eriksson & Koistinen 2014, 5). It is recommended to choose a case study as a research method when the following criteria is met: Research aims to answer what, how, and why. The researcher has limited control over the research occurrences, there exists only limited research on the subject, and the topic of research is current. (Eriksson & Koistinen 2014.)

It is essential to define the case in question when conducting case research, as the case can be many things. This can be, in some instances, a challenging task, especially defining it precisely before the empirical research phase. (Eriksson & Koistinen 2014.) To answer this, the existing literature has provided a method of "caseing". This means the case is defined throughout the research as a process. (Ragin & Becker 1992; Eriksson & Koistinen 2014.) The particular case should be selected with defined logic, and the expectations for selecting the case should be that we learn something from the particular case. The case should be representative, special, unique, or theoretically interesting. (Stake 1995.) This suits our research since, based on our literature review and the framework, we are able to define the case to be the unstructured textual data in the analytical processes of procurement. However, as we gather the empirical material, the case can be defined into further details of specific unstructured data in specific procurement processes and explore the means and whys related to this case in an explorative way.

Eriksson & Koistinen (2005, Chapter 5, p.5), using existing literature, distinct case studies into two: intensive and extensive case study. An intensive case study focuses on a unique case, and an extensive case study on generalized findings by testing the theoretical framework in multiple distinct locations. Our research falls into the latter, an extensive case study. Stake (1995, 3) classifies the case studies into three distinct types. The study could be an intrinsic, instrumental, or collective case study, depending on the nature, amount, goals, and orientation of the cases. Intrinsic case study tries only understanding one specific phenomenon and does not try to generalize the research from this one case. If it were beneficial and plausible to generalize the research to other cases as well, we would discuss an instrumental case study. As one example from our research context would be studying certain technology implications in a certain organization. We are also motivated to analyze what would be findings in other organizations as well. Therefore, we consider our case study as instrumental. Having a clear framework available implies us to see this study as instrumental (Eriksson & Koistinen 2014). The research will perform a case study on unstructured data utilization in the selected procurement organizations. Our case study uses these two organizations to provide perspectives on the subject in two different locations. With this approach, we can evaluate our presented theoretical framework and test how the findings are reflected in these case organizations in practice. Empirical research will focus on these two organizations only since they have explored and studied the possibilities of unstructured textual data in their organization and are clients of the software provider Sievo. The nature and topic of the research require case organizations to understand the subject in question in detail to study the possible approaches, capabilities, requirements, and value. Our research is also focusing on studying how Sievo could help its client approach the unstructured textual data, limiting the case organizations to clients of Sievo.

The case study has received criticism as a research method for the incompetence to create generalization from it. We acknowledge this to be true. It is not plausible to create statistical suggestions from only a few cases. (Eriksson & Koistinen 2014.) However, theoretical suggestions can be derived from the case study, keeping in mind the primary purpose of creating information and understanding the chosen case. Usually, studies that focus on studying and developing the existing theory try to create a systematic and replicable system for the case research. (Stake 1995, 7-8.)

We will analyze the trustworthiness of this study from four perspectives: credibility, transferability, dependability, and conformability (Eriksson & Kovalainen 2015, Chapter 19). Meaning the research should have strong logic with its' claims, existing literature should complement, the process is transparent, and the findings are understandable.

6.2   Collection of data

The case study focuses on finding practical information on how organizations utilize the unstructured textual data, more precisely, the supplier contracts. The primary method for gathering the information is semi structured interviews with the data and procurement expert and leader from both case organizations. Interview participants were selected with pre-assessment. This ensures we have reliable and precise information on the actual solutions to this relatively new and emerging phenomenon we are studying. Randomly selecting interview participants might result in inaccurate data and non-interesting results (Hirsjärvi et al. 1996). Prior to the interviews, participants were prepared for the interviews by sharing relevant information on the research and the questions to be used in the interview. This helped the participants gather the information from their organizations if they did not have the information themselves. The actual online interviews lasted approximately 60-90 minutes each, with follow-up discussions held offline.

Interview participants work at large global organizations. Both participants were responsible for digital tools in procurement and other procurement activities and development within their organizations, making them an excellent fit for our study, as they have the required knowledge and expertise on the subject from their organization. Both of the organizations where the interview participants work are global, Europe-based, multi-billion companies. The industry nature of these companies is fast-moving goods and distribution. These companies are both clients of Sievo, a Finnish procurement analytics software company, who is also a participant of this research. Both organizations have already introduced analytical processes and tools into their procurement and are actively developing their analytics processes and capability. Persons and organizations are anonymized in the research. The anonymization was chosen as the approach to increase the depth and quality of the answers, and also due to the fact our research is not interested in the actual companies, rather the solutions.

6.3   Data analysis

The interviews were recorded in English and transcribed to clean text. The gathered data is analyzed with qualitative content analysis. The content analysis can be divided into preparation, categorization, and reporting of the data and the results. We will prepare the data and analyze it by reflecting on our research framework and identifying common patterns, cause-effect, and other themes. For this exercise, coding is recommended. The cleaned textual data from both of the interviews is coded into each research framework topic. This is then cross-compared to existing literature and also to the other interview.

(Elo et al. 2014; Eriksson & Kovalainen 2015.) In the analysis, we will classify the two case organizations as organization A (OA), and organization B (OB).

7   RESULTS

In this chapter, we will analyze the results and findings from the case study interviews. The analysis is structured with the research framework. Organization and interviewees are anonymized, and referred only as Organization A (OA), and organization B (OB).

7.1   Unstructured data

7.1.1   Definition and format

Based on the expert interviews, unstructured data is something data that is not controlled in terms of what it is, it comes in many forms, and has lots of potential. It was also said that unstructured data requires manipulation to have it as meaningful information. This was compared as an opposite to structured data which has rows and columns, numeric values, variables, and is quantifiable.

Most interesting data for procurement was seen to be unstructured textual data. Interviewed also mentioned that unstructured data, that could help procurement, is found from images, videos, etc. but the textual had the most prominent value. Hence both organizations were focusing on analyzing unstructured textual data within their procurement function.

7.1.2   Source and capture

Main source for the unstructured textual data was seen to be the various procurement documents. These are contracts, news, reports, and other supplier related documents. And the interesting unstructured data, was the content from these documents. Supplier contracts are important for procurement. They define the terms of engagement with the suppliers.

> *OA: What I see in the legal contract, you define the rules of the game/engagement with the supplier.*

The source and location of the data was known in the organizations. There was some variation in the capturing process. Organization either had process for creating and storing the documents, holding the unstructured textual data, themselves, or the documents were

received from the suppliers. In both approaches, the access and availability to the information was limited. The organization A which created and managed the documents themselves, used standard templates for document creation. But at the end, the documents were customized to meet the personal needs, resulting into incomparability and non-standardized process. These documents were stored in digital solution. In the organization B, there were multiple distinct processes for receiving the data, from digital and automated solution, to less digital and manual solution. This also depended on the format and location of the data, whether it was product descriptions or supplier contracts, wished to be analyzed. It was mentioned that it requires a lot of knowledge and resource to locate, structure, and analyze the data.

> *OB: This [capture] is depending on the type of analysis I need to perform and the expected outcome, so I really take this project by project so if I do analysis I need to the outcome and what are working towards. and for every analysis it is defined the strategy what it takes to get there, and yeah that takes time.*

7.1.3   *Availability and management*

As said both organizations were able to locate the data, but accessing it, making it available, and managing it, was seen as a struggle. After locating the data, organizations pursued to analyze the data by querying relevant information or integrating it into other solutions. It was identified as a requirement to have sophisticated tools for this process, due to various challenges this format and location of data creates, such as the vast amount of information and its variety.

> *OA: If you have only one contract, it's easy to remember. But if it is tens of thousands of contracts, then it's difficult. So, there's a lot of knowledge, rules, constraints, liabilities, and things you are obliged to in these contracts, that you need be aware of. And cause it's in text form, it's not easy to quantify.*

The management relies on individual efforts and legacy tools. When starting the analysis, user would have to know for which purpose the analysis and data is used, and what is the expected outcome. This approach was seen as resource intensive. The expertise and knowledge were in this case siloed and not shared across organization. The tools in use at the moment for analysis were standard support tools, and not state-of-start designed for

the purpose. It was referenced to other companies, with better resource and tooling, how well they were able to analyze and make the unstructured data available. We also founded that even though the tooling might be in place, the results are not in evident. (Organization B)

In the organization A, we did have state-of-art tooling in place, but not the vision and motivation from the users to work with the data, resulting into not-expected results in terms of the availability of the data. These reasons existed both on the user level, and on the organization level. And to top this, we also faced problem of not knowing what to search, even if the tooling and motivation to run analysis would have been in place.

Both organizations recognized the need for structuring the unstructured data. Processes were in place, or were in design phase, for applying meta data for the documents containing the unstructured data, and running text analysis on the unstructured textual data. These processes were performed with own tools, with purpose to later combine the distinct data assets and tools together. Applying meta data to the supplier contracts and documents was seen as a requirement, in order to combine the unstructured textual data found from the contracts, with other supplier related data. This was seen as relevant and valuable.

> *OA: We would want to integrate with other data sources. For example, if you know this contract I have here, that I have analyzed with this supplier. What is written in that contract, you want to connect that, with I know on my supplier [in other sources]*

To increase the relevancy of the unstructured textual data, organizations did have text analysis processes in place. The motivation was to make the data comparable by formatting the unstructured data to more structured, and also to find the relevant information from the contracts, such as obligated parties and terms. The text analysis processes were done in multiple ways. It was reliant on manual efforts using standard help tools, such as Excel. This process was not considered great, as said before. In some instances, the organizations did have sophisticated tools to run the text analysis, using NLP or other AI algorithms. That was considered as a way to go.

