



VARIETY AND SENSITIVITY OF DATA IN PUBLIC PROCUREMENT PROCESS

Lappeenranta–Lahti University of Technology LUT

Master's Programme in Supply Management, Business Administration, Master's thesis

2024

Kaisu Hilska

Examiners: Professor Katrina Lintukangas

Post-Doctoral Researcher Elina Karttunen

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT Business School

Business Administration

Kaisu Hilska

Variety and sensitivity of data in public procurement process

Master's thesis

2024

63 pages, 3 figures, 8 tables and 3 appendices

Examiners: Professor Katrina Lintukangas

Post-Doctoral Researcher Elina Karttunen

Keywords: public procurement, data management, data classification, data sensitivity, data variety, data analytics, data protection

Public sector organizations such as procurement units face a growing demand to both use data beneficially and protect collected data. Data usage and protection require understanding of what data organization has, which can be reached with data classification. This study aims to recognize what kind of data is generated through public procurement process. Secondary aims include categorizing how this data can be classified according to sensitivity and variety, and how do these classifications impact the protection and usability of data.

Empirical data for the study was collected through semi-structured interviews with 13 public procurement professionals. Even though there is mostly public and non-sensitive data in tender phase, pre-tender and post-tender phases include a lot of sensitive data that require more protection from unauthorized use. Pre-tender data is mostly unstructured, whereas tender data is semi-structured due to standardization. Post-tender data is largely structured, which makes it simple to analyze and use in monitoring activities. To improve the usability of data from pre-tender and tender phases, procurement units should be supported in transition from traditional structured data analytics to newer analytics tools and methods.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT-kauppakorkeakoulu

Kauppatieteet

Kaisu Hilska

Tiedon rakenne ja sensitiivisyys julkisessa hankintaprosessissa

Kauppatieteiden pro gradu -tutkielma

2024

63 sivua, 3 kuvaa, 8 taulukkoa ja 3 liitettä.

Tarkastaja(t): Professori Katrina Lintukangas

Tutkijatohtori Elina Karttunen

Avainsanat: julkinen hankinta, tiedonhallinta, tiedon luokittelu, tiedon sensitiivisyys, tiedon rakenne, data-analytiikka, tietosuoja

Julkisilla hankintayksiköillä on kasvava paine sekä hyödyntää tietoa päätöksenteossa että suojata kerättyä tietoa asianmukaisesti. Edellytyksenä tiedon suojaamiselle ja hyödyntämiselle on, että organisaatio ymmärtää millaista tietoa sillä on hallussaan. Tämä voidaan selvittää tiedon luokittelulla. Tutkimuksen tavoitteena on tunnistaa, millaista tietoa eli dataa julkisessa hankintaprosessissa syntyy. Toisena tavoitteena on luokitella, kuinka tätä tietoa voidaan luokitella sensitiivisyyden ja rakenteen mukaan, ja miten nämä luokitukset vaikuttavat tiedon suojaamiseen ja käytettävyyteen.

Empiirinen tutkimusaineisto kerättiin puolistrukturoitujen haastattelujen kautta. Vaikka tarjousvaiheen data on enimmäkseen julkista ja ei-arkaluonteista, valmisteluvaiheen ja sopimuskauden ostodata vaatii enemmän suojaustoimenpiteitä. Valmisteluvaiheen data on enimmäkseen strukturoimatonta, mikä vaatii kehittyneempää analytiikkaosaamista ja -työkaluja. Tarjousvaiheen data on standardisoinnin vuoksi semistrukturoitua. Sopimuskauden ostodata on pääosin strukturoitua, minkä ansiosta sitä on helppo analysoida. Valmisteluvaiheen ja tarjousvaiheen datan analysointimahdollisuuksien parantamiseksi hankintayksiköitä tulisi tukea investoimaan edistyneempiin analytiikkatyökaluihin.

Table of contents

Abstract

1	Introduction	1
1.1	Background of the study	2
1.2	Research questions and objectives	3
1.3	Theoretical framework	4
1.4	Defining key concepts	6
2	Data.....	8
2.1	Features of data and data as an asset	8
2.2	Data management and data security management	9
2.3	Data classification	11
2.4	Data sensitivity	14
2.5	Procurement data analytics	15
2.6	Data variety	18
3	Public procurement.....	20
3.1	Principles of public procurement	20
3.2	Thresholds and procurement procedures	21
3.3	Procurement process	23
3.3.1	Pre-tender	23
3.3.2	Tender	25
3.3.3	Post-tender	26
3.4	Public procurement process data.....	27
4	Research design	30
4.1	Research methodology and process	30
4.2	Case organization	31
4.2.1	Data management and classification in VTT	32
4.3	Data collection	34
4.4	Data analysis	37
4.5	Reliability and validity	38
5	Empirical findings	40

5.1	Pre-tender data	40
5.1.1	Pre-tender data by variety	42
5.1.2	Data classification of pre-tender	43
5.2	Tender data.....	44
5.2.1	Data classification in tender.....	46
5.2.2	Tender data by variety	48
5.3	Post-tender data.....	49
5.3.1	Data classification in post-tender.....	50
5.3.2	Data variety in post-tender.....	50
6	Discussion and conclusions.....	52
6.1	Answering to research questions.....	52
6.2	Implications for practice	61
6.3	Conclusions, limitations, and suggestions for further research.....	62
	References.....	64

Appendices

Appendix 1. The interview invite (translated from Finnish to English).

Appendix 2. Purchasing/post-tender process description for interview.

Appendix 3. Tendering/pre-tender and tender process description for interview.

1 Introduction

As technology is developing and use of digital devices is increasing, the rate of data generation has exponentially grown. Large amounts of data have become accessible for companies, and this data is becoming easier and less expensive to store (Pawar & Paluri 2022, 1) This development concerns all aspects of society, including public sector organizations where there is an increased pressure to both use data beneficially and protect collected data. Before being able to use data efficiently and securely, an organization must understand the data they have and what kind of properties their data has.

Brous, Janssen, & Vilminko-Heikkinen 2016, 115, state that public organizations generally store a lot of data, which theoretically should help organizations make better operational decisions. However, usage and processing of stored data seems to be lacking. As an example, according to Langseth & Haddara 2021, 248-249, only 17% of Norwegian procurement practitioners analyze their spend and 26,5% of them do not apply any form of data analytics in the planning of their purchases. However, even though there are possibilities with data, there are also risks. As cyber and information security are rising concern with accelerated pace of data generation, recognition and classification of sensitive data becomes increasingly important.

This study researches what kind of data is generated in public procurement process. It examines how data from public procurement process can be classified and how do these classifications impact the usability and protection of public procurement data. First part of this paper gives an overview to relevant literature to the topics of data and public procurement. Empirical research is conducted as a case study that includes qualitative data from 13 procurement professionals. As a result of this study, suggestions to improve data usability and protection based on data classification are presented.

1.1 Background of the study

To obtain value from data, it needs to be stored, assessed, and managed. This processing of data creates information that can be used to make rational operational decisions. (Gordon 2022, 9-10) Data management creates the policies and processes that both prepares the data for usage and ensures protection of data (Gordon 2022, 12; Mahanti 2021, 68). Benefits of good data management can include direct benefits such as financial benefits (saving costs and improving efficiency) and decreased risks (reducing privacy violations, civil and regulatory liabilities, increasing data privacy) (Brous et al. 2016, 116). Unfortunately, many organizations struggle to reap the benefits of available data as it continues to be an often-neglected resource (Gordon 2022, 14).

Public procurement is a process where public organizations and authorities purchase work, goods, or services from companies (European Commission 2022). Public organizations include government organizations, municipalities, and congregations. The economical role of public procurement is significant in the EU-area: the annual spend of public authorities contributes to approximately 14% of GDP (European Commission 2022). Public procurement also has an important role in promoting jobs, economic growth, investments, and innovation (European Commission 2022). As public procurement has big volumes, its improvements can lead to great monetary savings that could be largely supported by data-driven decision making.

As public institutions have become dependent of electronic platforms, cybersecurity risks have also emerged (Issabayeva, Yesseniyazova & Grega 2019, 62). Many public institutions have already been impacted by data breaches and other similar digital security hazards (van Ooijen, Ubaldi & Welby 2019, 44-45). The focus on data security also in public procurement has grown. This is due to increasing digitization of procurement processes and improved recognition of potential risks associated with handling sensitive information. (van Ooijen, Ubaldi & Welby 2019, 44-45)

Most academic research about data analytics in public procurement is focused on its possibilities in ensuring compliance and effectiveness from public's perspective. In the

context of public procurement, main objective of data analytics has been to prevent and reveal corruption or wasteful spending. (Mynarz 2014, 21; Tan & Lee 2015, 70) This can be done by monitoring systems that are utilizing data analytics techniques to detect anomalies in public procurement data (Hammoudi, Quix & Bernardino 2021, 115). The focus on reducing corruption and crime can be understood due to public procurement's economic impacts: in 2021, the annual public procurement spend in Finland was 35 billion euros (Työ- ja elinkeinoministeriö 2021, 4).

However, research on public procurement unit's independent and proactive work in developing data analytics and data management is limited. It has also been argued that public organizations such as public procurement units are lacking behind in implementation of analytics solutions compared to private companies (Merhi & Bregu 2020, 606). This study aims to address this gap by offering insights to the topic from a single procurement unit's perspective.

1.2 Research questions and objectives

The aim of this study is to recognize what kind of data is generated through public procurement process and how this data can be grouped into different categories. The study takes a process-focused view on procurement data and compares data classifications among different procurement process phases. In this study, procurement process includes both tendering and purchasing processes, so it covers the whole chain from pre-tendering planning phase to operational phase of sending a purchase order to supplier.

The chosen perspective is focused on one specific organization. The case organization is VTT Technical Research Centre of Finland that is a state-owned and non-profit company. As public procurement tendering process is regulated, this process is similar in other public organizations. However, operational purchasing process and its phases can vary a lot, which impacts how well study's results can be connected to other organizations.

Main research question is:

What data is collected and created in the public procurement process?

Sub-research questions are:

How can this data be classified according to variety and sensitivity?

How do these classifications impact the usage and protection of data?

1.3 Theoretical framework

The main concepts of theoretical framework include public procurement and data classification as an integral part of data management. The amount of data used and collected has exploded in organizations in recent years, which emphasizes the importance of managing data as any other asset. With developing technology and digitalization, it can be expected that the importance of data and its management will further increase in all industries. As public organizations and procurement units have a lot of data and they spend a lot of money, possible benefits from better data management are significant. (Brous et al. 2016, 115) For example, it has been estimated that developing public procurement processes could bring 10 % savings from current costs (Valtiovarainministeriö 2020, 50).

The role of data in public procurement has been recognized in the Finnish Ministry of Finance that published a report discussing the state of public procurement in 2020. The importance of data and information management is highlighted, but it is also recognized that usage of data is very limited in public procurement units (Valtiovarainministeriö 2020, 20). According to the publication, the procurement units should categorize procurement data assets logically. The goal is that data assets are automatically utilized in all stages of procurement process. (Valtiovarainministeriö 2020, 23, 30) This can be done by data analytics that brings knowledge to management to develop decision-making, such as helping keep track of spend per category, ranking suppliers, improving risk management, enhancing

contract management, and assisting in supplier evaluations (Shao, St. Louis, Corral & Li 2022, 132).

Both data security and data analytics require data classification, which can be seen in Figure 1. Addressing data security in public procurement units starts with recognizing which data regulation is applicable and proceeds with determining a data classification policy and classifying data assets according to sensitivity (Kim 2017, 1). Finding the sensitive data and directing most protective efforts to those data assets is considered the foundation of data security. Data classification is also necessary in data usage and analytics: to use data, organization must recognize the form or variety of data to find the right tools and techniques for analysis.

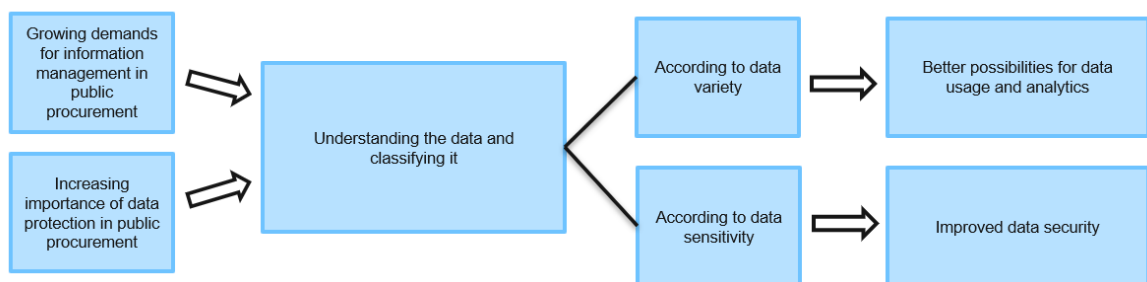


Figure 1. Theoretical framework

Still, it has been identified that many organizations do not have distinct data management policies, proper understanding of their data or defined data classifications. This confirms that data and factors related to it require much more attention among organizations. (Alhassan, Sammon & Daly 2018, 300-301; Thompson, Ravindran & Nicosia 2015, 317) As public organizations seem to be falling behind in data-related developments, it is safe to assume that this recognition is also applicable to public procurement units in Finland. This study aims to offer more insights into public procurement process data, how this data could be used better in the context of procurement analytics and how it should be protected according to sensitivity to ensure data security and compliance.

1.4 Defining key concepts

Data is representation of information or facts that has been collected to be used, stored, and examined. In modern world, data is generally stored in an electronic form, and it can be used and accessed by a computer or other electronic device. (Mahanti 2021, 8) Data can consist of numbers, letters, or other attributes, and it can be quantitative or qualitative. Data can only have a meaning if it is set in context, and this processing converts data into information. (Bhatia & Bansal 2015, 16)

Data classification is a way for an organization to organize their data into groups that have similar characteristics. This procedure is needed to ensure proper storage, management, and retrieval of data. Data can be classified by source, variety, entity, sensitivity, or some other way that is beneficial for the organization. (Mahanti 2021, 26, 28). In the context of this study, data is classified according to sensitivity and variety.