> *OA: We are using some AI and NLP tools. We have the contracts and meta data in the repository. Then we take the contracts and dump them into different tool that has NLP capabilities, that allows you to find which part of the text is mentioned start date, end date, contracting parties, governing law, all the relevant things, so it allows to query those contracts.*

7.2 Data factors

7.2.1 Data challenges

For assessing the challenges organizations are due to face, we had the framework from existing literature. In the below table, there are findings from the interviews.

Nature	Challenge	Comment
Organizational	Alignment with business strategy and clear understanding of the value proposal	Seen as one of the largest topics to address in the organizations
Organizational	Defined data–driven decision–making process	Seen as one of the largest challenge with unstructured data, due to unavailability of data and human-driven culture
Organizational	Distribution of information	Challenge, the challenge was the linkage of the data or and the systems
Techincal	Integration of the data	Seen as one of the largest challenge with unstructured data, due to complexity of unstructured data
Techincal	Interoperability	Seen as one of the largest challenge with unstructured data, especially the lack of vision in the strategy
Techincal	Variety and volume of the data	This topic was seen as a matter of perspective
Technical	Accessibility of the data	Challenge to have data available for users to access
Techincal	Cost– and resource–intensity	Seen as one of the largest challenge with unstructured data since tools were not optimized
Techincal	Searchability and terminology of the data	Seen as one of the largest challenge with unstructured data since was hard to unify the data and know what the data actually is
Techincal	Security and privacy of the data	Not an issue, software and service providers have their insurances on this
Techincal	Scalability, development, and maintenance of data environments	Not an issue, software providers have their insurances. Need for governance was identified
Human	Human motivation and acceptance to use new technologies and data	The technical capability, nature of work and motivation were seen as large challenges to overcome

Table 4 Challenges from empirical research

Both organizations were in the middle of solution implementations, hence some of these issues were to be solved in near future. From the organizational point of view, lack of vision in the integration of the data infrastructure and the implementation of data-driven culture were seen as the most present challenges at the moment. The lack of vision in the data ecosystem was due to implementation of various systems, such as contract

repository, to only serve the need of storing the contracts. It was not considered that these systems would generate new data in the future, such as meta data, which then could be linked to other systems. This problem is now limiting the possibilities organization could have combining the unstructured textual data with other data infrastructure entities.

> *OA: We build them just to cover the current need of having a place to dump the contracts, but we didn't make it easy to analyze them in the future. And this happened with contracts, and it happened with other procurement tools, systems and solutions. We just focused on getting the job done right now for me, and we did not think why there and that.*

For data-driven decision-making organization saw both technical and organizational limitation. Data was not entirely available, but there was strong mandate to have this data available, and not anymore use the unavailability of data as an excuse. This challenge and the shift from, what was before human-driven decision-making, was recognized as one pressing challenges in the organizations.

Alignment was considered as important answering these challenges, both organizations felt they had alignment between their organization strategies. Actual implementation and processes around the strategy might be challenging, but support and alignment across organization, business, IT, and management was recognized.

> *OB: Our director who's steering the data project sees really the value of having uniformed dataset that is complete and reliable that we can make use of to leverage our value propositions. It comes also from business strategy, they are really stressing on implementing the tooling and getting things done asap in terms of data.*

From technical point of view the interoperability, integration, distribution, accessibility, and cost- and resource-intensity were seen as the biggest challenges. It was interesting, that these challenges handle, both the data, and the technology around the data. In order to have data speaking to each other and have it as usable information, we need the technology to speak as well, and vice versa. It is not either or.

The interoperability and integration relate also to the organizational challenges of having the vision in the strategy, it is not purely technical matter to create the integration between unstructured and structured data and the technologies.

> *OA: That's the challenge we have now. We need to explain the value, we need to say it's not just the contracts, cause the [supplier] contract you can connect to [supplier] risk, and then you can do to the strategies. And yeah,*

you are the team that putted the contract management system in place, and you just thought about a list of keeping track of contracts, but you didn't think of risk, and that's what we are doing to now, connecting the dots in a digital way.

The distribution and accessibility are also issue if the data strategy does not include them as key points. If the tools are siloed, so is the information and access. Global organizations can be using for the same things, different tools and solutions, making the distribution and access to information limited on a global level. And the effect is negative to the local organization as well. With optimized sharing and access of the information, local organization could improve their efficiency with optimized tools, and would not need to repeat redundant tasks, and the information would not be dependent on individual efforts. It was interesting to find out that the accessibility was seen as the problem also because the implemented software was not seen as source of data, it was seen as a place for the data, and the extraction of data from the tool had become problematic.

OA: Accessibility bit of a challenge, because we did not think these systems as sources of data, we did not build them in a way it is easy to extract the data.

Searchability and terminology was identified as issue in both organizations. This challenge did not only concern data but also the way of working, processes and key business definition. With data, organizations did not have unified view or processes for managing the data. Therefore, it was not possible to have all data searchable with same terminology, one business unit considered data point x as y, as other considered it to be z. Uncertainty of what the data is, or is it complete for usage, was issue.

OB: What we have now in our current setting, if we request for same data, for instance the spend of a supplier, it can be that for the same time period, we get different answer

Cost- and resource-intensity was mutually raised as concern in current solution scope. The amount of work to cleanse and manipulate the unstructured data was at the moment resource intensive. Lot of manual steps were involved, and optimized tools were not available.

OB: Expensive people are spending too much time on structuring the data due to lack of tooling

Both organizations were handling majority of their data related processes by outsourcing the tools and services, therefore for example, the security and scalability were not seen as the most urgent issue. Those were considered to be covered by the software and service providers, of course not understating their importance.

Interestingly, from the interviews a human related emerged as one of the most pressing challenges. Both organizations saw that one obstacle in the digitalization transition, was the legacy human culture and human acceptance, and the change management related to it.

7.2.2   Data maturity and processes

For assessing the data maturity, the data maturity model and key data processes were presented to the interviewees. The data processes could be characterized as individual efforts, ad hoc, non-standardized, and manual for the most part. The results of various processes and analysis took long time, and the value is not evident in all cases.

> *OB: There is no uniformed classification of the data fully implemented yet. This can lead to incomplete data. Extractions are performed manually by different people with different capabilities and knowledge of the system. It is easy to make a mistake in such a complex environment. You have to have a lot of knowledge.*

Data-driven decision-making was not incorporated into organization culture resulting into challenges with data management and processes related to it, limiting the maturity of the organizations.

> *OA: This comes from the company culture, it's a human culture, people working there. It's not so much process optimized company and that has its advantages and disadvantages. And in this case, with data, it is disadvantage, because we have very few defined and standardized processes, and if you don't have you cannot embed technology software to generate data, because if some day you do it like this and some day you do it like that, you cannot agree what we need to build. So that's the situation, some cases repeatable, some cases ad hoc.*

Both organizations saw that their current data maturity is limited to the first stage, being ad hoc and non-repeatable. In the organization A, for some parts of the organization

higher maturity was recognized. Processes were repeatable and efficient, and not dependent on individual efforts. This concerns especially business critical functions, other functions are to follow with this path, still behind in the maturity though.

It was notable that, our interviewed organization were in a bit of different stage when it comes to maturity, and this was due to the stage of procurement digitalization. This gives us good chance to reflect and share the findings. It was noticed in the more mature organization, that having the defined data processes in place, software and use case agnostic, is critical for success. The whole lifecycle needs to be considered, from capture of the data, to sharing the data to other sources, as the tools generate new data. This is key. It is not only enough to plan the process to be one-off project, where data is cleansed to the platform. The platform needs to be sustainable and linked to each other, as the data will cycle across the platform. Organizations did acknowledge the need for more mature processes, and their future development aimed for the mature processes

7.3 Data strategy

7.3.1 Alignment

Both organizations saw their business and data strategy to be aligned. Although this has not the been the case always. But lately, the organizations had recognized the data as a primary source for competitive advantage, and this has been implemented across organization as a driver in the business strategy.

7.3.2 Involvement

The involvement of various functions was supported in the organizations. To have this, alignment is requirement, and that was in place. In the more data mature organization, the responsibilities were divided into data, IT, and business functions. All these entities were involved in the strategy planning and implementation. Data and IT functions were shared across different business functions, such as procurement and sales. Those two parties were driving the strategy planning, keeping the business functions involved in this process in a collaborative way. In the less mature organization, the approach was similar but with one IT party. The planning and implementation was supported by management, and also external partners.

7.3.3 *Vision*

Vision was interesting topic, it has come across in previous points as well. In one of the organizations, we identified the lack of clear vision as main reason for incomplete utilization of the data. From this learning, the organization had developed their vision. This vision emphasized the enablement of information for the business decision. The data was made available, and with this not basing the decision on it was no more acceptable. Secondly, organization aimed to create automated and more mature data processes, and enable the brilliant people working at the organization to focus on the value-adding activities more, instead of wasting time with the repeatable processes, and so maximizing the value of human factor. Their vision on solving these problems was with dual approach, by serving the immediate needs of the users, and at the same time creating and developing robust data environment. Key topics in this process are digitalization of the tools and processes, alignment with the data, availability, and transparency of the information.

Consolidating the vision of our other organization with a quote:

> *OB: One single source of truth per data type and to be able to access this full, up to date, cleaned & uniform data via the designated tools*

This sums quite brilliantly the vision for the organization. Trust with the one single source of truth of all data. Availability and actionability of the data for turning the data into usable information and value, in which the tools enable this process.

7.4 Data infrastructure and requirements

7.4.1 *Layers*

As the industry best practice is, organizations had their infrastructure visioned as an ecosystem, where all the entities would be connected. In the more mature organization, the infrastructure design consists of a software and data layer. The software layer included for example the ERPs, data repositories and various analytics tools. And to this software layer was connected the data layer. The data comes and goes between these layers. The data is stored, processed, managed, and analyzed within software layer, and then sent back to data layer. With this approach organizations can have multiple different tooling and systems for similar data. The data connects the organization. As we have discussed, there are challenges in making the connection, which the organization is now struggling. In the other organization, which has still their infrastructure in the implementation stage,

the vision was to move away from siloed data storing, to unified and shared data lake, which would be then connected to various tooling systems. The data flow would be from the various ERPs of organization, to centralized and global data management tool making it uniformed, and storing the uniformed data in the data lake, which would feed and eat the data from and to all the other applications and systems. They also wanted to make these systems linked and interoperable. So quite similar approaches taken for managing the data.

For the management of unstructured data in this infrastructure organization had similar approaches. The unstructured textual data, found from the various documents, was handled with specialized processes. The contracts were stored to contract repository with all their data included, the structured meta data and the unstructured textual data within the documents. The unstructured data would then be analyzed further suing advanced text analysis solutions, and also connected with other sources and ends using the meta data, and results from text analysis.

7.4.2 Requirements

When designing the infrastructure, organizations recognized following four topics as key requirements. It is necessity to know how the data is built, to which the data will be used for, to make the unstructured data more structured, and to store the data in way it can be used again. Organizations also emphasized making the data available for the whole organization with transparency and efficient data processes.

> OB: First, we need to understand a) how the data is built up, and b) what we are going to do with the data. And then, we need to make unstructured data 'structured', so it can be used. And lastly, store the findings in the system

Key activities for having the unstructured textual data more structured, was seen to be applying meta data, and running advanced text analysis. Meta data, such as owner and date, were given to the contracts and documents found from the system. This enriched dataset, the raw contracts and metadata, was then stored in the repository system. The purpose for metadata was improved data management and integration of the various data sets and tools. Text analysis was then performed to the actual raw unstructured textual data found from the contracts. Organization could use various techniques for this process. The interviewed organizations had few approaches. More mature organization had sourced a service from 3rd party providers to run advanced text analysis using NLP and other AI algorithms. They wanted to search relevant information found from the text,

such as dates, parties, terms, or laws mentioned. The other organization had a data management project in their pipeline, where the goal was to improve the quality and structure of the data at source. In addition, ad hoc analysis was seen as one process, where relevant information could be searched from the unstructured textual data using different types of tooling. Problem was raised on the quality of the tooling, it should be advanced and designed for the purpose, to receive full value of the unstructured data.

7.5 Capability

7.5.1 Tangible

From our research framework, data, technology, and basic resource, were the tangible resources of an organization.

Data was seen as available in both organizations, but dependent on the technology. The ownership and management of data was defined in the organizations. Previously mentioned data challenges concerned the data itself, but were related more to the technology according to interviewees. Majority of the technology resources were still in the implementation phase in both organizations. The first solutions were meeting the requirements mentioned before, thus fulfilling the capability requirements. Though, there was some limitations. Previous technology implementations resulted poor results due to the lack of vision in the data strategy. In addition, the misalignment between IT and business on the delivery methods and expectations of the technology were recognized as problems. This misalignment created challenges for organizations to develop their technology capability.

The basic resources, time and money, were seen as important resource in both organizations. Both organizations recognized the absolute need for the basic resources. Some misalignments were identified across organization in the available resources. Also, there was clear message of needing constantly more resources from IT and organization.

7.5.2 Human

From our research framework, human resources were classified into technical, managerial, and procurement skills. Technical related to capabilities on digitalization and usage

of the data. Managerial handling the capability of management to support their organization in implementing and developing skills. And procurement skills, assessed with the level of professionalism in procurement.

Technical skills were seen as a current challenge in the organizations. The legacy of human driven processes in the organization was seen as obstacle for data-driven decision-making adoption, and for the general usage and management of the data. Organizations were lacking, both the technical ability and motivation of the people, to use the data. It was raised that especially, in procurement roles, such as buyer, the technical skills were insufficient on a general level.

Managerial skills relate to how well are the implementations and development supported and enabled. It was noticed that some support is provided, but it is not sufficient. There did exist a general training from the shared service of the whole organization, but the actual support from local management, to ensure the business users receive the understanding of why, what, and how, was missing. The development of the technical skills, and was seen as important resource for the future, and prioritized in the management.

Procurement skills were on a sufficient level. Especially, concerning the core procurement activities. Procurement organization had mature strategies for maximizing the service they provide to their stakeholders. Information systems were included in the operations, also cross-organizationally and with key external parties. Procurement performance was measured through various indicators, such as total costs, provided service, value, and risks, but there was some misalignment on the global language of procurement, resulting into limited performance. Not all the functions and organization in and out of procurement were aligned globally, and local entities had their own ways of doing business. This limits the maturity level of the global procurement function.

7.5.3   *Intangible*

From the research framework, the intangible resources were data-driven culture, organizational learning, and actionability. These all relate to each other, and are referenced in the prior discussions of this research.

Currently both organizations recognized that they are in the middle of shifting towards data-driven organization, but not there yet. Topics affecting this shift were the legacy of human-driven culture, previously mentioned insufficient tangible and intangible resources, and the current nature and description of the organization work roles. The old legacy of human driven organization makes it difficult to create defined data processes and adopt users to work with the data. Tangible and intangible resources need to be in place, such as the data and software using that data. And the work role is important, since at the moment the role of a typical procurement people is buyer. Their work is to negotiate

and buy, their work is not to create and manage the data. This needs to be hurdled in some way to ensure people start better engaging with the data, thus enabling the shift to data-driven organization.

Organizational learning was referred in the same context as the data-driven organization. It was recognized as insufficient at the moment, but something that it is developed and focused on. Executives are emphasizing organization wide digitalization process, and the learning culture from the digitalization initiatives. Still, lack of resources and commitments were seen as concern. Organizations had sourced services from external parties to help with this journey.

Actionability roots to the data-driven decision-making. Limitations on actionability were the availability of the data, and the legacy culture and working habits with the data. The support and incentive were in place for turning the information available to actions from management.

7.6   Performance

For analyzing the expected value, we used the performance indicators of cost, risk, quality, and value since those resonated well with the interviewees. To start, we discussed why procurement would need to start analyzing the unstructured textual data, and what business problems it would solve.

7.6.1   Current business problems solution would solve?

Organizations were not able to capture the information found from the unstructured textual data in the various documents. Advanced and working solution for analyzing the unstructured textual data would solve this, and enable the organization to know the important information found from the documents.

> OA: But it's [textual contract data] a very long text, we write, we sign it, and we store it somewhere. And then, we don't review that often. So, it could be that you are doing something that it's not according to what you have signed in the agreement.

Key problem limiting the analysis now is the incomparability and integration of data. If the analysis is somehow done today, it is really resource intensive. People are spending

the time cleaning and manipulating the data. This limits the capable people in the organization to run their value-adding activities.

> *OB: The problem of incomparability of the data, and the time assigned making the incomparable data comparable*

The limitation also concerns the other functions than procurement, for example finance would need information from these documents, and currently the information sharing is not at the required level. What makes the sharing of information within different procurement units, and across functions, difficult is the information language used. All have their own tools, processes, and definitions. Having unified way of working with the data would solve this problem.

7.6.2 Cost

If organizations would be able to analyze the various unstructured textual data found from the contracts, in a uniformed and efficient way, they could obtain tangible cost savings. It could be that certain supplier is providing same goods or services, but with different descriptions or codes. Now, if the organization could know from the unstructured textual data which is indicating the goods or services in question, they could reach price savings with using the lowest price offered by the supplier. Cost savings could also be generated from capitalizing the terms, such as discounts and rebate terms, found from the supplier contracts in unstructured textual data format. Having the access to the information, organization could claim their rights.

> *OB: If we are able to structure unstructured data this can have huge possibilities for the future … you then can find the cheapest suppliers … you then move all your volume to one supplier you'll have a win-win situation, … results to volume bonuses or lower prices on your side.*

> *OA: It could be that you are doing something that it's not according to what you signed in the agreement, or you are missing part of the deal, so you could have if you buy 200 tons in the year from you, I get the a bit of discount. Well, if you know you already bought 200 tons you need to claim that discount. But you need to know it from your contracts.*

In general, having access to information would improve the negotiation power of the organization with suppliers, yielding into various cost savings in all of their procurement activities.

7.6.3 Risk

Organization saw that having access to all the unstructured textual data, would improve risk management. The unstructured textual data includes a lot of important information concerning the organization, such purchasing terms, laws, and obligations. Not knowing this information imposes a substantial risk. These could be related to non-compliant activities, resulting into regulation and reputation risks.

> *OA: We might non-compliant, one example is reputation. There are things you need to be compliant by law, because EU says that or local law says that. And, you need to able to say, are my contracts aligned with that law and jurisdiction. Then you need to go check, and well if you could do the analysis on the contracts you could get the answers*

Enabling the analysis of various unstructured external data, from internet or other documents, giving risk and other information on organization supply chain and its parties would also improve the risk management.