Data analytics is a process of collecting and analyzing data to extract value and significant insights. In practice, it can consist of identifying patterns between factors and noticing trends, which can be used in data-driven decision-making. (Langseth & Haddara 2021, 223).

Public procurement is a procedure where public organizations and authorities purchase work, goods, or services (European Commission 2022). Public procurement must follow a predetermined procurement process, excluding purchases that fall under a government-set threshold value. National threshold value for most goods and services is 60 000 euros, for construction work 150 000 euros, and 400 000 euros for social and health care services. Public procurement is regulated both on national (The Act on Public and Concession Contracts (1397/2016) and EU-level (Jääskeläinen & Tukiainen 2019, 5)

Procurement process refers to set of activities that are needed to ensure acquisition of needed goods, services, or other work. Traditionally, process starts from identification of need and setting specifications for a purchase and ends with follow-up and evaluation. (Nieminen 2016) In this study, procurement process is divided into three phases that are pre-tender, tender, and post-tender. Pre-tender phase starts from recognition of procurement need and

ends to. Tender phase covers the official public tendering process and ends to finalization of procurement contract. Post-tender phase includes all operational purchasing activities that take place after contracting

2 Data

This chapter presents the theory linked to data, data management and data classifications. First part covers why data should be treated as an asset, how it differs from traditional fixed assets, and how these differences impact its management. Second part includes an introduction to data management and data security management. Next, data classification as part of data security and data sensitivity are discussed. Last chapters cover procurement data analytics and data variety in analytics.

2.1 Features of data and data as an asset

According to Gordon 2022, 4, a common definition of information is that it is data placed in context. This definition states that data must be processed to become information. Bhatia & Bansal 2015, 16-17, state that data in itself has little value if it is not processed. Data can consist of numbers, letters or other attributes, but without context and organizing, it cannot be understood or used. Mahanti 2021, 8, states that data is both an interpretation of the objects it presents and an object that should be interpreted itself.

According to Al-Ruithe, Benkhelifa & Hameed (2019, 839), historically data has been seen as a by-product of business, which is why data has not been managed as an asset. This tendency lasted till 1990s, when the role of data started to expand, and many business decisions started to be driven by data. However, while financial and physical capital and assets have been and are valued and generally managed, data is often still not treated as a similarly valuable resource.

One possible reason for not managing data and neglecting its value is that it is difficult to place a specific financial value to data. Data is rarely used to its full potential, and it can also potentially be seen as another additional expense rather than an asset by the senior management. (Mahanti 2021, 2) According to Gordon 2022, today's information and digital data's connection to technology can also lead to data being seen as something mystical and overly complex. However, information or data is always needed to run a business, and it is

a resource that should be the most available for management as it supports strategic business decisions, day-to-day operations, and is generally required make appropriate plans.

As an asset, data is very different from fixed ones such as factories, machines, or money. It is intangible and non-fungible, meaning that data is not physical, and it cannot be replaced with another piece of data, whereas a fixed asset like a broken machine could be replaced with another one. Data is also accumulative and non-rivalrous, so it can be connected to other data and used by multiple parties at the same time, which is generally not possible with fixed assets. Data is not consumed by use, it can be transformed, replicated, or deleted easily. (Mahanti 2021, 15-16) These qualities make data a more complex and enterprise-wide asset compared to fixed assets, which is why traditional ways of managing it do not work.

2.2 Data management and data security management

As data is an asset and a valuable part of organization's success, it should be properly managed. Data management can be defined as a practice of capturing, storing, organizing, and governing of data across an organization. In addition to managing the data and being connected to most data-related activities, it is an umbrella term that has multiple sub-disciplines such as data architecture, data quality and data security management. (Mahanti 2021, 2, 53-54) Data management function is involved with setting data definitions, enabling data sharing and access across organization, and coordinating many data-related activities. It can have an operational and a strategic role as it both educates stakeholders about common data issues and offers insight about data to upper management. (Gordon 2022, 71,74)

One of the main sub-disciplines of data management discipline is data governance that ties all sub-disciplines together. Data governance is not just technology: it is a business initiative that established the guidelines, rules, principles, and responsibilities related to data. Even though data governance and data management are closely related, they are not the same. (Mahanti 2021, 2, 53-55) Data management is focused on specifying data elements, how they are stored and structured and moved, and it oversees all activities and practices related to data. Data governance can be characterized as high-level organization and control of data management, and the aim of it is to set policies and practices that ensure effective data

management, but data governance does not have an official definition. As the definitions of data management and data governance can change by user, there is a lot of confusion between these terms, and even some researchers use them interchangeably. (Al-Ruithe et al. 2019, 841)

Good data governance requires the participation and dedication of the whole organization. It demands expertise and input from different departments to form a clear and organized framework that considers all different aspects of organization's data. Example benefits of good data governance include increased understanding of organization's data and confidence in its quality, improved data-based decision-making, reduction of uncertainty and ineffectiveness in operations, and better protection of data. (Al-Ruithe et al. 2019, 846)

Data security management is a data management function that deals with protection of data and data assets throughout their lifecycle. It consists of policies, rules, processes, and other practices that protect data by providing necessary authentication, authorization, access, and auditing of data, whereas data governance ensures that data security policies are implemented and enforced. (Mahanti 2021, 114-115) Data security management is connected to information security management that is one part of the organizational security management and a part of organizational management system.

Compared to data security management, information security has a broader scope concerning organization's information systems and infrastructure, such as protecting organization from cyberattacks or compliance with relevant regulations. (Ključnikov, Mura & Sklenár 2019, 2082-2083) However, data security and information security are sometimes used as interchangeable terms. In the context of this paper, it can be summarized that information security provides the foundation for data security, and data security is one part of information security.

Data security is a high priority for all organizations as the potential impacts of not securing data are too severe. Data breaches – incidents where data is disclosed, accessed, or stolen by an unauthorized party – have increased in the last ten years, which has caused individuals to be increasingly worried about the safety of their personal data. From business perspective,

an ordinary cost of a data breach can be counted in millions of dollars, and it takes on average 280 days to identify and contain it. (Bradford, Taylor, Seymore 2022, 201; Mahanti 2021,113-114) In addition to direct costs related to failed data security, data breaches and violations have a negative reputational impact to any organization, which can lead to loss of business and trust from stakeholders (Mahanti 2021, 66).

2.3 Data classification

The first step of data security management is data classification. This is a process that puts different labels on data assets. In the context of security, data classification places data and information into categories related to how these should be handled when it comes to confidentiality. (Axelrod, Bayuk & Schutzer 2009, 60) Classification class is based on an assessment of the impact on the organization if the data is disclosed, altered, or destroyed without authorization (Agrawal 2017, 265). Even though the importance of data classification is widely recognized, it is an understudied area. Most research on the topic is focused on its technological aspects and automated classification systems: especially practical studies on how organizations actually classify and recognize sensitive data is limited. (Bergquist, Tinet & Gao 2022, 154)

Data classification offers visibility to the amount of sensitive information that should have the most protective measures from unauthorized use (Bradford et al. 2022, 201). It assists in the implementation and determination of suitable security controls. Possible controls for data security include practices such as encryption, network access control, and role-based access controls. (Tankard 2015, 9) Data classification is mandatory activity for government agencies in some countries, but it is also endorsed for private organizations by different information security standards such as ISO/IEC 27000 (Bergquist et al. 2022, 155).

Classification of data plays a central role in risk management, helps with compliance and regulatory mandates, and offers a competitive advantage over organizations that do not classify their data. Increased employee understanding regarding the need for data security is another benefit of data classification. As the reason for most data breaches is a staff-related error, this can be very useful in preventing data-related leaks. (Al-Ruithe et al. 2019, 850;

Bradford et al. 2022, 201-203; Tankard 2015, 1, 4) However, in public sector, there is a tendency to “over-classify” data, which can contribute to increased costs in the field of data protection and classification (Gleeson & Walden 2016, 687).

In practice, data classification is often neglected. Data classification is a fairly new discipline, so the importance might not be clear for everyone. (Everett 2011, 5) In addition to having little academic literature and research in the area, there is also little detailed assistance on how to do it in practical guides, even though there are many professional frameworks and standards that mention data classification (Bradford et al. 2022, 203). Textbooks about the topic are generally directed at security professionals, but they seem to forget the end-users of data, subject-matter experts, and labelers of data. Collaboration that requires shared understanding of the used terms between labeler and security professionals is needed for successful data classification, but this is often overlooked in the literature. (Axelrod et al. 2009, 63)

Another reason for dismissing data classification is the growing complexity of the IT landscape. Information is increasingly stored and processed in multiple devices, locations, and programs, which makes it challenging to locate, identify and classify all data correctly. Lack of resources is another possible reason for overlooking data classification as it requires a lot of time to create, implement and maintain data classification policies. (Bradford et al. 2022, 210-212)

In addition to being time-consuming, data classification is sometimes considered to be tedious by people. So far, automation has failed to provide sufficiently competent tools for data classification, so it cannot be completely trusted in the process. (Everett 2011, 5) Automated classification is especially unreliable with unstructured data, and on top of this problem, it is expensive (Bradford et al. 2022, 211). It should also be repeated that many issues related to data security are often human-related, which means that technical solutions do not get to the root of the problem (Everett 2011, 5). Increasing awareness with data classification and security trainings does provide some positive results in decreasing non-compliance, but this requires a change of culture and organizational outlook to data security and classification (Bradford et al. 2022, 211-212).

Data classification process starts with establishing a data classification policy that includes a data classification framework, its scope and principles, descriptions of data classification levels, roles, and responsibilities for data identification. Data classification policy should also include a description of its purpose: what is the organization trying to achieve by implementing a data classification policy? (Bergquist et al. 2022, 156; Mahanti 2021, 116)

After data classification policy has been established, organization must decide if only new data will be classified or will all existing data be classified. If existing data contains sensitive data, it needs to be located, and a data discovery process is required. This is both expensive and time consuming, which is why the benefits and risks should be assessed before starting the exercise. (Mahanti 2021, 116)

At this point it is also necessary to identify the roles and responsibilities of data classification policy (Mahanti 2021, 123). This refers to roles such as data owner and data steward that are sometimes used interchangeably. Roles and responsibilities can also mean naming designated people to committees or project workshops related to data governance and security. Generally, data owner is a senior executive or manager who has authority and is responsible for the quality of data. (Alhassan, Sammon & Daly 2019, 106; Bradford et al. 2022, 207) Data steward could potentially be anybody in the organization but is usually someone who is a subject-matter expert, is involved with everyday operations around the data and has good understanding of the characteristics of data in said area (Seiner 2014, 68).

After these steps, data should be classified according to determined rules and policy (Mahanti 2021, 123). This action is also called labelling. This should be done by authorized individuals that are subject-matter experts or data stewards in different functions and understand the content and context of data assets. Information security professionals should concentrate in ensuring technical protection of data, not labelling the data unless they are experts in the specific area. (Axelrod et al. 2009, 63) However, close collaboration between labelers and information security is needed to ensure smooth flow of this process (Bradford et al. 2022, 207).

Once data has been classified and risks associated with it have been recognized, organization can review if all data is protected appropriately or if changes should be made, whether it is to strengthen or loosen protection in specific data assets. Organization should also revisit the classifications regularly to assess if they are still valid or if they should be changed. This can happen due to regulatory changes or contractual obligations but also because of changes in use of data. (Mahanti 2021, 124-125)

2.4 Data sensitivity

Organizations can have their own specific models to categorize security levels of their information, which impacts to available classification categories. Data classification model often includes different evaluations according to decided principles. Commonly used principles consist of confidentiality, integrity, and availability. Most popular principle is confidentiality as it is the easiest to comprehend and understand by data labelers. (Bergquist et al. 2022, 155) Confidentiality refers to protection of data from unauthorized access and privacy of data (Bradford et al. 2022, 203). In this case, data is classified according to confidentiality and risks of disclosure, which is a usual practice for public sector organizations (Gleeson & Walden 2016, 687).

Data groups related to data confidentiality and sensitivity can include categories such as public, private or internal, confidential, regulatory, restricted, or top secret. It is frequent to provide classification categories with additional details on how this data category can be described. (Mahanti 2021, 35; Shaikh & Sasikumar 2015, 494) Many organizations have three classification categories: one for public data, another for internal data, and third for most sensitive data. However, there is no definitive answer to how many classification categories there should be, and two or four data classification categories are not rare. Number of classifications is always a balancing act between having enough to manage data sufficiently and making classifications easy enough to comprehend for everyone in the organization. (Bradford et al. 2022, 207-208)

Data can be classified as restricted if unauthorized actions to it cause a very high risk, whether it is to an organization as compliance or reputation. It is very sensitive and requires

the highest level of privacy and protection, meaning that restricted data is not available for everyone in the organization. This data can contain usernames, passwords, or personally identifiable information such as social security numbers. (Mahanti 2021, 37-38). Trade secrets and other crucial business data is also in this category of sensitivity. Restricted data should be used and shared only when needed as breach of this data would be disastrous for an organization. (Bradford et al. 2022, 208)

Confidential data is moderately sensitive, but the level of harm caused due to misuse is less severe and the protection is not as high as with restricted data. As examples, this kind of data can include home address or religious orientation. (Mahanti 2021, 37-38). Confidential data can be related to specific team where this data is processed (Bradford et al. 2022, 208).

Private or internal data is moderate-level risk, and it is often the default assumption of data security classification if not otherwise specified. As name suggests, private or internal data is only for internal use. (Mahanti 2021, 38-39). Examples of this data can include employee handbooks, company policies, and company-wide memos (Bradford et al. 2022, 208).

Public data is the least sensitive, and this data can be found relatively easily by anyone (Mahanti 2021, 38-39). In some cases, this information can be found from organization's website. Examples of this kind of data include published financial statements, company background or locations. (Bradford et al. 2022, 208)

2.5 Procurement data analytics

Procurement data analytics can be defined as a process of collecting and analyzing procurement data to extract value and significant insights from it (Langseth & Haddara 2021, 223). Procurement data analytics offers a data-driven approach to combat different procurement-related problems. Traditional data analytics operates on structured data that has been categorized into variables and parameters. Usual tools of data analytics consist of statistics, graphical visualization tools, simulation, and mathematical algorithms. Typical topics of procurement data analytics include purchasing categories, spend and costs, risk

management, and supplier management and evaluations. (Handfield, Jeong & Choi 2019, 974)

There are multiple types of data analytics that can be used for different purposes. The starting point and foundation for data analytics is descriptive analytics. It is the simplest data analytics type that states what has happened in the past or what is happening at the moment. Data source is usually historical business transaction data, and descriptive analytics present the key aspects of this data. (Mahanti 2021, 134). Commonly used methods for descriptive analytics include simple statistical techniques and applications that describe the contents of data in a database. Presentation of descriptive analytics can include dashboards, pivot tables, and other visualizations of data such as charts. (Sedkaoui 2018, 45; Shao et al. 2022, 134) Descriptive analytics in procurement can be used in spend analysis and category management (Wang, Gunasekaran, Ngai, & Papadopoulos 2016, 103).

Predictive analytics is the next level of data analytics, and it uses data to assess what is going to happen in the future (Shao et al. 2022, 134; Sedkaoui 2018, 45). Compared to descriptive analytics, it uses more advanced statistical techniques and machine learning algorithms, and the goal is to discover predictive variables to build models that identify trends and relationships between different factors (Mahanti 2021, 135; Sedkaoui 2018, 45). Sources of data can include historical and transactional procurement data, contract data, market data, financial data and/or operational data. In the context of procurement, for instance, predictive analytics can be used to predict supplier's future behaviors based on previous actions. This can be helpful in reducing inventory or predicting supply chain flows by demand planning and forecasting. (Shao et al. 2022, 134; Souza 2014, 597)

Prescriptive analytics is the third level of data analytics. It uses data to predict what is going to happen and gives guidance on what is the optimal response in this situation, so it is more context- and action-oriented. It can also provide explanations on why this scenario is going to happen by revealing cause-and-effect relationships. (Mahanti 2021, 135; Shao et al. 2022, 135) Prescriptive analytics both merge data from descriptive and predictive analytics and use real time-external data that can include management and decision science. Prescriptive analytics can help to allocate organization's resources optimally, but it requires techniques

such as smart experiments and simulation algorithms. In procurement, it can help to optimize sourcing strategies and inventory. (Sedkaoui 2018, 45)

Cognitive analytics refers to use of advanced analytics methods such as machine learning, artificial intelligence, or robotics. By these methods, it incorporates human-like thinking in decision-making, which decreases or even removes the need for actual human involvement. (Shao et al. 2022, 135) Cognitive analytics can be summarized as developed data processing that focuses on identifying and understanding diverse, varying, complex and qualitative data. In addition to processing information, cognitive analytics can understand fundamental ideas and concepts, and it can further develop with time as it handles more data. (Handfield et al. 2019, 974)

Overall, to use procurement data analytics, organizations need to have both sufficient data management and data analytics capabilities (Rafati & Poels 2015, 30). An organization cannot analyze data that it does not have, and organization must have both technological capabilities and personnel understanding about data analytics to be able to utilize data. Also, the quality of analytics is always related to the quality of data: if the quality of data is low, the results are also untrustworthy. (Shao et al. 2022, 135-136)

It can be summarized that the first step of any form of data analytics is ensuring solid processes for data collection (Shao et al. 2022, 131). There are multiple issues that organizations face in this matter. One of common challenge in data collection is that procurement data can be located in multiple information systems as most organizations have separate systems in different organizational layers. In this scenario, it is challenging to comprise a unified view of what data organization has. In public procurement, tendering process, contract execution and contract monitoring can be managed by other units, which may lead to a situation where information is located in different systems and units do not know what data others have. Another problem is related to procurement processes themselves: some process phases might be handled outside of applications, which leads to a situation where electronic data is not available for analysis. This can happen if organizations' current applications and information systems do not adequately support procurement process. (Cocciolo et al. 2023, 272; Rafati & Poels 2015, 31-32)

2.6 Data variety

One way to categorize data is by variety. This refers to type of data by form. This group consists of three types that are structured, semi-structured and unstructured data. (Mahanti 2021, 31) Even though pros and cons of data varieties are often assessed, it should be stated that different data varieties mix and complement each other in providing sufficient insights (Harriott & Isson 2012, 363). A diverse set of data can offer a more holistic and complete view that utilizes each data varieties strengths and decreases their limitations.

Structured data is a defined data type, and it can be stored in rows and columns in a database. It is in predictable in form, so it has a standardized format and follows a pattern. This kind of data can consist of dates, numbers, and short character strings. (Mahanti 2021, 31; Gordon 2022) It generally does not require a deep level of expertise to comprehend or analyze: it can be used easily by statistical analysis or other traditional methods. However, it does have limited usage as it is not very flexible, thus it can offer limited insights. (IBM 2021)

For many organizations, most generated data is unstructured, and the role and importance of this data is expected to increase (Salo 2013). Unstructured data does not have a detectable structure or predefined data model. It cannot be stored in rows or columns in a traditional database, so it stored in non-relational data base or data lake. This makes unstructured data difficult to search and manage. As an example, unstructured data can be text in a document, an image or audio. (Mahanti 2021, 31; Gordon 2022) Compared to structured data, unstructured data can be described as more qualitative data that requires more processing. Unstructured data cannot be processed and analyzed with traditional analysis tools and methods, so it requires preprocessing, advanced tools and expertise in data science. Possible approaches to analyze unstructured data can include techniques such as natural language processing (NLP) or text mining. Even though it requires more effort in analysis, unstructured data can be very adaptable and used for many different purposes. (IBM 2021; Harriott & Isson 2012, 361)

Data variety can sometimes be challenging to assess especially as all data variety is evaluated in a continuum (Salo 2013, 25). Semi-structured data is located between structured and

unstructured data. It does not exhibit the same degree of predictability as structured data, but it generally has some recognizable patterns or tags, which makes it easier to analyze compared to unstructured data. (Mahanti 2021, 31; Gordon 2022) Semi-structured data can include information that is stored in a document that uses metadata such as tags or semantic markers. These identifiers make it possible to place semi-structured data into records and specific fields, which makes it easier to catalogue, search and analyze compared to unstructured data. (Mahanti 2021, 159; IBM 2021) However, there can be a lot of variation between different semi-structured data assets. Issues also arise when some part of semi-structured data becomes crucial after its metadata has been created and needed data has not been tagged. (Salo 2013, 25)

3 Public procurement

In this section, public procurement is presented. First, general principles of public procurement are introduced. Second part explains public procurement thresholds and procurement procedures. Third part of this section discusses open public procurement process, and it has been divided into pre-tender, tender, and post-tender phases. Fourth and final part describes procurement process data.

3.1 Principles of public procurement

Public procurement has four main principles that impact the procurement procedure. These include transparency, non-discrimination, transparency, and relativity. The principles have been stated in The Act on Public and Concession Contracts (1397/2016), and these must be followed at every stage of the procurement process. (Julkisten hankintojen neuvontayksikkö 2016a)

Transparency means that procurement process must be appropriately transparent, procurement unit's actions are predictable, and sufficient and clear information is provided to tenderers. This principle is very present in the demand to publish invitation to tender in an official manner. (Pekkala et al. 2022, 39-40) Procurement process must be public, information regarding procurement procedure cannot be withheld, tenderers are informed about the results of tender competition, and documents regarding procurement are generally available for the public. (Julkisten hankintojen neuvontayksikkö 2016a) Invitations to tender have to be informed publicly (Ministry of Economic Affairs and Employment of Finland 2021c), and also public contracts have to be sufficiently publicized once they have been established. (Ministry of Economic Affairs and Employment of Finland 2021a).

Equality means that invitation to tender other procurement documents and/or criteria cannot be set in a manner that puts tenderers in an unequal position. The principle of equality requires that same criteria apply to all tenderers and similar cases have to be assessed alike unless there is an objectively acceptable reason to do otherwise. (Julkisten hankintojen

neuvontayksikkö 2016a) If a tenderer is excluded from procurement process, it must be based on clearly indicated criteria that is available for all.

Non-discrimination is related to principle of equality. It aims to ensure that suppliers have equal possibilities to participate in a tendering competition and they are treated equally with another. (Pekkala et al. 2022, 41) Procurement decision must exclude all assessment factors that are not related to procurement in question. These factors can include nationality or location of tenderers. (Julkisten hankintojen neuvontayksikkö 2016a) Contracts must be awarded based on criteria that have been stated in the invitation to tender, and these criteria cannot be changed later in the process. (Ministry of Economic Affairs and Employment of Finland 2021a) Principle of non-discrimination ensures that the competition in the markets is used optimally, and that the winner of tendering competition has the most economically advantageous offer for the procurement unit (Pekkala et al. 2022, 41).

Relativity means that requirements of the acquisition must be in proportionate in relation to the character, objective, and value of the procurement. Even though all selection criteria set some limits and constraints, procurement unit should be able to give valid reasons why determined qualifications have been set for a specific procurement process and understand how these qualifications potentially impact the markets. The aim of relativity is to ensure that public procurement does not exclude any potential tenderer without a justifiable reason. (Pekkala et al. 2022, 42) For example, it is not acceptable to have as strict qualitative requirement criteria for a high-value expert service and low-value office supply procurement contract. In addition to qualifications, relativity must be considered during other phases as well, for example when deciding the procurement procedure and preparing invitations for tenders. (Julkisten hankintojen neuvontayksikkö 2016a).

3.2 Thresholds and procurement procedures

Public procurement must be conducted by following a pre-determined process (Kontio, Kronström, Kumlin & Mäki 2017, 105). One factor that impacts public procurement process is threshold values, which means that different procurement procedures can be separated into three categories according to value (Kontio et al. 2017, 69-71).

Small-scale procurement includes purchases that fall under the national threshold value. In these cases, Act on Public Procurement and Concession Contracts does not apply, and these purchases do not have to be tendered publicly. Procurement units have their own principles and guidelines regarding small-scale procurements, but as there are no specific regulatory rules on how these procurements should be organized, there is less bureaucracy and notification obligations. However, principles of public procurement still apply to small-scale procurements. (Julkisten hankintojen neuvontayksikkö 2022b; Kontio et al. 2017, 71) It should be noted that even though it is not mandatory to use official tendering procedure in small-scale procurement, procurement unit can voluntarily follow the official process as well.

National procurement contains procurements that exceed national threshold value but fall under EU-threshold value. Threshold values vary according to type of procurement. (Kontio et al. 2017, 71) In goods and services, national threshold value is 60 000 €. Threshold value is 150 000 € for building contracts and 400 000 € for health and social services. In these cases, procurement unit publishes an invitation to tender in official national channel. In Finland, Hilma is the official electronic channel for public procurement notifications. (Julkisten hankintojen neuvontayksikkö 2022a) After releasing the official invitation to tender in Hilma, procurement unit follows an official public procurement process according to legislation.

EU-procurement applies to procurements that exceed EU-threshold values. As with national thresholds, the levels differ between procurement categories. (Kontio et al. 2017, 70) EU-thresholds vary between central government authority and other contracting authorities that include government organizations, municipalities, and congregations. For other contracting authorities, value is 221 000 € for supply and service contracts and 5 538 000 € for building and public works contracts. (Julkisten hankintojen neuvontayksikkö 2022b) EU-level procurements are published in Hilma and Tenders Electronic Daily (TED) that is an EU-wide information channel. Both invitations to tenders and contract awards are published in TED. (Duguay, Rauter & Samuels 2023, 1160; Julkisten hankintojen neuvontayksikkö 2022a.)

3.3 Procurement process

There are multiple ways to group different process phases of public procurement. According to Holma, Vesalainen, Söderman & Sammalmaa (2019, 1), public procurement is generally divided into three different main phases that are pre-tender, tender, and post-tender. Pekkala et al. (2022, 21-27) have taken a more detailed approach and are able to recognize fourteen different process phases. However, all these fourteen phases can also be grouped into pre-tender, tender- and post-tender.

In national procurements, it is possible to use multiple different public procurement procedures. Procurement unit must clearly state which procedure is used in invitation to tender. The most used procedure is open procedure where invitation to tender with other tender materials is published in Hilma, and all companies can participate in bidding. (Kontio et al. 2017, 106-107; Heijboer & Telgen 2002) In this study, procurement process follows open procedure of national public procurement. The process has been divided into three main phases that are pre-tender, tender, and post-tender, which can be seen from Figure 2.

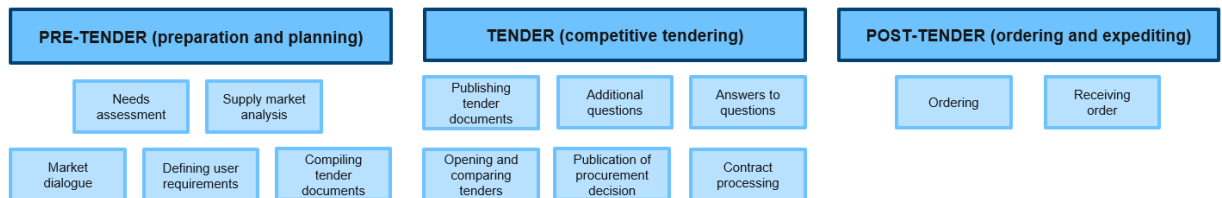


Figure 2. Public procurement process divided into main phases.

3.3.1 Pre-tender

Overall, pre-tendering stages set the groundwork for successful public procurement project. Pre-tender of public procurement includes all planning and preparing tasks for procurement project. These activities can include needs assessment, defining user requirements, looking into the supply markets for available options, and preparing and processing tender documents. (Holma et al. 2020, 2) Deciding the most suitable procurement procedure is also part of pre-tender (Kontio et al. 2017, 90). Pre-tendering phase is subjected to less regulation

compared to later stages of public procurement, which enables a bit more flexibility in the pre-tendering process and collaboration with potential suppliers (McKevitt & Davis 2015, 79).