> *OA: You could a look bit broader and say find relevant information on a supplier, website of suppliers, could be finance report, could be news, you go to yahoo business and get latest information on the supplier, but it could be announcement, PR releases, I don't know you might even know about risk, I know a lot of this suppliers in this region, and this region is going to have typhoon, earthquake so there's going to be some disruption, so you should be able to capture that information and turn it into risk for a group of suppliers, so there is some unstructured data, text based mostly, but it's sort of scraping and scavenging in the internet information and condensing it.*

To value organization expect to receive in their risk management from having the unstructured textual data available is firstly, knowing what is said in contracts, which in essence define the rules. And then, improving the risk position the added information found from the contracts.

7.6.4 Quality

Improving the solutions for the analysis of the unstructured textual data, would make the analysis processes more efficient and higher quality. Time spent cleansing and analyzing the data is now wasted resource, as the tools are not optimized.

> *OB: Of course, you can do this in excel, however it's so time consuming and you don't have the insights in a click of button yet. And there we are not yet at the moment. Reaching the potential value in the same amount of time we should be able to get when once all data is stored correctly.*

Then also, having the unstructured data available, would enable the new possibilities on how to improve the quality of procurement function. New valuable information could be used in the decision-making more efficiently. This could be for example capability to manage disruption.

> *OA: Something we came up during Covid, when the situation becomes really bad, we need to know which contracts have certain clauses that allows me to change them to more council or do something, and then you look for force major clause in contracts, acts of good, or things completely unexpected on a massive scale, that can make this contract invalid. That's written in the contracts, but you need to find which of the contracts supplies. This is something that analyzing those contracts will add value, because it helps you really quickly measure your exposure, or which supplier you going to have problems, or with which supplier you have flexibility because that contract already allows that flexibility.*

7.6.5 Value

Value propositions obviously relates to all previously benefits, they all provide value to the organization. Though, it is important to handle it as own dimension, to ensure all possible value propositions are included. Interviewees identified unstructured textual data to have effect on the overall value procurement can offer internally and externally. Important internal value was the enablement of digitalized organization, which can utilize all available information. External value was seen as larger theme, with the correct solutions, organizations can start driving their external their supply chain partners, for better performance and value, such as sustainability or compliancy related.

OA: For instance, we say we want to go one step forward in our sustainability agenda.

Meaning it is not only internal costs or risks, but also external value that is enabled. Organizations could enable, for example all the purchasing done is compliant with the sustainability requirement found from the supplier contract, or other external data sources. This would increase the positive effect of the procurement to many parties, and increase the overall value provided to the network.

8 FINDINGS, CONCLUSIONS AND DISCUSSION

In this chapter, we will summarize our findings from the interviews. Then, we will answer to our research questions, using the framework composed from the literature review, and the key findings from our case study interviews. After this synthesis, we will discuss the research trustworthiness, its contribution, possible future research.

8.1 Findings from the interviews

The interviewees were both well informed on the current capabilities and possibilities of the utilization unstructured textual data in their organizations. The discussion followed great structure, and interesting insights were identified. To consolidate the interviews, it is good to use following themes of our research: discussion on the actual data, capabilities and requirements related to it, and the potential value. These will help us to construct the findings, which we will later use to assess our research questions and make the conclusions of our case study research.

Our case study was set to explore the unstructured textual data in procurement organizations. Organizations defined the unstructured data as something that is not controlled in terms of what it is, comes in many forms, and has lots of potential. The source and format of the unstructured textual data was supplier related procurement documents. These documents were seen as extremely important, since in essence they define the rules of engagement with your suppliers. Therefore, organization were motivated to understand how this data could and should be managed in the procurement organizations. The initial capture of the various documents, and the unstructured textual data they hold, was identified to require mature data processes. After locating and storing data, organization pursued to making this data available for analysis. This was considered as challenging task. Largest challenges organizations faced were complexity of unstructured data, structuration of the unstructured data, integration of the various data sources and assets, incomparability and interoperability of the data and the related tools, resource intensity of the data management processes, lack of vision in the data strategy, and embedding data-driven culture to the organizations code. The maturity of data processes was important for addressing these topics. All data management processes ought to be repeatable, standardized, automated, and uniformed across organization. Organizations pursued to develop their less mature data processes to more mature.

After defining the data, and addressing the management of it, we discussed the requirements and capabilities organization should have to enable the analysis of the procurement documents and the unstructured textual data found from those. The key themes were data

strategy and infrastructure, and the capability using resource-based-view. The data strategy needed to be aligned with business strategy, and the involvement of the whole organization was required. The vision was identified as critical, since the missing vision and goals with the data strategy had previously limited the capability and performance in one of the organizations. The vision and goals needed to include all the previously mentioned challenges. The governance and support of mature data processes were seen as enabler for the implementation of the data strategy. Key requirements of the data infrastructure were integration of all the data into one place, alignment on the purpose of the data, making the data uniformed and usable, and enabling efficient storing and sharing of the data. The infrastructure needed to be defined with mature data processes, and not to be dependent on individual efforts. The technical infrastructure of the organizations was designed to be an ecosystem, in which all the data and software are interconnected. The infrastructure emphasized enabling the information for all users, with transparency, and efficient processes. The unstructured textual data is processed with its own advanced tools, such as contract repository supporting meta data, and advanced text analysis tools with NLP, text mining, or other AI algorithms. The processed unstructured textual data is integrated with other data and systems in the ecosystem, for organizations to run value-adding analysis. There still are certain capabilities needed from the organization to turn the data into value. These we assessed through resource-based-view. There are tangible, human, and intangible resources. Summarizing the most important resources and the capabilities of the organizations, tangible resources are data and technology, discussed above, and the basic resources of time and money allocated for the information systems. Human resources were the technical, managerial, and procurement skills. More digitalized, visionary, and mature the organizations are, better is their capability with the human resources. Lack of technology skills was seen as important issue to be solved in the organizations, currently limiting now more advanced analysis, such as unstructured data utilization. Intangible resources were the data-driven culture, continuous learning and actionability of the data. These were considered important resources to have in to improve the overall capability in the frontier of digitalization.

Lastly, we want to summarize what business problems organizations expect to solve, and what is the deriving value, with the utilization and analysis of the unstructured textual data in their procurement. Main business problem organizations have now is the unavailability of the business-critical information found from the procurement contracts. They do not have the visibility to this unstructured textual data and cannot use it in their decision-making processes. People are either spending a lot of time managing the data, or then not using it at all. We already have discussed how important the information found from the procurement contract could be. They defined the rules of the game with the suppliers. But also, they do offer more information and value to procurement. It would create tangible cost savings, as they could improve their sourcing efficiency, for example

optimize their material prices with suppliers or claim their rights for discount and rebates. Having the information available, would also improve their risk management. They would be ensured that they are following the terms stated in the contracts, and this way mitigate risks related to compliancy. Or they would have all possible information available related on their suppliers, enabling more informed decisions. Advanced and robust solution for the analysis of the unstructured textual data, would yield into improved internal processes and efficiency. Organizations could focus more on value-adding activities, and maximize the human factor in the decision-making, as they would not have to waste money on repeatable and manual data management processes and analyses. The overall quality of procurement would improve, as they could use the information more efficiently in their data-driven decision-making process and share it across organization. This could increase external value propositions of procurement, such as driving sustainability agenda forward in their network.

8.2   Conclusions

Our research was motivated to study how procurement organizations should approach unstructured textual data. Furthermore, identify the expected value of the unstructured textual data in procurement analytics.

To answer the first research question, we first defined the unstructured textual data, what is the source of it, how it is managed, and what are the factors affecting its availability. From the 5V definition (Demchenko et al. 2013), other literature (Inmon & Nesavich 2007), and the interviews, we defined the unstructured data as a data asset that comes in many forms and high volume, is hard to manage and control, requires extensive efforts to have it as meaningful information, and possess a potential value. It could be considered as an opposite to structured data, which is managed in relational data models. Most interesting format for procurement was unstructured textual data. The most prominent source for the unstructured textual data was procurement documents related to their suppliers, such as supplier contracts. The unstructured textual data was not per se these documents, but the actual written text within those documents. After locating the data organizations need to make the data as available information they can use. We studied the management process of the unstructured textual data. Starting with the capture of the data, pursuing for standardized and mature processes (Korpela et al. 2017; Kasravi et al. 2003; Kruk et al. 2007), which was seen as challenge based on our case study, resulting into not-optimized processes in the capture process. Organizations could apply metadata into the documents for improved integration possibilities or standardize the documentation processes. Alternatively, advanced text analysis processes can be performed on the data to make it

more searchable and meaningful. The goal of all these processes is to have the unstructured textual data available in the same environment as the structured data, thus enabling more informed business decisions. There exist though challenges, both technical and organizational. Most present technical challenges were the integration of the data, interoperability of the data and systems, resource intensity, and searchability and terminology of the data. The nature of unstructured data, context-sensitivity, uncertainty, and complexity, the requirement for advanced technology created these challenges, and some of the organizational challenges created these challenges. From organizational point of view, the alignment between business and data strategy was considered important factor, in addition to data-driven decision-making process, and distribution of information. Our research also included the data maturity and processes as key factors when making the unstructured textual data available. The more mature and advanced data processes organization have, the better capability organization would have. Both case organizations recognized themselves as not that mature but were motivated to increase the maturity of their data management processes, to improve their capability with management of the unstructured textual data. It was considered as a requirement to improve the data management processes towards more repeatable and automated direction. To summarize, procurement has unstructured data available, the most prominent being the unstructured textual data found from the supplier related procurement documents. To have this data available, as usable information, organizations need to integrate the data into their current data infrastructure and make it comparable. This task includes both technical and organizational challenges, in which more mature data management processes can help.

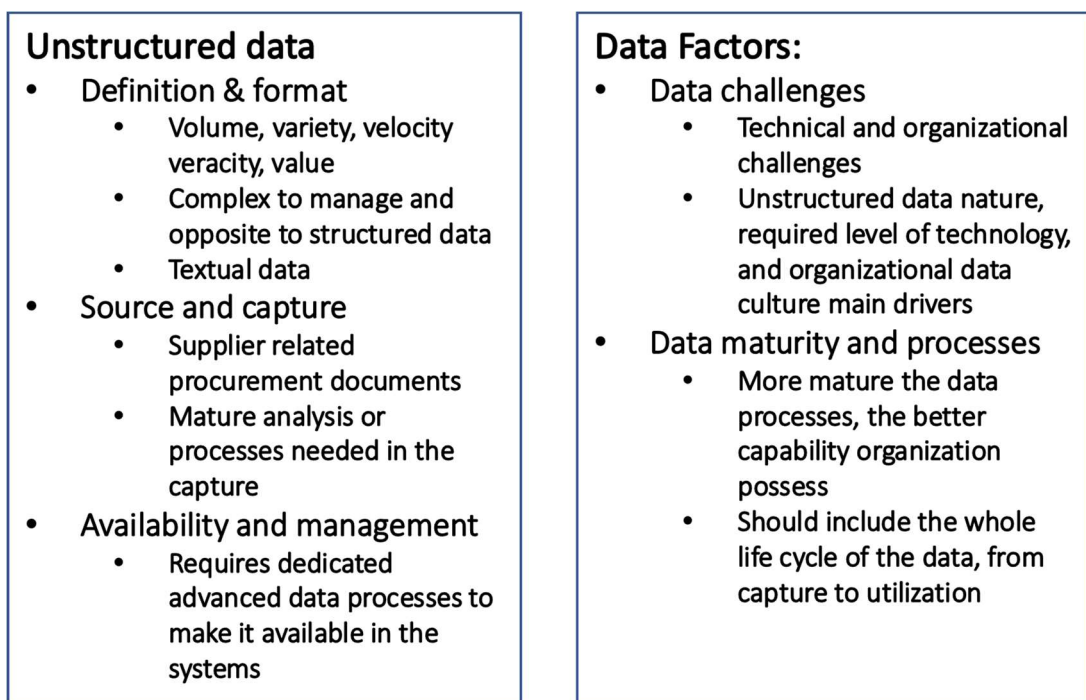| Unstructured data	Data Factors:

• Definition & format
   • Volume, variety, velocity veracity, value
   • Complex to manage and opposite to structured data
   • Textual data
• Source and capture
   • Supplier related procurement documents
   • Mature analysis or processes needed in the capture
• Availability and management
   • Requires dedicated advanced data processes to make it available in the systems
	• Data challenges
	   • Technical and organizational challenges
	   • Unstructured data nature, required level of technology, and organizational data culture main drivers
	• Data maturity and processes
	   • More mature the data processes, the better capability organization possess
	   • Should include the whole life cycle of the data, from capture to utilization

Figure 10 Overview on the available unstructured textual data, RQ1

In Figure 10, we have summarized definition and format of the unstructured textual data, its source, and capture and management process. The figure also highlights the key factors which should be considered, which are the data challenges and maturity and overall data processes of a organization.

As we have now defined the unstructured data, and the basic principles related to the management of this unstructured textual data, we can continue answering the first research question: How should procurement organizations approach unstructured textual data? In our research framework, we had the key requirements for an organization's data strategy. Those were the alignment of the business and data strategies, involvement of the whole interested organization, governance and support of the mature data processes, and clear vision and goals of the data strategy (McAfee et al. 2012; Deloitte 2019; Loshin 2013, 31-36). Our case study interviews supported these to be key topics. The infrastructure layers and requirements were discussed in the literature by Baars & Kemper (2008), Land et al. (2009), and Kassner et al. (2014), and in Figure 4, all recommending for integration layer with defined ETL-processes for each format of data, data repository for the integrated data, and analytics layer for value-adding analysis using both structured data and unstructured data. The infrastructure design should include definitions for rights to access, use, share, and reproduce the data (Munoz-Arcentales et al. 2019). From our interviews, we would still emphasize the requirement of connecting these layers to each other. The systems need to be designed for their special purpose, but with the linkage to other systems, since data is being created in these systems all the time, and this data should be used in other systems as well. The actual architectural design can be designed accordingly to the data strategy. The key requirements are the integration of all the data and enabling the required information for the whole organization.

For capability assessment, we used resource-based-view, which is practical for assessing the capability (Wade & Hulland 2004, 131). The resources were tangible, human, and intangible. The tangible resources were the data and technology, which were discussed, and then the basic resources, time and money, both seen as necessary for any project (Peppard & Ward 2004, 176). The human resources we classified into technical, managerial, and procurement skills. Technical skills were considered as important (Deloitte 2019) and seen as key issue in our case study organization, as well as a provider of competitive edge (Davenport 2006). To achieve the full value, technical skills needed to be included into procurement skills, which was not seen as an easy task. The primary job of a procurement person is not a data analyst, but a buyer, and the nature of buyer the role needs to be changed to be more data-driven. For that, procurement maturity development is needed (Rozemeijer et al. 2003) and management support. The more mature procurement is, the better capability it has, to achieve the full value with advanced technologies (Schiele 2007). We consider the organization's intangible resources, data-driven

decision-making, organizational learning, and actionability to be critical with any data project or strategy, especially with new technologies (Mikalef et al. 2020; LaValle et al. 2011).

RQ1: How should procurement organizations approach unstructured textual data?

Data strategy
- Alignment of business strategy with data strategy
    - Needs to be fully aligned
- Involvement of whole interested organization
    - Management, IT and business need to collaborate across organization
- Governance and support of the mature data processes
    - Requirement for success
- Clear vision and goals of the data strategy
    - Needs to be embedded to strategy and well balanced

Data infrastructure requirements
- Whole lifecycle of the data and relating rights considered
- Connected systems and data
- Integration of all the data into the infrastructure
- Heterogeneous approach with mature data processes
- Enabling value adding analysis for the users

Capability
Tangible
- Data
    - First and obligatory resource for the capability
- Technology
    - Advanced technology improve capability
- Basic resource
    - Required for successful projects

Human
- Technical skills
    - Important resource affecting the capability
- Managerial skills
    - Support and relation to data strategy contributes to capability
- Procurement skills
    - Maturity contributes, with dependency on technical skills

Intangible
- Data-driven culture
    - Important
- Organization learning intensity
    - Important
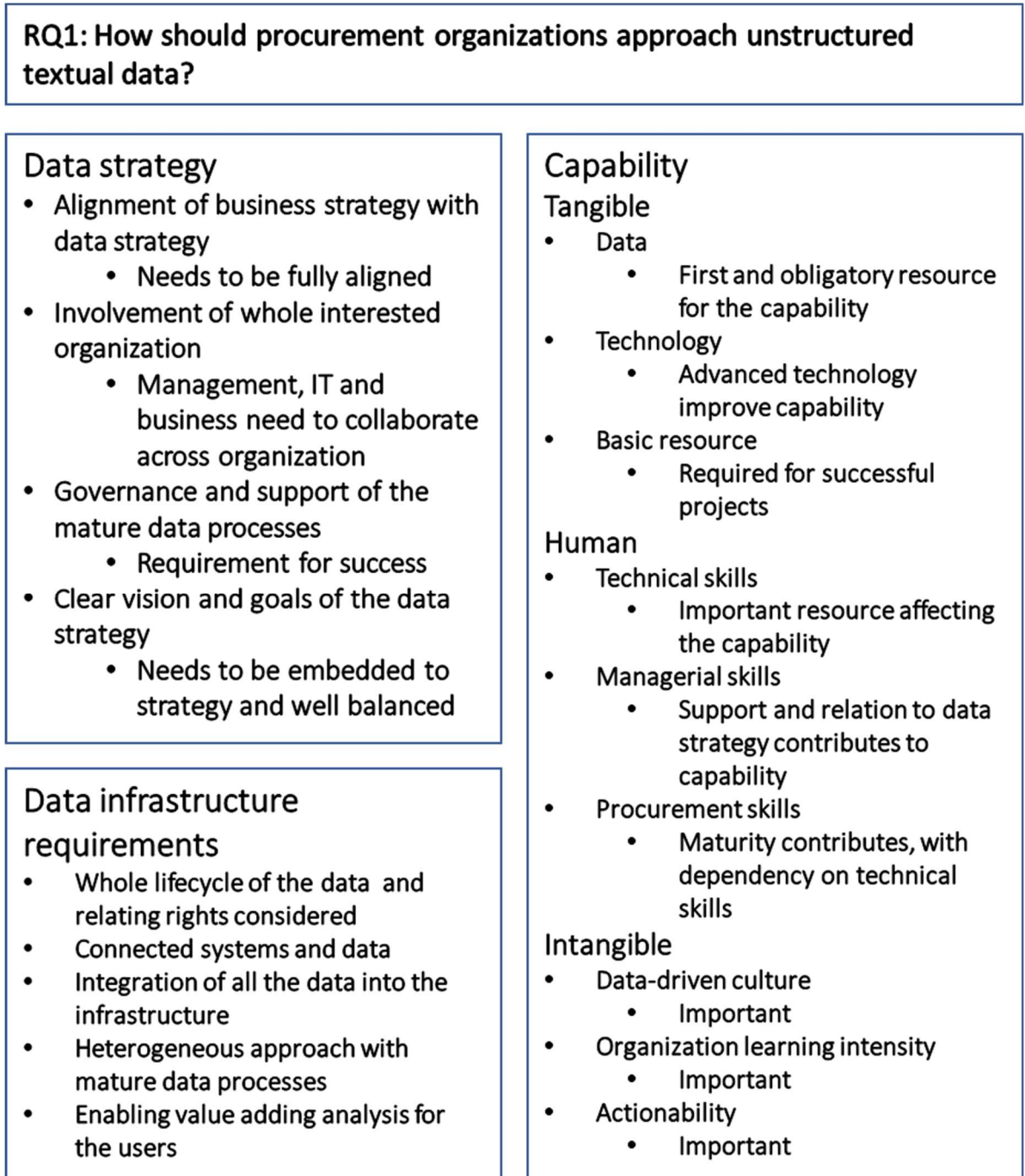- Actionability
    - Important

Figure 11 Overview on the approach to unstructured textual data, RQ1

Figure 11 highlights the main dimensions organizations should consider when approaching unstructured textual data: data strategy, data infrastructure, and overall capability. Under each, we list the key information, which answers our second research ques-

tion. Data strategy has four important factors which should be considered. The data infrastructure should have the mentioned requirements covered. And the capability of the organization is assessed with the three resources, which are tangible, human, and intangible resources, all of which are important for successful utilization of the data.