Pre-tender phase starts with internal planning that includes defining specifications of what the organization wants to buy (Van Weele 2014, 387). This phase requires dialogue between procurement unit and internal customer that is requesting the purchase. Internal customer can be defined as an individual who will use what procurement unit buys. Procurement professionals have to understand the object of the purchase to formulate a clear invitation to tender and to be able to carry out the tendering process according to regulation. (Holma et al. 2020, 4; Kontio et al. 2017, 93-94)

After internally setting initial goals for the procurement, it is possible for procurement organization to engage in market analysis and dialogue with potential tenderers. These actions aim to increase buyer's knowledge and understanding of the markets, which helps to set appropriate specifications for the official invitation to tender. Market dialogue can include discussions about possible technical properties of the wished product. (Holma et al. 2020, 4) Through this phase, procurement unit discovers what kind of options are available in the market and what kind of properties and specifications should be included in the invitation to tender (Pekkala et al. 2022, 21). Without market analysis and dialogue, procurement unit carries a bigger risk of releasing an invitation to tender that does not sufficiently attract potential suppliers.

Market dialogue can be done by collecting brochures from potential tenderers, using expert services, visiting events related to the field, or some other way. It is also possible to publish a request for information in Hilma and directly invite potentially interested tenderers to share information about their products, services, prices, and other related topics. (Kontio et al. 2017, 92) These responses are non-binding and preliminary, but they help procurement unit to assess how many options there are and what kind of pricing is realistic in the markets (Kontio et al. 2017, 93).

Preparing materials for invitation to tender is the next phase of tendering process, main document being invitation to tender. This includes specific criteria for the product or service that procurement unit wants to purchase. Invitation to tender must also include specifications on how these criteria will be measured and how much they impact the procurement decision. (Pekkala et al. 2022, 21) For public procurement unit, selection criteria can be grouped into exclusion criteria and suitability criteria. Exclusion criteria are further grouped into mandatory (e.g. participating in criminal organization) and optional (e.g. guilty of professional misconduct) grounds for exclusion. (Van Weele 2014, 389)

3.3.2 Tender

Official tendering competition begins with publishing the invitation to tender. Procurement unit has to publish specifications and other attached documents for tenderers in the invitation to tender. Other documents can include drawings, technical descriptions, or calculations. Contract terms or a contract draft are common attachments in the invitation to tender. (Pekkala et al. 2022, 345). Invitation to tender includes the main content and specifications of the procurement object. It is generally not acceptable to amend invitation to tender after publishing, and contract award decision must be done according to the specifications that have been released to the public. (Pekkala et al. 2022, 22)

Invitation to tender and other related information are published in Hilma that is an official and electronic channel for public procurement (Pekkala et al. 2022, 332). In the tendering phase, communication between procurement unit and tenderers is stricter and more procedural compared to pre-tender. The aim of this approach is to ensure that principles of public procurement such as transparency and equality will hold. (McKevitt & Davis 2015, 80) However, procurement unit can give some additional information regarding the invitation to tender after publishing it if potential tenderers want to ask additional questions about the procurement. This can be done if invitation to tender is unclear or somehow insufficient. Still, this information cannot intrinsically impact the terms of procurement, and all additional information must be released for all potential tenderers. (Kontio et al. 2017, 93)

All received tenders are opened at the same time after the deadline. This event is generally not public, but it is recommended that procurement unit should document this. Minutes about this event should include time, place, participants, number of received bids, and a list of all bidders. (Kontio et al. 2017, 237) Tenders cannot be amended or supplemented after they have been submitted, but procurement unit can ask additional questions from a tenderer if the received bid is unclear in some parts. However, this is procurement unit's right and not an obligation, and answers to questions cannot impact negatively other tenderer's position in competition. (Kontio et al. 2017, 240)

Before comparing tenders, procurement unit needs to check if all tenders fulfil the minimum criteria and/or exclusion criteria. If the tenderer does not pass this phase, the tender itself is rejected from further assessment as procurement unit can only compare tenderers that fill the minimum criteria. (Pekkala et al. 2022, 22) Tenders that have not been excluded will be evaluated according to comparison criteria. Procurement unit can only use criteria that have been included in invitation to tender. (Kontio et al. 2017, 251) After comparing these tenders according to published criteria, procurement unit makes the procurement decision. Procurement unit publicly informs about the initial award of the bid and informs non-selected tenderers about the decision, including the winner and why the contract has been awarded to selected party. (Kontio et al. 2017, 257-258)

Non-selected tenderers can complain about the procurement decision during the waiting period, and the instructions for this are included in the procurement decision. If there are no complaints, selected supplier and procurement unit must abstain from signing a contract until waiting period has passed. After waiting period, finalization of the contract takes place. (Van Weele 2014, 389-390; Pekkala et al. 2022, 22)

3.3.3 Post-tender

Traditional tasks related operational purchasing are identification of need, selecting a supplier, contracting, and ordering (Nieminen 2016). In the context of public procurement, most of these tasks have been done during official tendering process. Only ordering phase is left and post-tender process includes the same ordering steps as other purchasing functions.

When terms and conditions have been agreed on, ordering phase can take place. Purchaser sends an official purchase order with specifications (what and when) to supplier. Suppliers usually send back an order confirmation that confirms the terms and delivery time for the purchase order. (Van Weele 2014, 42; Nieminen 2016)

Expediting means securing that ordered products and services are delivered according to what has been agreed. If all goes well, there is little to do for purchasing after ordering and receiving an order confirmation. However, situations change, and expediting phase can require a lot of time and focus from purchasing department. Possible cases include exception expediting (internal customer informs that product or service has not arrived in time), routine status check (buyer contacts supplier in advance to confirm delivery time), and advanced status check (buyer checks delivery time and progress regularly with supplier). (Nieminen 2016; Van Weele 2014, 43)

Follow-up and evaluation is a more systematic process that aims to improve purchasing operations. These are generally not related to specific purchases but the whole collaboration with suppliers. Measuring supplier's performance can be approached from multiple points of view, whether it is purchasing costs, quality, or delivery accuracy. (Nieminen 2016)

3.4 Public procurement process data

Procurement process generates a lot of data, and most of it is stored in databases. Generally, a lot of data that is used in data analytics is transactional such as purchase orders and invoices. (Chae & Olson 2013, 11-13) For example, analyzing spend per category is often based on purchase order and/or invoice data. In addition to prices, purchase orders and invoices include information about suppliers, delivery times, payment terms, delivery terms, and shipping costs. These data sets can be connected to common and traditional types of procurement data analytics such as spend analysis and forecasting, supplier performance monitoring, and purchase order analysis. These can be characterized as descriptive analytics that aim to identify problems in existing processes. Other sources of data from procurement process are negotiations, quotations, contracts, and supplier assessments. (Wang,

Gunasekaran, Ngai & Papadopoulos 2016, 103) It can be assumed that the variety of this data is mostly unstructured as it is text-heavy and does not follow a specific pattern.

Public procurement typically follows a pre-determined process. These phases generate data on tendering process, bidding process, bid evaluation, contract awards, and contracting. All these phases generate some form of data, which can be seen from Figure 3. Pre-tender data includes data on planning, budgeting, and tender preparation (Cocciolo et al. 2023, 273). Planning includes planned purchases and yearly budget allocations, and analyzing this data is useful in assessing the quality of planning (World Bank 2022, 7). Tender preparation in the form of market analysis and dialogue can be compared to negotiations and quotations in the private sector, so these phases generate data about suppliers, products and services, prices, and contract terms in the field. However, it has been established that usage of planning, budgeting and tender preparation data is lacking in public procurement analytics compared to traditional forms of analytics such as analyzing spend and prices (Cocciolo et al. 2023, 273).

Tender data is easier to understand as invitation to tender and other official tendering documents are available for everyone and are the foundation of the process. The structure of documentation includes quantities of items, section names, conditions, and specifications. Even though the content is similar in most cases due to regulation and available ready-made forms, procurement unit practically chooses how tender documents are organized and worded. (Cocciolo et al. 2023, 273; Hammoudi et al. 2021, 120-125) It should be noted that tendering process generates also internal data that is not publicly available, and assessing this data externally is not possible.

Public tender data can be utilized to assess effectiveness of procurement procedure: this can be done by analyzing tender processing time, price, and contract implementation. Transparency and competition are also important dimensions in public procurement, which can be measured by assessing process and contracting data. (World Bank 2022, 7) Even though these are relevant aspects to assess, it has been recognized that public procurement data analytics could be developed to incorporate more than these traditional indicators that focus on immediate monetary impact. For example, sustainability, quality, and different

social and environmental objectives are increasingly important for governments and public organizations. This development leads to question if current indicators are sufficient or if they should be further developed. (OECD 2019, 15; Cocciolo et al. 2023, 260)

Post-tender data consists of contract execution data such as payments, purchase orders and possible complaints (Cocciolo et al. 2023, 273). This is similar to transactional procurement data from private sector. In addition to analyzing spend, payment data can be used to study the frequency of payments and the time for payment completion. Data on complaints help in connecting procurement processes and contracts to these redressal mechanisms, but this is not as used in practice as traditional public procurement data analytics. (World Bank 2022, 8; Cocciolo et al. 2023, 260) As some of this post-tender data is processed in databases, its variety is more structured, and it is easier to analyze with traditional methods and tools (Cocciolo et al. 2023, 272).

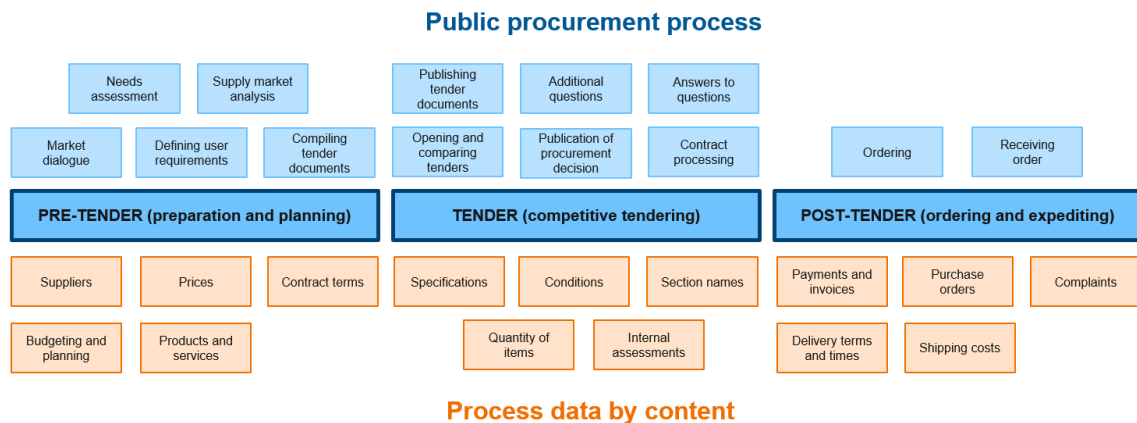


Figure 3. Public procurement process and process data by content.

4 Research design

In this section, the empirical study is being presented. First, research methodology and process of the study are explained. Second part includes a description of the data collection and analysis of this study. In the final part of this section, reliability and validity of the study are discussed.

4.1 Research methodology and process

The empirical part of this study was conducted as qualitative research. According to Hammersley 2013, 10-14, the qualitative research and its methodology are commonly focused on verbal statements and their description and analysis, whereas quantitative research focuses on numerical and statistical elements. This means that qualitative research generally generates a lot of unstructured data that must be organized by the researcher, which is why qualitative research usually includes a smaller number of study subjects compared to quantitative research.

In this paper, the chosen research strategy is case study. As a research method, case study provides a detailed and in-depth investigation about a specified case or cases. These can consist of empirical units like organizations or theoretical constructs such as resilience. (Patton 2015, 259). Case studies can be explanatory, exploratory, or descriptive in nature (Kähkönen 2014, 34). Case study research can generate data that is both rich in description and includes real-life examples and quotes. This research type works especially well for exploration of unfamiliar topics, theory building, and theory testing. As the research strategy aims to capture the reality of a specific case or cases, they often have a lot of value for the practitioners of the topic. (Voss, Tsiriktsis & Frohlich 2002, 195-196; Kähkönen 2014, 31)

Research process is illustrated in Figure 4. The research process started with defining preliminary research themes and questions with the case organization. Literature review on the relevant topics was conducted next, which assisted in planning the interviews and

deciding on a suitable interview approach. After data collection, interview data was analyzed. Final part of the study, conclusions, includes answers to research questions.

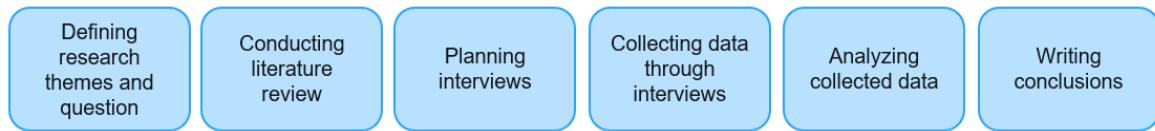


Figure 4. Research process.

4.2 Case organization

VTT Technical Research Centre of Finland, commonly known as VTT, is a research organization and a non-profit limited liability company. It is one of Europe's leading research organizations in the field of applied technological research, innovation, and development. VTT is fully state-owned and is under the steering of Ministry of Economic Affairs and Employment. VTT is an independent and impartial organization, but most of its funding comes from government grants. It has over 2000 employees and the main building is located in Otaniemi in Espoo, but it has locations and offices in five other Finnish cities as well.

The core of VTT's strategy is to find solutions to global challenges and its research is focused on three key areas that are carbon neutral solutions, sustainable products and materials, and digital technologies. VTT's customers are mainly private sector companies that want to develop their business and find new sustainable solutions through technology and innovation. VTT has strong ties with universities and other research institutions to advance collaboration and information sharing in areas that are connected to Finland's priority areas.

VTT's procurement team consists of 19 employees. As VTT is a research organization in multiple areas, it has varying purchasing and tendering categories. Possible tendering and purchasing needs can be almost anything such as office supplies, subcontracting, mechanical components, different services, and research equipment.

There are also defined procurement processes and instructions in tendering and purchasing that are available for all VTT employees. Tendering process covers a description and explanation of planning, procedures, and thresholds in public procurement. Purchasing process covers guidelines on how purchasing works in VTT and which factors employees should consider when requesting a purchase.

4.2.1 Data management and classification in VTT

VTT has started a data classification process and updated its data management principles where data is recognized as an asset. All departments have started the process of identifying what data assets they collect and use in their work and where this data is located. The scope is to classify all data assets in the organization as they could have sensitive data. Next steps include defining responsibilities and roles around the data and starting data classification process.

Along with data management principles, VTT also released data classification and processing guidelines that are applicable to all departments and personnel. The guide introduces the procedures that are related to confidentiality and data processing, and it explains the factors that impact the data classification. Guidebook has instructions on defined security levels, distribution levels, and different IT-systems. It also includes steps to classify information.

First step of classifying data is to define the information security level. This involves the assessment of financial loss or reputational damage in the case of disclosure. Case organization has five classification levels that are Unclassified, Restricted, Confidential, Secret and TL-, EU-, NATO-classified information. As per rule, information is classified as sensitive by VTT. Data labelers are given examples of each category to assist with the evaluation of the classification level. Data classification levels, principles for assigning the information security level and some examples are included in Table 1.

Table 1. Classification levels of data.

Name	Principles for assigning this information security level	Examples
VTT-UNCLASSIFIED	<ul style="list-style-type: none"> Disclosure does not cause financial loss to VTT, or disclosure does not cause reputational damage to VTT 	<ul style="list-style-type: none"> Publishable research results and press releases Publishable personal data, for example work contact details of employees operating in customer interface
VTT-RESTRICTED	<ul style="list-style-type: none"> Disclosure causes significant losses or disclosure causes moderate reputational damage to VTT 	<ul style="list-style-type: none"> Contact details from public sources or otherwise non-confidential contact details
VTT-CONFIDENTIAL	<ul style="list-style-type: none"> Disclosure causes high losses or disclosure causes significant reputational damage to VTT Default – used if there is no information on security level 	<ul style="list-style-type: none"> Customers' confidential information VTT's trade secrets Pseudonymised special categories of personal data and particularly sensitive data
VTT-SECRET	<ul style="list-style-type: none"> Disclosure causes extremely high losses or disclosure causes critical reputational damage to VTT Legislation or customer requirements restrictions Requirements for processing VTT-Secret information shall be assessed case by case 	<ul style="list-style-type: none"> Special categories of personal data Otherwise confidential personal data (e.g. non-disclosure)
TL-, EU-, NATO-classified information	<ul style="list-style-type: none"> Only used with the national security classification (TL), or EU and NATO classification at the customer's request 	<ul style="list-style-type: none"> Information with national TL IV, TL II, TL II classification

Second phase includes the evaluation of suitable IT-system for processing of data. This is based on the classification level that has been determined in the earlier phase, so it does not require further assessment from data labeler. The aim of this is to ensure that employees are aware which systems can be used for different data classifications based on system's security level, which is a protective measure in data security.

Third step is to consider the distribution level of data. There are four distribution levels that are Public, Internal, Limited, and Individually. Similarly, as with security level, the guide includes examples and descriptions of these levels to help with labelling. The distribution levels are also included in Table 2. The combination of these three phases form organization's data classification scheme.

Table 2. Distribution levels of data.

Classification	Access rights	To expand scope of distribution	Examples
VTT Distribution-Public	<ul style="list-style-type: none"> No restrictions Information is typically already available from public sources 	<ul style="list-style-type: none"> VTT employees can expand the scope of distribution independently 	<ul style="list-style-type: none"> Research results/outputs that have been approved for publication VTT press releases and publications Work contact details of employees operating in customer interface
VTT Distribution-Internal	<ul style="list-style-type: none"> Information can be freely shared within VTT 	<ul style="list-style-type: none"> VTT employees can expand the scope of distribution independently within VTT. Information may not be published on the internet or disclosed to parties outside VTT without the owner's permission 	<ul style="list-style-type: none"> Internal instructions and news Basic information describing a customer project Work contact details of VTT employees not operating in customer interface
VTT Distribution-Limited	<ul style="list-style-type: none"> Assigned to designated groups or persons in accordance with the need-to-know principle. 	<ul style="list-style-type: none"> Distribution can be extended at VTT by the "owner" of the information If necessary, the person disclosing the information can set additional restrictions or liberties for the processing of the information in connection with the classification 	<ul style="list-style-type: none"> Identifying customer data VTT's trade secrets Project materials, agreements, project results before publication Regular personal data used in research purposes All other information subject to the confidentiality obligation imposed on VTT
VTT Distribution-Individually	<ul style="list-style-type: none"> Information is disclosed individually only to the recipient present in the situation 	<ul style="list-style-type: none"> This person may not further disclose the information in question even within VTT Only the original owner of the information can share it 	<ul style="list-style-type: none"> Development discussion materials and other information concerning employee evaluation Information under a personal NDA (specifying individual distribution)

4.3 Data collection

In qualitative research, there are few limits in method of data collection. First, the researcher must decide if they want to use existing secondary data, collect new primary data, or mix these methods to answer the research questions. Secondary data can include raw data, published reports, summaries, statistics, or other documents. Primary data is usually collected through interviews, participant observations, and surveys. (Hammersley 2013, 50; Saunders, Lewis & Thornhill 2000, 188-189)

In this study, primary data is used, and this data was collected in interviews. As a data collection method, interviews can help collect valuable data that is relevant and connected to research question and objectives. Interview is a focused discussion between two or more people, and the interaction is meant to provide information about a specific topic. As interview is based on interaction, its clear advantage as a data collection method is flexibility. The interviewer can repeat, rephrase, offer additional information, or change the order of questions if necessary. In addition to recognizing what is being said, it is also possible to note how it is said, which can lead to more nuanced findings, but these interpretations must be carefully assessed. (Saunders et al. 2000, 242; Tuomi & Sarajärvi 2009, 73)

There are multiple methods of conducting a research interview, and the most beneficial way depends on the purpose of research, research questions and objectives. In this study, the interviews were semi-structured, and they included a list of themes and questions that were covered to respond to research questions. Semi-structured interviews are useful in exploring new topics, finding out in detail what is connected to the phenomenon and answering to why, what, and how questions (Saunders et al. 2000, 242-245). According to Koskinen, Alasuutari & Peltonen 2005, a semi-structured approach allows more room both for interviewees and interviewer compared to a structured one. Semi-structured approach has an outline and ready-made questions by interviewer, but interviewees can answer very freely and even present own questions. Hammersley 2013, 54, states that the aim of this style of interview is to allow interviewee to speak in length and their own terms, which helps to capture interviewee's deep knowledge, perspectives, and own experience about the topic.

In this study, the interview questions and themes were developed based on case organizations official procurement process description and guidelines, which guaranteed that all process phases were sufficiently covered. The focus points of interviews and order of questions somewhat varied between interviewees: additional questions and flow of conversation influenced the interview process. However, all semi-structured interviews also included the same pre-determined questions and themes. The interviews were conducted with case organization's procurement employees, which were held in Teams. Even though interviewees to some extent understand each other's tasks, all have their own areas of expertise. The interviewees were divided into two groups according to these fortes. First group was interviewed about pre-tender and tender processes, and second group's interview questions centered around post-tender process phase (Appendix 2 and 3).

Interview can consist of multiple kinds of questions, and different question types have their own purpose. Closed questions are handful in collecting a specific piece of information or confirming a fact, whereas open questions allow participants to express their views descriptively and extensively. Probing questions from interviewer either aim to direct the conversation to a particular focus of direction or seek a further explanation from an answer. (Sanders et al. 2000, 260-261) In this study, the interview structure included open-ended questions that offered flexibility and additional insight to answers. Open ended questions included "what kind of data is generated from this process phase?", "where is this stored

and/or located?”, and “in what form is it stored and/or located?” Closed and probing questions were used when necessary to confirm certain facts or rising aspects during interview.

In total 13 interviews were conducted in Teams between July and September in 2023. The lengths of interviews varied from 38 minutes to 1h 14 minutes. Tables 3 and 4 show the length of each interview. The interviews were held in Finnish. The interviewees received a description of the interview topics in the invite, which confirmed that interviewees knew the theme of the interview beforehand (Appendix 1). Interview objectives and questions were also explained in the beginning of the interviews. The aim was to have a conversational approach to the interview and not to follow process description too strictly. The reason for this was that many interviewees had expressed that some of the topics related to data felt difficult and complex. A more relaxed and less structured approach allowed interviewees to speak freely and if the conversation included related themes, following precise process flow and questions was not necessary. However, if the discussion started to center around unrelated topics, it was directed back to relevant themes.

Table 3. Interviewees and lengths of interviews for pre-tender and tender phases.

Interviewee	Length of the interview
Interviewee A	1h 14 min
Interviewee B	1h 9 min
Interviewee C	59 min
Interviewee D	1h 3 min
Interviewee E	44 min
Interviewee F	52 min

Table 4. Interviewees and lengths of interviews for post-tender phase.

Interviewee	Length of the interview
Interviewee G	1h 7 min
Interviewee H	1h 8 min
Interviewee I	51 min
Interviewee J	49 min
Interviewee K	46 min
Interviewee L	38 min
Interviewee M	1h 2 min

4.4 Data analysis

To analyze the data in this study, notes were taken during interviews. Approximately 26 pages of interview notes were written during interviews. The notes and their analysis were done in Finnish, and the empirical results of this paper were translated to English. Initial ideas and thoughts were written right after the interview took place, but more in-depth analysis was done a bit later after the interview. Data was analyzed by content analysis. It is a research method that uses different techniques to draw valid conclusions from text such as open-ended interview answers. Even though there are no definitive rules for content analysis, the key feature is that text and words are categorized into smaller content categories. (Weber 2011, 3)

Preparation phase of analysis usually starts with selecting the units of analysis. These are generally relevant themes in the context of the study. (Elo & Kyngäs 2008, 109) First, interview transcripts were examined and familiarized to look for recurring and repeating themes and words. In practice, these key words and themes consisted of different data topics and aspects linked to them. Based on these initial findings, data was further organized by coding. This is the practice of labelling research material with a short word or description that recaps its content, which makes it easier to summarize and connect to other data (Linneberg & Korsgaard 2019, 259).

Selected codes were chosen directly from previously identified themes and popular words from interviews. This approach left researcher with a list of tens of detailed and scattered

codes. To develop a better understanding of data, a second round of coding was conducted after the first one. Second round of coding helps to reduce the number of codes to a more manageable level, figure out structures and connections between codes, and develop more progressed concepts from data (Gioia, Corley & Hamilton 2013, 20). This was done by seeking similarities and differences between codes, which resulted in providing higher level categories that were connected to first-round codes. For researcher, this provided a data structure that helped with the outline of procurement process and related data.

Later, codified data was filled to an outline of case organization's procurement process map. After general outline of the process and connected data was complete, process phases and data themes were filled to a table where their classification and variety were further assessed. This was done by simple color coding, which works well with a smaller amount of data and codes (Linneberg & Korsgaard 2019, 260). In practice, each data variety and sensitivity had their own color in the table, which helped to gather a unified view to data classifications in each procurement process phase.

4.5 Reliability and validity

Reliability of the study measures how reliably the research results can be replicated. A study has a good reliability if the study results can be repeated with the same indicators. Tuomi & Sarajärvi 2009, 136). The interview process, data collection and data analysis has been described in the previous chapters, which improves reliability as the same research process can be repeated.

Validity evaluates if the research results and conclusions match, and whether study results accurately answer to research questions. (Tuomi & Sarajärvi 2009, 136). Lack of generalizability has been the main area of criticism in case study research (Ellram 1996, 104). Case studies can include either one or multiple organizations, and this choice impacts the results. The positive aspect in single case studies includes in-depth observations, but it can impact generalizability of the study. (Voss et al. 2002, 201-202) As this study is focuses only on one organization, studying more different public procurement units could improve

both the validity and generalizability. However, as the public procurement process is regulated by law, it is likely that many aspects of the results are still applicable.

5 Empirical findings

This chapter will introduce the findings that are based on conducted interviews. Empirical findings are presented by process-phase, starting from pre-tender and ending to post-tender. First, data is introduced according to their general content and topic. After general introduction data is classified according to case organization's data classification steps and data variety of each process phase is assessed.

5.1 Pre-tender data

In the pre-tendering phase, a lot of external data is gathered. This data is collected through market analysis and dialogue, and it can include both written text and oral information. Collected data is often connected either to the object of procurement or to potential tenderer(s) in general. The data is needed to determine which aspects and specifications are relevant in this specific procurement project and how these aspects and specifications should be measured and assessed.

In general, interviewees agreed that the role of data in the pre-tender phase is vital, and to rush through this step would be a huge mistake as the collected data in pre-tender guides the whole procurement process. Internal customer is generally encouraged or to some extent expected to do market analysis independently before the procurement team becomes involved. However, as interviewee A stated, in practice procurement has a lot of responsibility that the specifications fit the markets. They also confirmed that this data supports and guides procurement unit's decisions moving forward. They also mentioned that pre-tender data helps procurement unit to prioritize which requirements and focus points are most important, determine limitations to the object of procurement, and set a realistic timeline for the project. Interviewee E stated that the importance of market analysis and dialogue has evolved to be even more important than before as the markets in most fields have got more fragmented and complex. They also confirmed that pre-tender data is crucial in formulating an invitation to tender that fits the markets well, which ensures that procurement unit receives enough suitable candidates in the tendering competition.

All interviewees mentioned technical information about the object of procurement as one of the main focuses of data collection in pre-tender phase. Needed data and its volume varies significantly according to nature of the procurement object: it is very different to compare specifications between office supplies and expert services. This generated data is crucial in formulating fitting requirements for invitation to tender.

Interviewees B highlighted that especially with technical aspects collaborative dialogue with the markets plays an important role in pre-tender. As procurement unit can start to prepare preliminary requirements for object of procurement based solely on internal customer's wishes, these specifications may sometimes be unappealing or even unrealistic to tenderers. In practice this means that when procurement unit describes their preliminary requirements and wishes to potential tenderers, it can become clear that there is no such product or service available in the market. Interviewee E also mentions that internal customer's wishes must be often developed to match the markets. Collaborative discussion about these requirements leads to a situation where tenderers have an opportunity to tell what kind of specifications their products can meet, and procurement unit can have a further discussion about this information with internal customer to re-formulate requirements.