Lastly, we answer the second research question: What is the expected value of the unstructured textual data in procurement? Our case study with the procurement organizations provided us an excellent opportunity to assess this question. We identified the large potential with the unstructured textual data in procurement and the current problems in managing it. Our literature supported our findings from the interviews on the prominent value of new technologies in procurement (Handfield et al. 2019), and the struggle to utilize this type of data asset (Accenture 2014, DalleMule & Dav-enport 2017). Turning this vast amount of various unstructured textual data into usable information for business decisions would yield various benefits. The core value proposition is having more meaningful information available in the business decisions, resulting in improved performance (McAfee et al. 2012). Procurement especially needs all the possible information, with its large impact on business profitability and operations. The supplier contracts basically hold the rules of the engagement with the suppliers. If the organization cannot access and use the data within those, they miss a large portion of the information they could and should use in their procurement activities. Having access to the data and making it usable information, organizations could gain value in important procurement levers of cost, risk, quality, and value. (Deloitte 2019, Handfield et al. 2019). Cost savings would be generated from the improved negotiation power with the suppliers. Sources for these savings would be pricing optimization, improved margins from rebates and discounts, and all the other tangible savings deriving from improved negotiation power with the suppliers. The risk management would improve as the organization would actually know what the contracts say, what the agreed terms are, and what is their exposure to those, thus, mitigating their risks related to compliance. The quality of procurement activities would also improve. People working and cleaning the data with manual processes are wasting a lot of valuable time. This process could be automated, enabling the people to focus on the more value-adding activities, where human factor has a larger impact, using the previously unavailable contract data to their advantage. Procurement would increase the overall value of their network by sharing the information across and enabling new possibilities for all the parties.
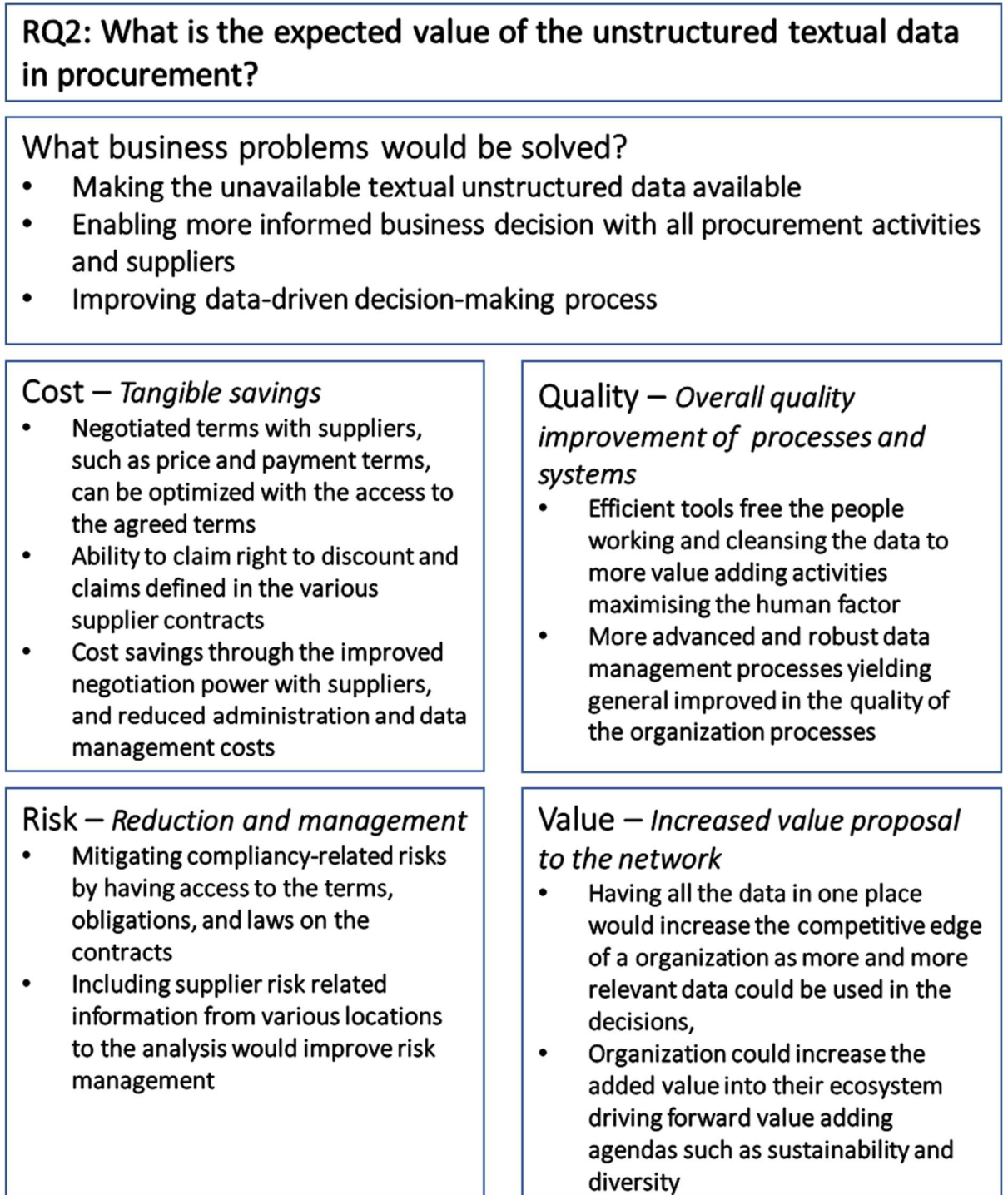
RQ2: What is the expected value of the unstructured textual data in procurement?

What business problems would be solved?
- Making the unavailable textual unstructured data available
- Enabling more informed business decision with all procurement activities and suppliers
- Improving data-driven decision-making process

Cost – *Tangible savings*
- Negotiated terms with suppliers, such as price and payment terms, can be optimized with the access to the agreed terms
- Ability to claim right to discount and claims defined in the various supplier contracts
- Cost savings through the improved negotiation power with suppliers, and reduced administration and data management costs

Quality – *Overall quality improvement of processes and systems*
- Efficient tools free the people working and cleansing the data to more value adding activities maximising the human factor
- More advanced and robust data management processes yielding general improved in the quality of the organization processes

Risk – *Reduction and management*
- Mitigating compliancy-related risks by having access to the terms, obligations, and laws on the contracts
- Including supplier risk related information from various locations to the analysis would improve risk management

Value – *Increased value proposal to the network*
- Having all the data in one place would increase the competitive edge of a organization as more and more relevant data could be used in the decisions,
- Organization could increase the added value into their ecosystem driving forward value adding agendas such as sustainability and diversity

Figure 12 Overview on the expected value of the unstructured textual data, RQ2

In Figure 12, we can see the main business problems organization can solve with the utilization of the unstructured textual data, and the deriving value propositions. Assessing this, we can definitely understand why procurement should be interested in including the unstructured textual data in their analytical processes.

8.3   Limitation and future research

We assess the trustworthiness of our research from critical point of view. Research methodology uses credibility, transferability, dependability, and conformability as the critical perspectives. Our research framework is based on the extensive and transparent literature review. The framework presents a logic process, studying first the subject and its characteristic, then the critical aspects related to it, and lastly the expectations relating to it. This framework was used as a base in our case study interviews with the selected organizations. The empirical research was instrumental case study, seeking how the themes in the research framework, exist in the selected example organizations. This information was then used together with the literature to draw conclusions on the case subject. These conclusions could be used in other locations as well, since our research framework is logical, and transparent in terms of what was studied and why. All though, we comment that the purpose of this research was to study how the unstructured textual data could be utilized in these specific organizations. We tried to provide the conclusions in understandable point of view, opening the technical definitions, sharing information on the models used, focusing on practical implications and the expected results with business use cases. We though note that case study approach imposes limitation on the generalization of the conclusions. Our scope was limited to two organization, and the procurement documents relating to supplies, such as contracts. This implies that not all organization have similar problems, processes, or expectations, but with the literature framework, organizations are able assess their status when fitting. We also note the limitations in our empirical research findings that case organization were in the implementation stage, and did not have yet finalized their solution, resulting into more conceptual discussion, though the motivation was to get practical learning on the discussed solutions. The results from the empirical interviews are aligned with the research framework, supporting our chosen methodology and approach to be logical.