Market analysis and dialogue also generates data about prices and pricing models in pre-tender phase. As public procurement aims to ensure efficient use of public spending, price is always included in the selection criteria. Interviewee D states that information about prices and pricing models is especially important if the object of procurement and its markets are new and unfamiliar to procurement unit. Pricing models also vary according to industry, and as invitation to tender generally includes an attachment that offers additional information about target prices, this must be understood in the procurement unit before releasing invitation to tender.

Procurement unit also wants to gather data about the potential tenderers as companies. This includes information about credit ratings, financial situations, and insurance policies. Contractual specifications can be discussed with tenderers: as procurement unit has a contract draft attached to published invitation to tender, it is necessary to assess which terms are suitable both for the markets and for the procurement unit. Interviewee E confirms that

contractual information should be always included in market analysis and dialogue. Tendering stage should not be started if there is uncertainty about the contractual obligations fitting the markets. In a worst-case scenario, tendering process is carried out but contracting parties cannot come to an agreement in the end of it, which would be a huge waste of time and resources for everyone involved.

Tenderer's competence and references are another subject of data in pre-tendering phase. Procurement unit is interested in knowing how much experience tenderers and their personnel have in providing services or products especially to other similar organizations, which helps to assess their capability and trustworthiness. Competence can be assessed both through references and experience. Interviewee D mentioned that procuring expert services generally includes more detailed competence and experience specifications directed at tenderer's personnel, which means that these specifications must be planned carefully.

Most interviewees mention data about tenderer's information security and systems as a crucial matter in pre-tender. Interviewees agree that even though setting requirements regarding cyber and information security is extremely important, it is very challenging. This holds especially true in tendering objects that center around sensitive information, such as procuring a new payroll system. Case organization has strict policies regarding security, which can lead to a situation where potential tenderers cannot fully fill or commit to pre-set requirements from procurement unit. Collaborative approach to pre-tender is needed to find the right wordings and terms between procurement unit and tenderers.

5.1.1 Pre-tender data by variety

There is no regulation or specific instructions on how and where pre-tender data can and should be collected. Data can be tenderer's PowerPoint-presentations, different Word-files, and emails, which do not have a predetermined structure and it cannot be stored in rows and columns in a database. It requires a lot of processing and is very text-heavy, so data in pre-tender is mostly unstructured in variety.

Data on companies' competencies, references, pricing models, and information security systems are often described with own words by potential suppliers. Even though pre-tender data can include price and numerical information that is generally connected to structured data, this information can also be communicated in multiple ways and forms, which make it difficult to process without human involvement and analyzing it with traditional statistical methods is not possible.

However, some data may be also semi-structured. For example, if case organization has left an official request for information in Hilma and there are ready-made closed-ended questions in the sheet, this data can be semi-structured. In this case, the data includes tags and patterns that make it easier to dissect and place data into specific fields. Still, the content of request for information can vary a lot depending on the object of procurement, and the document includes free-form text from potential tenderers, which complicate the automated search and analysis of this data.

5.1.2 Data classification of pre-tender

Pre-tender data is used internally for next phases of procurement process, but is it not directly released or published from case organization. Pre-tender data consists of information regarding the markets that includes data about services, goods, and potential tenderers. Pre-tender data is collected by case organization directly from potential tenderers, meaning that data collection is based on a specific need that is communicated to the tenderer. This gives the tenderer an opportunity to formulate their response specifically for case organization's needs.

The collected data can include information that is not available to public. For example, all references from potential tenderers may not be public information. Preliminary pricing may also vary according to customer. When it comes to personal information, tenderers can also give information about organization's employees and their work history. Technical aspects and other limitations are also areas where tenderers may not disclose all information publicly.

Due to these reasons, pre-tender data could not be classified as Unclassified. Classification of confidentiality has been labelled as Confidential, and the used IT-systems are marked accordingly. Even though potential tenderers are not customers, they can be considered as partners that are linked to the same category. Distribution of pre-tender data should be classified as Limited. It is not necessary to share this information internally and it should be assigned to appointed project group. It includes project information and identifying data about potential business partners, and both are linked to Limited distribution level. An overview of pre-tender data, classification levels, IT-systems and distribution levels can be seen below on Table 5.

Table 5. Pre-tender-phase data by confidentiality, IT-system, and distribution classification.

Data in pre-tender	Classification of confidentiality	IT-system	Classification of distribution
Object of procurement	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Technical aspects	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Limitations	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Pricing	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Competence, references, and experience	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Information security	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited

5.2 Tender data

In the tendering phase, procurement unit publishes contract notice, invitation to tender, contract draft and other possible tendering documents that are based on data from pre-tender phase. Received offers from tenderers are new data that is processed by procurement unit. Tender data is collected and created in the context of regulated public procurement process, and it is done through text, mostly through official forms and channels. Even though this data is needed to proceed to final phase of official tendering process (contracting), it is also collected and assessed because of legal obligation to do so.

Contract notice and invitation to tender have similar content. Invitation to tender has an official template that can be used for all public procurement projects. First part of invitation to tender includes general information about the procurement. Sub-sections of general information include information about procurement unit such as official name, address, and contact information. Second sub-chapter explains where tendering takes place. Two last sub-chapters include information about the type of procurement unit and what is their main

industry. Procurement object is described after general information. This part includes a short descriptions and sub-chapters about the procurement object, price, and contract. Next part of the document contains procurement procedure and administrative information, and final section covers other complimentary information about the procurement such as appealing.

Invitation to tender can include attachments that can provide additional information and guidance to tender. Possible attachments can be either formal documents or forms to be completed by tenderer. Examples of formal documents include a detailed description of object of procurement, contract draft and/or terms, and code of conduct. Fillable forms can be used for customer statements, prices, or references. Interviewee D mentions that the goal is to decrease the number of open-ended questions or answers, and make these forms as simple and structured as possible, so that tenderer can tick a box or answer yes/no. This is helpful both in terms of efficiency and clarity.

During tendering phase and after tendering documents have been published, it is possible for tenderers to ask additional questions about the tendering project. Interviewee C states that questions are usually linked to unclarities or contradictions in the tendering documents. However, interviewee F mentions that most of the time there are no additional questions, but if there are inquiries from tenderer's side, they are mostly connected to technical specifications. Interviewee E summarizes that surplus of questions in this phase is a sign that something has gone wrong in the earlier stages of procurement process. Possible reason for multiple additional questions can be that market analysis was done poorly, and there hasn't been enough data and information to formulate sufficient tendering documents. Responses to additional questions are worked together with internal customer and project group, and they are released for all tenderers.

After offer period has passed, begins the evaluation process. During the evaluation, different internal documentation is created. First, received bids are opened at the same time. Procurement team comprises a formal document regarding the time of opening bids. After this, it is checked if received bids fill determined and published qualification criteria. This is done in a qualification matrix where all qualifications and tenderers are listed. Project team

marks if the tenderer fills a specific qualification in the matrix, and they can also leave open-ended comments in the document. Interviewee D states that the minimum requirements vary a lot between procurement projects, which means that the collected and processed data may also be very different both in content and volume. In case there are unclarities regarding tenderer's answers, procurement unit can send a tenderer a clarification request.

If there are multiple tenders that fill the qualification criteria, project team continues with further evaluation of tenders. Evaluation is completed with a similar evaluation matrix as qualification matrix. Evaluation matrix includes both written comments and given points by project team to tenderers and specific criteria. Interviewee A mentions that there are both internal and external matrixes in tender phase. First one is connected to internal implementation process of comparing tenders, second one will be published as official evaluation matrix for the Award decision.

Based on evaluation, procurement decision is made, and contract award notice is released. It contains the general information of the procurement, winner of the tendering competition, and contract information. Attachments of contract award notice contain detailed instructions on appealing process. Another attachment is the justification for the award decision. This attachment includes a detailed description and reporting of tendering decision, received tenders, tendering procedure, and evaluation criteria of tenders. Individual tenderer's points and procurement unit's comments to these are also published as an attachment to contract award notice.

As mentioned earlier, contract draft is usually published with invitation to tender. This is why interviewee E states that there should not be much new data at this stage, and often tenderer's general information, contacts, and prices are filled to ready-made contract. In addition to main contract, there can be other contracting documents created in this stage.

5.2.1 Data classification in tender

In the start of tender phase, case organization releases multiple public documents. These consist of contract notice, invitation to tender, contract draft, and possible other attachments

to tendering documents. Attachments can include forms about prices, references, or customer statements. As these are available online and can theoretically be accessed freely, data is considered as Unclassified. The distribution can also be determined as Public because information can be shared freely.

Act on Public Procurement and Concession Contracts 64§ states that data security of tenders and requests to participate must be ensured when sharing and storing information on a procurement procedure. Procurement unit must ensure that the contents of tenders or requests to participate are not disclosed before the time limits for submitting them has expired. This sets additional requirements for case organization's data security and processing with tenderer's offers and requests of additional information. In addition to this, tender data can include information that is not available to public such as detailed employee information. This is why tenderer's offers are classified as Confidential data, and their distribution level is Limited. Additional questions become public and Unclassified data as these will be published, However, the identity of questioner will not be released. As answers to additional questions are released to all tenderers, information is also classified as Unclassified, and distribution level is Public.

Case organization has own qualification and evaluation documents that can include internal comments and questions to other project team members, and this data can be sensitive. As this data is related to internal evaluation process of tenders in the procurement unit, it is not public data. Also, additional questions to specific tenderers can include personal data. All this data can be classified as Limited, and distribution level Limited. However, contract award notice, justification for the award decision and other related documents are released publicly in the final phases of tender process. These are classified as Unclassified and distribution level is Public.

All classification and distribution levels are summarized in Table 6. However, it should be noted that none of the publishable documents should not be classified as Unclassified while they are still in progress. In practice it means that these documents should not be disclosed or distributed publicly or processed in Low-level IT-systems before the official release of contract award notice.

Table 6. Tender data by confidentiality, IT-system, and distribution classification.

Data in tender	Classification of confidentiality	IT-system	Classification of distribution
Contract notice	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Invitation to tender	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Contract draft	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Customer statements (company) document	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Price list document	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
References (personnel) document	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Additional questions from tenderers	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Answers to additional questions	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Tenderer's offers	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Opening bids	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Qualification criteria matrix (internal version)	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Additional questions from VTT	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Evaluation criteria matrix (internal version)	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Contract award notice	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Justification for the Award Decision	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Evaluation matrix for the Award Decision	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public
Information on Appeals	1. VTT-UNCLASSIFIED	1. Low (VTT- unclassified)	1. VTT Distribution-Public

5.2.2 Tender data by variety

In tender phase, there are different data varieties, but most common data variety is semi-structured data. Documents published by procurement unit such as contract notice and invitation to tender follow a predetermined pattern. Contract notice includes tags and headings such as type of procurement, short description of procurement, procurement vocabulary code (CPV), and type of procedure, but there is also free-form text in some of these fields. Invitation to tender includes similar fields, tags, and information, but also additional data about requirements and specifications. Some requirements and specifications can be closed-ended fields where “yes” answer is required to proceed in the process, other parts may involve detailed text-form descriptions of object of procurement. These characteristics make invitation to tender and contract notice semi-structured data.

Similar approach is applicable to most attachments that are linked to invitation to tender: both open-ended and closed-ended sections are included in customer statements and references. However, price lists are generally more straightforward as they do not require text but just numbers to a ready-made table, which make them easy to compare with statistical methods and store the data in a database. In this case, it is possible to categorize price lists as structured data, and other attachments as semi-structured data.

Clarification request to tenderer, possible additional questions from tenderers and answers to these questions are generally unstructured data. Questions can be either straightforward yes or no -questions or long and detailed. Answers by procurement unit are generally kept as short and simple as possible. The same approach applies to possible clarification request from procurement unit to tenderer if the received bid is unclear in some part.

Opening bids is semi-structured data document that include time, place, participants, number of received bids, and a list of all bidders. Received offers follow the same pattern as invitation to tender, because they are filled to tender form. Tenders include the same tags and headings as invitation to tender, and parts can include closed-form “yes” answers, numbers, or free-form text. All matrixes (qualification matrix, evaluation matrix, and evaluation matrix for the award decision) are also categorized as semi-structured data. They are in tabular form and include insights both directly from received offers and comments from project team regarding the offers. Evaluation matrix and evaluation matrix for the award decision also include given points and numerical assessment for specific criteria.

Contract Awards Notice, Justification for the Award decision, and Information on Appeals follow a similar pattern. Documentation includes metadata that can be used to place data into fields and columns according to topic, but especially Justification for the Award decision is very text-heavy, and it is not possible to analyze it with traditional tools.

5.3 Post-tender data

In the post-tender phase, contract has been signed by procurement unit and supplier, and general terms between parties have been set. Post-tender data is focused on contractual follow-up and operational activities that fulfil the contract. In this phase, collected and created operational data is often specific and focused on singular transactions.

Post-tender data starts with a purchase requisition from internal customer. This request is made in internal system. It includes data about requester, approval flow, project, supplier, product, product category, quantity, price, delivery address, and required delivery time. In some cases, there may be a quotation or price list attached to requisition.

Purchase order request is being processed by a buyer that creates a purchase order regarding the requisition. All the data from requisition transfers to the purchase order. In the case of contract purchase, delivery and payment terms are generally linked to supplier information. From supplier side, order confirmation confirms the terms of the purchase. It generally includes information about the shipping and delivery as well. Attached to order confirmation, there may be warranty details or other terms or conditions.

5.3.1 Data classification in post-tender

In post-tender, all data is classified as Confidential. This data is related to a business transaction that can include sensitive data. Requisition also includes project information that can be sensitive and related to customer's information. Post-tender-phase data by confidentiality, IT-system, and distribution classification are specified in Table 7.

Classification of distribution is Limited. Even though the requisition is made by a specific person, there may be valid reasons to share the delivery information with a team member, and requisition and purchase order do not include information that would require Individually-level distribution.

Table 7. Post-tender-phase data by confidentiality, IT-system, and distribution classification.