In our research, we did not discuss the technical implementation of such an application to keep the scope meaningful. We would recommend the field to turn focus on the practical application and research in this domain, using the approach and framework of this research as guidance when assessing or starting a project or research. As digitalization evolves and becomes more mature, more organizations are able to utilize various kinds of data assets more efficiently. Studying the value received from the unstructured textual data with a larger sample of case organizations would be interesting. Most interesting technical applications are related to metadata, advanced text analysis processes, and general management tools of the big data, whether it is unstructured or structured. We assessed the organizations' capabilities from a resource-based-view. It could be worth assessing further which resources are the most critical for adopting new technology, such as unstructured data analytics, and how the characteristic of those resources affects this

capability. We would also recommend studying the challenges, issues, and costs that might come with the unstructured textual data. Important topics for future research could be an assessment of the data rights, meaning the access, usage, transfer, and reproduction rights, of the unstructured textual data. In this research domain, privacy issues are important, especially with unstructured textual data found from the contracts. More research on the trustworthiness and quality of the unstructured textual data found from the procurement documents, such as supplier contracts, would be beneficial. As well as the research on the maintenance and administration of the contract data in the continuous lifecycle of the contracts and other similar documents. Our research focused more on studying the challenges of making the unstructured textual data available in the organizations and its expected value. The logical next step would be to study the utilization of the data in a continuous process in procurement organizations from a critical point of view, assessing the resources required versus the value expected. The unstructured textual data is a relatively new type of data that procurement organizations start to utilize in their operations; thus, more research from all aspects related to the unstructured textual would be needed.

9   SUMMARY

Our research was motivated to study how procurement organizations should utilize unstructured textual data, and the value organizations expect to receive from this data asset. What makes the topic attractive is the increasing importance of data, and the significant impact which procurement has on organizations' performance. The research reviewed the existing literature and composed a logical theoretical framework. This framework starts from the definition of the data and key principles related to it. After which, it presents the key factors which should be considered when approaching the data. It also includes key indicators to assess the expected value. As part of the research, we conducted a case study with selected procurement organization which were assessing how they could start utilizing unstructured textual data in their procurement analytics processes.

We define unstructured data as a data asset that comes in many forms and high volume, is hard to manage and control, requires extensive efforts to have it as meaningful information, and possesses a potential value. For procurement, textual data, found from example supplier contracts, seems most interesting. To have this data available as usable information for their business decisions, organizations need to integrate the data into their current data infrastructure and make the data assets and tools comparable. The most prominent technical methods for integrating the data would be utilization of metadata and advanced text analysis processes. This process includes both technical and organizational challenges, in which more mature data management processes can help. We should also acknowledge the complex nature of unstructured textual data as problematic, which should be researched further. Proper design of data strategy and data infrastructure needs to be in place in the organization. Our research used existing literature to create a listing of the most important factors in each. The data strategy should align and involve the whole organization and include a clear vision and governance. The main requirements for the infrastructure are dedicated and heterogeneous processes for integration and management of the data and connection across the various data and tools existing in the environment. Resource-based-view was used to assess the organization's capability to utilize the unstructured textual data in their procurement analytics processes. The resources are tangible, human, and intangible resources, which all affect the capability of an organization. We started the research by stating the importance of data to an organization's competitiveness. This is also the main value proposition of the unstructured textual data in procurement. It would enable more informed business decisions and improve the data-driven decision-making process if the currently unused unstructured textual from the various documents could be made available as information for business decisions. Organizations would receive cost savings, mitigate their risks, improve overall quality, and add value to their organizations. All these performance indicators are fundamental for procurement for increasing the competitiveness of their organization.

REFERENCES

Accenture Global Operations Megatrends Study. (2014) Big Data Analytics in Supply Chain: Hype or Here to Stay?

Baars, H. – Kemper, H. G. (2008) Management support with structured and unstructured data—an integrated business intelligence framework. *Information Systems Management*, 25(2), 132-148.

Bakshi, K. (2012) Considerations for big data: Architecture and approach. *IEEE aerospace conference* (pp. 1-7).

Baldwin, H. (2018) When Big Data projects go wrong. *Forbes.* <https://www.forbes.com/sites/howardbaldwin/2015/01/22/when-big-data-projects-go-wrong/#4b11e3a26231> accessed 03.10.2020

Ball, A. (2012) Review of data management lifecycle models. *University of Bath, ID-MRC.*

Batenburg, R. – Versendaal, J. (2008) Maturity Matters: Performance Determinants of the Procurement Business Function. *ECIS* (pp. 563-574).

Becker, J. – Knackstedt, R. – Pöppelbuß, J. (2009) Developing Maturity Models for IT Management - A Procedure Model and its Application. *Business & Information Systems Engineering* (3), pp 213-222.

Berg, B. L. – Lune, H. – Lune, H. (2004) Qualitative research methods for the social sciences (Vol. 5). *Boston, MA: Pearson.*

Blumberg, R. – Atre, S. (2003) The problem with unstructured data. *Dm Review*, 13(42-49), 62.

Carlsson, C. (2018) Decision analytics—Key to digitalisation. *Information Sciences*, 460, 424-438.

Chae, B. – Olson, D. L. (2013) Business analytics for supply chain: A dynamic-capabilities framework. *International Journal of Information Technology & Decision Making*, 12(01), 9-26.

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

Crowston, K. – Qin, J. (2011) A capability maturity model for scientific data management: Evidence from the literature. *American Society for Information Science and Technology*, 48(1), 1-9.

DalleMule, L. – Davenport, T. H. (2017) What's your data strategy. *Harvard Business Review*, *95*(3), 112-121.

Davenport, T. H. (2006) Competing on analytics. *Harvard business review*, 84(1), 98.

Davenport, T. – Harris, J. (2017) Competing on analytics: Updated, with a new introduction: The new science of winning. *Harvard Business Press*.

Demchenko, Y. – Grosso, P. – De Laat, C. – Membrey, P. (2013) Addressing big data issues in scientific data infrastructure. *IEEE International Conference on Collaboration Technologies and Systems* (CTS) (pp. 48-55).

Denton, J. W. – Kirchenbauer, T. W. – Lewis, R. F. – Mallinckrodt, I. J. W. – Richer, P. R. – Sweeney, D. C. – Willey, S. M. (2013) Method, apparatus, and computer program product for providing contract analytics U.S. Patent Application No. 13/835,878. *U.S. Patent and Trademark Office.*

Dhar, S. – Mazumdar, S. (2014) Challenges and best practices for enterprise adoption of big data technologies. *IEEE International Technology Management Conference* (pp. 1-4).

Elo, S. – Kääriäinen, M. – Kanste, O. – Pölkki, T. – Utriainen, K. – Kyngäs, H. (2014) Qualitative content analysis: A focus on trustworthiness. *SAGE open*, 4(1), 2158244014522633.

Eriksson, P. – Koistinen, K. (2014) Monenlainen tapaustutkimus. *Kuluttajatutkimuskeskus.*

Eriksson, P. – Kovalainen, A. (2015) *Qualitative methods in business research: A practical guide to social research.* SAGE.

Gartner. Big Data, < https://www.gartner.com/en/information-technology/glossary/big-data> accessed 07.02.2020

Gartner. Dark Data, <https://www.gartner.com/en/information-technology/glossary/dark-data> accessed 24.3.2020

Gupta, M. – George, J. F. (2016) Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049-1064.

Handfield, R. – Jeong, S. – Choi, T. (2019) Emerging procurement technology: data analytics and cognitive analytics. *International Journal of Physical Distribution & Logistics Management*.

Henderson, R. M. – Clark, K. B. (1990) Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

Hu, H. – Wen, Y. – Chua, T. S. – Li, X. (2014) Toward scalable systems for big data analytics: A technology tutorial. *IEEE*, 2, 652-687.

Inform. (2020) <https://www.informs.org/Explore/Operations-Research-Analytics> accessed 23.11.2020

Inmon, W. H. – Nesavich, A. (2007) *Tapping into unstructured data: integrating unstructured data and textual analytics into business intelligence*. Pearson.

Janssen, M. – van der Voort, H. – Wahyudi, A. (2017) Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338-345.

Kadadi, A. – Agrawal, R. – Nyamful, C. – Atiq, R. (2014) Challenges of data integration and interoperability in big data. *IEEE international conference on big data* (pp. 38-40).

Kasravi, K. – Quigney, P. – Kruk, J. – Varadarajan, V. S. (2003) U.S. Patent Application No. 10/279,189. *U.S. Patent and Trademark Office.*

Kassner, L. – Gröger, C. – Mitschang, B. – Westkämper, E. (2014) Product life cycle analytics-next generation data analytics on structured and unstructured data. *CIRP Conference on Intelligent Computation in Manufacturing Engineering* (Vol. 33, pp. 35-40).

Korpela, K. – Hallikas, J. – Dahlberg, T. (2017, January) Digital supply chain transformation toward blockchain integration. *50th Hawaii international conference on system sciences.*

Kraljic, P. (1983) Purchasing must become supply management. *Harvard business review*, 61(5), 109-117.

Kruk, J. M. – Quigney, P. P. – Tehrani, S. – Richards, P. J. (2007) U.S. Patent No. 7,165,036. Washington, DC: *U.S. Patent and Trademark Office.*

Lang, A. – Mera Ortiz, M. – Abraham, S. (2009) Enhancing Business Intelligence with unstructured data. *Datenbanksysteme in Business, Technologie und Web (BTW)–13. Fachtagung des GI-Fachbereichs" Datenbanken und Informationssysteme"(DBIS).*

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

Lomotey, R. – Deters, R. (2016) *Unstructured Data, NoSQL, and Terms Analytics. Big Data Applications and Use Cases* (109–143). Springer International Publishing.