Data in post-tender	Classification of confidentiality	IT-system	Classification of distribution
Requisition	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Quotation	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Contract	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Purchase order	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Order confirmation	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited
Delivery	3. VTT-CONFIDENTIAL	3. Medium (VTT-confidential)	3. VTT Distribution-Limited

5.3.2 Data variety in post-tender

There is a lot of structured data in post-tender. It includes numerical data in the form of quantity, price and required delivery time. Supplier, project, delivery address, and product category are also selected from requester's side, and these are in ready-made structured

format. Product name is the only field that can include completely free of form text, which makes it unstructured.

As the requisition is directly connected to purchase order, data in these is similar. Purchase order can include additional data such as incoterms and payment terms. These are also selected from selection bar by purchaser in system's columns and rows, so the variety is structured as it can be easily tracked and analyzed in the system.

Still, some data is not structured such as order confirmations or delivery information that may have some structure but varies between suppliers. The same can be applied to quotations: even though there are tags and headings, their structure between suppliers is not the same. Contract form is generally the same for all suppliers, and it is semi-structured as most tendering documents.

6 Discussion and conclusions

This chapter concludes the main findings of this paper. Results of empirical part are summarized, discussed, and evaluated together with relevant literature and conducted literature review, and research questions are answered. The main objective of the study is to investigate what kind of data is generated through public procurement process.

6.1 Answering to research questions

The main research question is “What data is collected and created in the public procurement process?” Two supporting research questions were created to respond to provide additional insights to main objective. The answers to sub-research questions are presented after main research question.

What data is collected and created in the public procurement process?

Pre-tender data includes data collection through market analysis and dialogue with potential tenderers. This data is used in the creation of tendering documents that are published in next process phase. Collected data includes information about the object of procurement and potential tenderers as companies. Object of procurement can include data about specifications like technical specifications. Price and pricing models are also an important data topic in this phase. When it comes to potential tenderers, there can be information about financial aspects or contractual terms. References and competence as a data or information topic were emphasized in interviews. Also, information security was considered a very important topic in data collection.

It has been well established that procurement units need to articulate their needs, collect data from the markets, and process this data in the pre-tender phase. Communication can include open dialogue between procurement unit and companies to gain mutual understanding of the requirements for eventual invitation to tender. (Holma, Østensen, Holmen, & de Boer 2022, 52-53) The role of market analysis and dialogue is especially important with procurement

projects where the object of procurement is complex and unordinary (Pekkala et al. 2022, 21), and the level of interaction is often dependent on object of procurement. In some cases, tenderers and procurement unit do not need to have a mutual discussion in pre-tender. This might be the case with standard services or products that do not require tailoring or altering. (Holma et al. 2022, 53) In these situations, there may be fairly little collected data in pre-tender phase.

During interviews, the importance of collected pre-tender data was highlighted. As the case organization is heavily focused on innovation and applied research on new and trending areas, many procurement projects can have these characteristics as well, which can partly explain the emphasized role of pre-tender data. According to Johnsen, Mikkelsen & Paulraj 2016, 50, many poorly specified and failed public procurement purchases have been reported in the media, and common denominator for these cases has been complex nature of the object of procurement. In these cases, internal customer's needs and wishes may be difficult to conceptualize to the markets in a clear and straightforward way, which requires more conversations and collaboration in formulating fitting requirements. This high number of stakeholders can also be another reason for importance of pre-tender data collection and processing.

In tender phase, procurement unit releases the official tendering documents that are the most prominent part of public procurement process. These include contract notice, invitation to tender, contract draft and other attachments such as customer statement documents, price lists and references (personnel) document. Contract notice and invitation to tender form the main data content of procurement and public procurement. As these documents are subjected to regulation, they often include very similar data and there is little variation between contents. (Pekkala et al. 2022, 345-348)

Collected data includes tenders and possible additional questions from tenderers. After answering possible questions and receiving tenders, internal assessments are done based on received data. These are needed to check if tenders fulfil the minimum criteria and/or exclusion criteria (Pekkala et al. 2022, 22) As communication in tender phase is formal between procurement unit and tenderers, the collected and created data is procedural and

somewhat limited to constraints of public legislation (McKevitt & Davis 2015, 79). Due to these reasons, there are not many surprising elements in tender data, and the contents of tender data are generally well-known.

Post-tender data comes from contractual follow-up and operational activities. Most of it can be defined as operational purchasing data, and some of this data is similar to private company's purchasing as can be seen from Table 8. Created data includes a purchase requisition from internal customer. This request includes data about internal customer, supplier, and product. In some cases, there may be a quotation or price list attached to requisition that has been received from supplier. Collected data from supplier includes order confirmations and delivery information.

Table 8. Collected and created data in public procurement process phases.

Pre-tender data	Tender data	Post-tender data
<ul style="list-style-type: none"> • Data collection through market analysis and dialogue • Creation of tendering documents to be released on next phase <p>Collected data contains information about</p> <ul style="list-style-type: none"> ◦ Object of procurement: technical specifications, prices and pricing models ◦ Potential tenderers: financial information, contractual specifications, references and competence, information security and systems 	<ul style="list-style-type: none"> • Publishing of data and official tendering documents based on previously collected data <p>Created data</p> <ul style="list-style-type: none"> ◦ Official tendering documents such as contract notice, invitation to tender, other attachments, contract award notice, justification for the award decision, information on appeals ◦ (Answers to additional questions) ◦ Internal assessments and discussions regarding offers and questions <p>Collected data</p> <ul style="list-style-type: none"> ◦ (Additional questions from tenderers such as content of invitation to tender) ◦ Tenders 	<ul style="list-style-type: none"> • Contractual follow-up and operational activities that fulfil the contract <p>Created data</p> <ul style="list-style-type: none"> ◦ Purchase requisition from internal customer ◦ Purchase order <p>Collected data</p> <ul style="list-style-type: none"> ◦ Order confirmation ◦ Quotation ◦ Delivery information

How can public procurement process data be classified according to variety and sensitivity?

Overall, data variety is different between public procurement process phases. Pre-tender data is very unstructured as it does not have a clear structure and it cannot be stored in rows and columns in a database. It includes a lot of text from multiple sources and in various forms such as PowerPoint-presentations, PDF- and Word-files, emails, and web pages. Some data

can be semi-structured if it has been collected by forms with ready-made headings and tags, but the content can still be text-heavy and include long sentences. The surplus of unstructured data is not surprising as most organizational data is of this variety (Salo 2013). Overall, the amount of data variety is considered as one of the most prominent challenges in data management (Mynarz 2014, 22).

The amount of unstructured data in pre-tender is impacted by the collaborative and interactional nature of pre-tender phase. Also, the reason for pre-tender data collection is to understand the markets better and collect enough information about varying aspects referring to object of procurement. This often requires vast amounts of raw qualitative data that must be processed and assembled to form clear understanding of the content of data. Another reason for vast amount of unstructured data is the lack of unified approach to data collection in pre-tender. This is due to lack of regulation in pre-tender phase in public procurement. This can be positive in the form of flexibility, but as there are no regulatory requirements on how the data should be handled, there are no specific instructions on the process either. Data can also be collected and processed by several internal customers and/or procurement professionals, which can lead to diverse methods in data collection and documentation. Potential tenderers may also have different ways to communicate and present needed data.

Tender data includes mostly semi-structured and unstructured data, and possibly structured data in the form of price lists. Clarification request to tenderer, possible additional questions from tenderers and answers to these questions are generally unstructured data. Invitation to tender, Contract notice, Contract Awards Notice, Evaluation Matrixes, Justification for the Award decision, and Information on Appeals follow a predetermined layout that has metadata and classification systems, which puts these into semi-structured data category. These are official tendering documents that align with official forms, and according to Mynarz 2014, 22, this standardization is the key for ensuring possible re-use of public procurement data, whether it is for supervisory, analytical, or business purposes. For example, CPV code can be used to extract data from public procurement documents (Duguay et al. 2023, 1171). It should be noticed that these standards and instructions generally only apply to published tender documents. Internal documentation and data such as case organization's evaluation matrixes cannot be directly compared with other procurement

units. An exception to this is opening bids document that is an internal document but is encouraged by regulators.

Also, current standardization is far from perfect, which can result in increased variety of data in tender documents. Possible reasons for this include underspecified standards that leave room for inconsistent approaches and different interpretations among procurement units. This also linked to another cause of variety in tender data: as it is generated by different people and using different forms, deficiencies and irregularities are inevitable to some extent. (Mynarz 2014, 22) Even though there are standards and instructions on what must be included in the documentation, it can be demanding to locate relevant content within these documents as structure can vary between different files according to procurement unit's approach. (Hammoudi et al. 2021, 120) However, even if tackling this issue in larger scale is challenging, it could be possible for individual organizations to make their data be more cohesive and have less variety.

Post-tender data mostly consists of structured and semi-structured data. Structured data includes numerical data regarding prices, quantities, and delivery times. Also, supplier information, payment information and other data directly linked to purchase order are in structured format. Only product name is a field that can include completely free of form text and unstructured data in purchase order sheet. Order confirmations can include either semi-structured or unstructured data, depending on supplier. As purchase order systems are generally built on relational databases, it is expectable that most purchase order data is structured (Chae & Olson 2013, 11-13; Cocciolo et al. 2023, 272). Generally, there is little difference between private and public sector purchase order data. This can be explained by similar terminology, processes, and transactional nature of purchase orders on both sectors.

Data sensitivity varies between different process stages: pre-tender and post-tender data are classified as Confidential, but most data assets in tender phase are Unclassified. Classification class based on sensitivity is based on the estimation of damage on the organization if the data is disclosed, altered, or destroyed without authorization (Agrawal 2017, 265). However, considering the potential damages or losses of a disclosure is rarely

simple. In addition to having little practical research on the area (Bergquist et al. 2022, 154), assessing damages can be very speculative in cases such as losing trust or business partners.

Legal penalties of disclosure can be easier to evaluate as public procurement is subjected to laws that govern how data should be handled and secured. The fear of these consequences risks easily lead to “better safe than sorry” -approach, meaning that public procurement data is classified as more sensitive than it actually is. Even though this tendency is understandable especially with personal data and GDPR, it can be problematic if it risks the principle of transparency in public procurement. (Gleeson & Walden 2016, 685-687; Alexe & Sandru 2021, 224)

Pre-tender data is collected by case organization to a specific need, which may include information that is not available for public, whether it is about specifications or personnel information. Due to these reasons, collected pre-tender data is not publicly disclosed, but it is used to formulate invitation to tender and other tendering documents. This data could be considered similar as customer’s data, which confirms that data assets in pre-tender phase are Confidential. According to case organization’s data classification scheme, Confidential data means that disclosure causes high losses or significant reputational damage. In the context of public procurement, loss of reputation could lead to decreased willingness from the markets to collaborate in pre-tender phases.

Principle of transparency ensures that public procurement process is public. Tendering process data mainly consists of tendering documents that can be examined externally, which is why this data is Unclassified. As these documents will be accessible in public platforms, there is no need to place excessive protective measures on these data assets after their official release. Additional questions become public and Unclassified data as the answers to these questions will be published. However, the identity of questioner will not be released.

Tenderer’s offers and internal evaluation documents are classified as Confidential data. According to act on Public Procurement and Concession Contracts 64§, data security of tenders and requests to participate must be ensured when sharing and storing information on a procurement procedure. Procurement unit should especially pay attention to management

of sensitive personal data that can be included in tenders or requested attachments such as CV's. Access to this information should be limited to personnel who are directly involved with procurement process. (Alexe & Sandru 2021, 230)

Post-tender data is classified as Confidential. This data is connected to execution of procurement contract, but it includes information that can be considered sensitive such as detailed pricing. Also, purchase order data can be used to attempt phishing that can be defined as attempting to appear as a business partner or a reputable company in order to prompt people to reveal sensitive information (Jansson & von Solms 2013, 584). Other security threats connected to disclosing purchase order data can include false invoices or using details from purchase orders to make convincing requests to buyers.

How do these classifications impact the usage and protection of data?

Data variety can cause many issues in the usage and analysis of data. Main one of these is data integration that is defined as the ability to integrate and connect different data types from heterogenous sources and locations (Arunachalam, Kumar & Kawalek 2018, 425). Data integration requires technical components and technology that handles and presents the processed merged data. As many well-known procurement platforms operate on structured data, this is the foundational challenge in integrating data from different varieties and sources. (Handfield et al. 2019, 974-975) Public procurement process data varies between market dialogue data to transactional data, and integration of these is not currently possible for most procurement units. Data integration has already been recognized as an important development area in public procurement and overall public finance management: this would help to provide a more holistic and thorough understanding of the state of public procurement. (Cocciolo et al. 2023, 278) Integration and technical tools are also needed to move beyond descriptive analytics to predictive and prescriptive analytics in public procurement.

As post-tender data is mostly structured, it is the easiest to use and analyze. Post-tender data is most commonly used for descriptive analytics that is the first and least complex level of data analytics. It provides an overview of used suppliers, prices and bought products.

(Cocciolo et al. 2023, 267-268) From procurement unit's perspective, this should be fairly simple to do based on purchase order, invoice, and/or other transactional data using simple statistical measures. However, capabilities for data analytics can vary between organizations depending on their system infrastructure: some may even struggle to implement sufficient spend analysis due to disparate nature of information systems (Rafati & Poels 2015, 31). This is why public procurement data should be collected and maintained in centralized databases that enable better use of data for improved decision-making and follow-up (World Bank 2022, 16).

Using and analyzing of tender data has generally been focused on developing models and indicators that predict likelihood or aim to recognize corruption or crime in public procurement procedures and contract awards. This is usually done by utilizing predictive analytics. Still, it has been recognized that preparing and processing tender data for data analytics is challenging. (Lalic, Gracanin, Tasic & Simeunović 2022, 473, 479) Even though tender data has some form, it still has more variety than hoped as it often includes a lot of text. This makes using analytics tools more complex, and analyzing tenders manually can take relatively long, whether it is from procurement unit's or supervising body's point of view. If the variety was even more structured, analytics tools could be utilized better. One possibility for automated analysis could be in analyzing tender's specifications. This would help to reduce labor costs and working time in tendering phase for procurement units. (Artamonov, Vasilev, Tukumbetova & Ulizkob 2022, 97)

Similar approach to decrease costs and more automated analysis could also be applied to pre-tender data. However, as pre-tender phase is less structured in itself, simplifying data variety may be challenging. In the context of conducted interviews, the role of pre-tender phase and data collection is emphasized, and manual processing is needed for this data. This leads to wonder if data analytics could be implemented to pre-tender as well. As the data is unstructured and textual, it would require tools such as natural language processing, text mining, or other intelligent content extraction methods (Harriott & Isson 2012, 361). These are helpful tools in summarizing and categorizing topics or extracting specific pieces of information from large volumes of text and unstructured data. Because many of these techniques are still considered innovative and require both technological investments and capabilities, they can have high implementation costs. (Rejeb, Süle & Keogh 2018, 79-84)

Singular procurement units may not have the abilities or finances to invest in new infrastructure as this, but this development could potentially be supported by public funds as the probable cost savings pay these investments back in longer timeframe.