Marr, B. (2017) *Data strategy: How to profit from a world of big data, analytics and the internet of things*. Kogan Page Publishers.

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

Mena, C. – Christopher, M. – Van Hoek, R. (2014) *Leading procurement strategy: driving value through the supply chain*. Kogan Page Publishers.

Mikalef, P. – Krogstie, J. – Pappas, I. O. – Pavlou, P. (2020) Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management,* 57(2), 103169.

Moretto, A. – Ronchi, S. – Patrucco, A. S. (2017) Increasing the effectiveness of procurement decisions: The value of big data in the procurement process. *International Journal of RF Technologies*, *8*(3), 79-103.

Munoz-Arcentales, A. – López-Pernas, S. – Pozo, A. – Alonso, Á. – Salvachúa, J. – Huecas, G. (2019) An Architecture for Providing Data Usage and Access Control in Data Sharing Ecosystems. Procedia Computer Science, 160, 590-597.

Myers, M. D. – Avison, D. (2002) *Qualitative research in information systems*. SAGE.

Negash, S. – Gray, P. (2008) *Business intelligence. In Handbook on decision support systems 2*. Springer, Berlin, Heidelberg (pp. 175-193).

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

Peppard, J. – Ward, J. (2004) Beyond strategic information systems: towards an IS capability. *The Journal of Strategic Information Systems*, 13(2), 167-194.

Ransbotham, S. – Kiron, D. – Prentice, P. K. (2016) Beyond the hype: the hard work behind analytics success. *MIT Sloan Management Review*, 57(3).

Rejeb, A. – Sűle, E. – G Keogh, J. (2018) Exploring new technologies in procurement.

Rozemeijer, F. A. – Van Weele, A. – Weggeman, M. (2003) Creating corporate advantage through purchasing: toward a contingency model. *Journal of Supply Chain Management*, 39(4), 4-13.

Russom, P. (2011) Big data analytics. *TDWI best practices report*, fourth quarter, 19(4), 1-34.

Sharma, R. – Mithas, S. – Kankanhalli, A. (2014) Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations.

Schiele, H. (2007) Supply-management maturity, cost savings and purchasing absorptive capacity: Testing the procurement–performance link. *Journal of purchasing and supply management*, 13(4), 274-293.

Stake, R. E. (1995) *The art of case study research*. SAGE.

Syed, A. – Gillela, K. – Venugopal, C. (2013) The future revolution on big data. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(6), 2446-2451.

Tan, M. H. – Lee, W. L. (2015) Evaluation and improvement of procurement process with data analytics. *International Journal of Advanced Computer Science and Applications*, 6(8), 70.

van Raaij, E. (2016) Purchasing value: purchasing and supply management's contribution to health service performance.

Van Weele, A. J. (2009) *Purchasing and supply chain management: Analysis, strategy, planning and practice.* Cengage Learning EMEA.

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

Von der Gracht, H. – Giunipero, L. C. – Schueller, M. (2016) Future-Proof Procurement. Now or Never: The Big Procurement Transformation.

Wade, M. – Hulland, J. (2004) The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS quarterly*, 28(1), 107-142.

Wang, G. – Gunasekaran, A. – Ngai, E. W. – Papadopoulos, T. (2016) Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98-110.

Wang, Y. – Kung, L. – Byrd, T. A. (2018) Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.

Watson, M. (2014) Three things you should know about big data and analytics. Supply Chain Quarterly Q, 3, 44-48. < https://www.supplychainquarterly.com/articles/904-three-things-you-should-know-about-big-data-and-analytics> accessed 24.11.2020

INTERVIEW QUESTIONS

Main research questions:
1. What type of unstructured data procurement has available?
2. What are the requirements and capabilities for procurement organization to utilize the available unstructured in analytics?
3. What is the expected value of available unstructured to organization?

9.1 Research question #1: What type of unstructured data procurement has available?

9.1.1 Unstructured data

1. Which sort of data assets are you currently utilizing? Name few, historical development of these, which is seen as most interesting?

2. Which type of unstructured data would you have available?
   a. Please provide examples and definitions

3. How do you see unstructured data in your environment?
   a. How would you define it?
   b. Which possibilities it might have?
   c. Which current business problems it could solve?

4. Are you managing the unstructured data as of now how?
   a. Is it in own environment? Or integrated with structured data?
   b. Are you able to locate the unstructured data?
   c. In which stage it is formed and received?
   d. Are you able to access the data, how about search within the data?
   e. Are you applying some methods to the data? Such as metadata, text analytics techniques
   f. Are you able to transfer the unstructured data into available format and link it to structured environment, how?

9.1.2 Data challenges

1. Which these challenges would you recognize currently to exist in your organization?
   - o Data/Procurement alignment with business strategy and clear understanding of the value proposal
   - o Defined data–driven decision–making process across organization
   - o Distribution of information
   - o Integration of the data
   - o Interoperability
   - o Variety and volume of the data
   - o Accessibility of the data
   - o Cost– and resource–intensity
   - o Searchability and terminology of the data
   - o Security and privacy of the data
   - o Scalability, development, and maintenance of data environments

- Would there be additional / which you would see as the most critical for your organization to solve and manage?

9.1.3 Data maturity and processes

5. Would you describe your data management processes to be: (can identify on many levels)
   a. Ad hoc and only few processes are defined. The success of projects depends on individual efforts.
   b. Repeatable, data management processes are established. Planning and managing data management processes and projects is based on experience of similar projects.
   c. Defined, data management processes are documented, standardized and integrated into a universal organization data management process
   d. Quantitatively managed, data management processes are predictable since they are measured and controlled within the measured limits.
   e. Optimizing, organization data management processes are continuously improved with feedback and piloting of new innovative ideas, processes, and technologies.

6. Could you please describe more detailed your current data capturing process, especially with unstructured data.
    a. Is it reliable, transparent, high-quality?

7. Could you describe your data manipulation process, meaning the management of datasets, giving meta-information to datasets, ensuring compliancy to standards.

8. Could you describe your data sharing processes? Can the data be accessed?

9. Could you describe your data storing processes? Are they secure?

9.2 Research question #2: What are the requirements and capabilities for procurement organization to utilize it in analytics?

9.2.1 Data strategy

10. Does your organization have a defined strategy for utilization of data in your business processes?
    a. Are new emerging technologies included, such as unstructured data?
    b. Which would you see as key topics with your data strategy?
    c. How this comes up on a) organization b) your workday level

11. Would you consider your data strategy to be aligned with the business strategy of your organization and how this shown (themes as example):
    a. Themes: Resource, processes, goals

12. Who from your organization is involved in the strategy, planning & implementation?
    a. Themes: Functions in organizations, and roles of people and entities
    b. Would you consider based on your experience that all stakeholders are suited with resource and capability?

13. What would you describe as vision and goal for the data strategy in your organization?

9.2.2   Data infrastructure

14. Could you please describe your current data infrastructure layers?

9.2.3   Capability analysis – Tangible resource

15. Refer to earlier discussion on the data challenges and evaluate your capability to manage the challenges and the data maturity.

16. Would you consider your organization information system (analytics tool for example) resource aligned across organization?

17. Who is responsible for the organization's information systems? Is this IT-department, or is the responsibility aligned across organization?

9.2.4   Capability analysis – Human resource

18. *Technical skills*
    a. Are you expecting to receive business value and competitive edge from analytics?
    b. Are you basing your decision available data, is analytics main decision-support tool?
    c. Does your organization provide analytics/digital tool trainings?

19. *Managerial skills*
    a. Who in your organization is involved in analytics processes?
    b. Is your management supporting in the use?

20. *Procurement skills* – Recognize your procurement organization on each maturity level:
    d. First stage
        i. Primary goal is to ensure the availability of goods and services, and that is the main measure of performance
        ii. No strategy applied
    e. Second stage
        i. You are seeking for best price, negotiations and strategy is part of process

       ii. Price and delivery key measures

      iii. Cost savings the driving factor

  f. Third stage

       i. Unified purchasing cross-organization

      ii. Price, cost and quality as measure

      iii. Supplier management important

      iv. IS in place, but not integrated across organization

  g. Fourth stage

       i. Strategically important, cross-functional purchasing

      ii. TCO of the system, not only cost of unit

      iii. IS integrated across organization

  h. Fifth step

       i. External integration

      ii. Supplier management and development is important topic

      iii. Efficiency with external parties, and internal learning

  i. Last step

       i. Value to end customer as key measure, not the internal cost/value but the external value.

9.2.5   *Capability analysis – Organizational resource*

21. Would you describe your organization as data-driven? What does this mean to you? Which ways this shows in your organization?

22. Would you describe your organization as capable to continuously learn? What does this mean to you? Which ways this shows in your organization?

23. Would you describe your organization has capability to turn available information into actions? Which ways this shows in your organization?

9.3   Research question 3: What is the expected value for their organization?

24. Could you please share what type of current business and analytics problem processes you would see unstructured data could solve that structured data is limited to solve?
    a. Explorative discussion
    b. Refer to previously discussed format of unstructured data and its possible value
    c. What is the value for having the unstructured data as available information for your decisions?

Information: Research has identified following value dimension to help assessing the expected:

25. How you are now measuring the performance of procurement?
    a. Which indicators would you recognize as key and most important for you?
        i. Research framework has identified: cost, risk, quality and value

    b. Which is the expected value you would expect to receive when utilizing the defined unstructured data in procurement analytics domain of yours?
        i. Explorative discussion.