When it comes to data sensitivity, the simple act of classifying and labelling data is not enough to ensure data security with sensitive data. As data classification is a part of data protection and data governance, it has other layers that should be included in ensuring data security. Once data has been classified, organization should review its security policies and procedures to determine if data is protected according to classifications. Technical policies are needed to implement and enforce protection of data, but also the human-component of data security must be considered. Users of data have to be educated what kind of data they are processing in their work and what protective measures it requires such as where it can be shared and with whom. (Mahanti 2021, 125; Bradford et al. 2022, 202)

Public procurement process data sensitivity varies between process stages from Unclassified to Confidential. Even though there is diversity between process phases and their data sensitivity, all data is considered Confidential at some point: official public procurement documents are considered Confidential before their official release. Once procurement process data has been classified and risks associated with it have been recognized, it should be evaluated if all data is protected according to classification (Mahanti 2021, 124). In this case, all data has been shared and archived in platforms that have been approved by security.

In theory, this would mean that it is possible to handle tender documents in lower-level IT-systems after publishing them to tenderers. However, it is questionable whether it is useful in practice to handle same process data in different level systems. It always requires assessment whether a particular file includes some data that should be handled in higher security level system. This leaves a lot of room for human-made mistakes in judgement and accidents, which could lead to risk in compliance, especially as staff-related errors are the most likely reason for data leaks (Tankard 2015, 4).

As part of data security and data governance, organization should set roles for data-related activities. Data owners are accountable for collection of data and has approving authority

over data, whereas data stewards carry out the risk-related assessments about data assets based on their expertise. In collaboration, data owners and stewards create the needed processes and rules for handling data. (Mahanti 2021, 126) As public procurement data is handled by many people and two different sub-teams, these roles are needed to ensure that all team members have mutual understanding how procurement data should be managed.

6.2 Implications for practice

Data classification is an exercise that should be done in all organizations. The process of labelling and classifying data assets assists in recognizing what data organization has, why it is collected, where it is located, and how it should be protected. This helps to understand the overall data landscape and figure out if there is an overlap in systems in the form of duplicates. Data classifications and overall data classification policy should be clear and possible to understand for employees. Special attention should also be paid to providing sufficient instructions to data labellers. Even though labellers are generally subject-matter experts in their own field, topics related to information security and risk management may be more unfamiliar to some. Training in assessment of risks is needed to ensure homogeneous approach to data labelling.

When it comes to data usability, procurement units should pay attention and invest in data integration as this is one of the main challenges in data analytics. Data integration enables connecting data from different departments, silos, and teams, which would help to increase understanding and knowledge. As an example, connecting purchase invoice data to procurement contract data would help to automatically monitor if payment terms mentioned in contracts are followed in purchase invoices. In addition to developing better data integration, procurement units could benefit from more centralized IT-infrastructure that facilitates easier sharing of information.

Also, data variety could be simplified in areas where this is feasible to do. In the context of tender data, improved standardization through regulation is one possible solution. As pre-tender phase has a lot of unstructured data, this would be a good place to try to gather more structured data. This could be done by collecting data in a more planned way and using

methods that promote simplification of data variety, such as using structured forms and close-ended questions that can be answered simply. However, it should be noted that this approach may be difficult or even impossible with some procurement projects and pre-tender data collection. Procurement units should also be supported in transition from traditional data analytics to use of more advanced methods that can enhance efficiency and improved decision-making. As use of procurement analytics tools in public sector strengthens efficiency and decreases costs by reducing workload in manual tasks and supporting data-driven management, it is in public interest and should be supported by the public funds. Otherwise, there is a risk that smaller procurement units will be further left behind from this development.

6.3 Conclusions, limitations, and suggestions for further research

This study investigates what kind of data is generated through public procurement process and how this data can be classified according to sensitivity and variety. This study contributes to limited research on the area and takes a process-based approach to data creation and collection in the context of public procurement. It also provides insights on how these classifications can impact the possibilities and challenges in usage and protection of data. In summary, data variety and sensitivity are different between process phases of pre-tender, tender and post-tender, which impacts both data usage and protection.

Pre-tender data can be characterized as unstructured, which requires advanced analytics tools that are not readily available for all organizations. Tender data is semi-structured due to standardization and regulation, but preparing and processing data for analytics is still challenging with traditional techniques. Developed standardization is needed to simplify the variety, and procurement units should be supported to implement new analytics methods and tools to improve data integration. Post-tender data is structured in variety as it can be stored in relational databases, which makes its analysis fairly simple and easy for procurement units. Pre-tender and post-tender data can be classified as Confidential and sensitive data, which is why data protection should be considered when sharing and storing pre-tender and post-tender data. Some tender data is categorized as public and Unclassified

as it includes official tendering documents, but also this data should be appropriately protected before the official release of these documents.

As of limitations of this study, this work focuses on one organization, and findings may not be applicable to all procurement units in Finland. VTT is an organization that does not directly buy products or stock-keeping units into warehouse, which could also potentially limit generalization of this study's results when it comes to order process. Most of the procurement team was interviewed, so case organization's perspectives should be sufficiently covered from different perspectives. Interviews were not recorded, which could also impact the results. In case of recording interviews, empirical part of this study could have included more direct quotes to provide additional insights. More perspectives could have been added from re-listening interviewee's answers, and these could bring more credibility to results.

This work focuses on data that has been generated from public procurement process that starts from tendering and ends to receiving the product. The focus of this study is on procurement process that follows a standard process flow. Therefore, possible cases of reclaims or other disruptions are not discussed in this study. Also, tendering process covers open procedure, not other possible public tendering approaches such as limited procedure or planning competition.

There are various possible directions for future research to gather more in-depth understanding about public procurement data. As this work focuses on one organization, it could be beneficial to compare if utilization, processing, and protection of data varies between procurement units. This could provide interesting insights if there are differences in capabilities, understanding, and or willingness to improve data utilization, or if data classifications are different between organizations. As the nature of this study is qualitative, it would be interesting to conduct a more quantitative analysis of the phenomenon. It would also be valuable to study what are public procurement professional's perspectives of the state of data management in their organizations. For example, what are biggest barriers and obstacles to utilizing procurement data better from their perspective? This could provide more practical improvement suggestions to procurement units.

References

Agrawal, V. (2017) A framework for the information classification in ISO 27005 standard. 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud) 2017, 264-269.

Alexe, I. & Sandru, M. (2021) Data Protection in the Public Procurement Process. *European Journal of Law and Public Administration* 7, 2, 224-239.

Alhassan, I., Sammon, D. & Daly, M. (2018) Data governance activities: a comparison between scientific and practice-oriented literature. *Journal of enterprise information management* 31, 2, 300-316.

Alhassan, I., Sammon, D. & Daly, M. (2019) Critical Success Factors for Data Governance: A Theory Building Approach. *Information systems management* 3, 2, 98-110.

Al-Ruithe, M., Benkhelifa, E. & Hameed, K. (2019) A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing* 23, 5, 839–859.

Arunachalam, D., Kumar, N. & Kawalek, J. P. (2018) Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation research. Part E, Logistics and transportation review* 114, 416-436.

Artamonov, A., Vasilev, M., Tukumbetova, A. & Ulizkob, M. (2022) Multiagent System for Monitoring, Analysis and Classification of Data from Procurement Services. *Procedia Computer Science* 213, 96–100.

Axelrod, C. W., Bayuk, J. & Schutzer, D. (2009) *Enterprise information security and privacy*. Boston : Artech House.

Bergquist, J., Tinet, S. & Gao, S. (2022) An information classification model for public sector organizations in Sweden: a case study of a Swedish municipality. *Information and computer security* 30, 2, 153-172.

Bhatia, A. & Bansal, V. (2015) *Database management system*. Oxford, England: Alpha Science International Ltd.

Business Finland. (2020) The Act on Public Contracts, EU state aid rules, and funding granted by Business Finland [online website] [Accessed 20.9.2023] Available: https://www.businessfinland.fi/49eaf3/globalassets/finnish-customers/01-funding/08-guidelines--terms/instructions/en_hankintalaki_eu_n_valtiontukisaantely_ja_bf-rahoitus.pdf

Bradford, M., Taylor, E. & Seymore, M. (2022) A View from the CISO: Insights from the Data Classification Process. *Journal of Information Systems* 36, 1, 201-218.

Brous, P., Janssen, M., Vilminko-Heikkinen, R. (2016) Coordinating decision-making in data management activities: A systematic review of data governance principles. *International Conference on Electronic Government* 9820, 115-123.

Chae, K. & Olson, D. (2013) Business analytics for supply chain: A dynamic-capabilities framework. *International Journal of Information Technology & Decision Making* 12, 1, 9-26.

Cocciolo, S., Samaddar, S. & Fazekas, M. (2023) *The Government Analytics Handbook. Government Analytics Using Procurement Data*. [online document]. [Accessed 19.12.2023]. Available: <https://thedocs.worldbank.org/en/doc/5958d8e2b7f26b4f817027245b0ba2e2-0050042023/original/GAH-CHAPTER-12.pdf>

Duguay, R., Rauter, T. & Samuels, D. (2023) The Impact of Open Data on Public Procurement. *Journal of accounting research* 61, 4, 1159-1172.

Ellram, L. M. (1996) The use of the case study method in logistics research. *Journal of business logistics* 17, 2, 93–138.

Elo, S. & Kyngäs, H. (2008) The qualitative content analysis process. *Journal of Advanced Nursing* 62, 1, 107–115.

European Commission (2022) Public procurement. [online website]. [Accessed 12.8.2023] Available: https://single-market-economy.ec.europa.eu/single-market/public-procurement_en

Everett, C. (2011) Building solid foundations: the case for data classification. *Computer fraud & security* 2011, 6, 5-8.

Gioia, D., Corley, K. & Hamilton, A. (2013) Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational research methods* 16, 1, 15-31.

Gleeson, N. & Walden, I. (2016) Placing the state in the cloud: Issues of data governance and public procurement. *Computer Law & Security Review* 32, 5, 683-695.

Gordon, K. (2022) *Principles of data management : facilitating information sharing*. Swindon, England : BCS, The Chartered Institute for IT.

Hammersley, M. (2013) *What is qualitative research?* London : Bloomsbury Academic.

Hammoudi, S., Quix, C. & Bernardino, J. (2021) Intelligent Public Procurement Monitoring System Powered by Text Mining and Balanced Indicators. *Data Management Technologies and Applications* 1446, 115-133.

Handfield, R. & Jeong, S. & Choi, T. (2019) Emerging procurement technology: data analytics and cognitive analytics. *International journal of physical distribution & logistics management* 49, 10, 972-1002.

Harriott, J. & Isson, J.P. (2012) *Win with Advanced Business Analytics*. Hoboken, NJ, USA: John Wiley & Sons, Inc

Heijboer, G.J & Telgen, J. (2002) Choosing the open or restricted procedure: a big deal or a big deal? *Journal of public procurement* 2, 2, 187-215.

Holma, A-M., Vesalainen, J., Söderman, A. & Sammalmaa, J. (2020) Service specification in pre-tender phase of public procurement - A triadic model of meaningful involvement. *Journal of Purchasing and Supply Management*.

Holma, A-M., Østensen, M. W., Holmen, E. & de Boer, L. (2022) Market dialogue in public procurement: Buyer-supplier interfaces and relational abilities. *Industrial marketing management* 104, 51-67-

IBM Cloud Education (2021) *Structured vs. Unstructured Data: What's the Difference?* [online website]. [Accessed 15.12.2023] Available: <https://www.ibm.com/blog/structured-vs-unstructured-data/>

Issabayeva, S., Yesseniyazova, B. & Grega, M. (2019) Electronic Public Procurement: Process and Cybersecurity Issues. *NISPAcee journal of public administration and policy* 12, 2, 61-79.

Jansson, K. & von Solms, R. (2013) Phishing for phishing awareness. *Behaviour & information technology* 32, 6, 584-593.

Johnsen, T. E., Mikkelsen, Ole S. & Paulraj, A. (2016) The character and significance of Nordic purchasing and supply management research: A systematic review of the literature. *Journal of purchasing and supply management* 22, 1, 41-52.

Julkisten hankintojen neuvontayksikkö (2016) Mikä on julkinen hankinta? Hankintojen periaatteet. [online website]. [Accessed 30.9.2023] Available: <https://www.hankinnat.fi/mika-julkinen-hankinta/hankintojen-periaatteet>

Julkisten hankintojen neuvontayksikkö (2022a) Ilmoittaminen, HILMA. [online website]. [referred 17.12.2023] Available: <https://www.hankinnat.fi/eu-hankinta/ilmoittaminen/hilma>

Julkisten hankintojen neuvontayksikkö (2022b) Kynnysarvot. [online website]. [Accessed 27.1.2024] Available: <https://www.hankinnat.fi/mika-julkinen-hankinta/kynnysarvot>

Jääskeläinen, J. & Tukiainen, J. (2019) Anatomy of public procurement, VATT. Working paper 129.

Ključnikov, A., Mura, L. & Sklenár, D. (2019). Information security management in SMEs: factors of success. *Entrepreneurship and Sustainability Issues* 6, 4, 2081-2094.

Kim, D. (2017) Cybersecurity strategies for public procurement organizations. *Government Procurement* 25, 3, 18.

Kontio, A., Kronström, S., Kumlin, A. & Mäki, L. (2017) *Julkiset hankinnat : käsikirja*. Edita Publishing Oy.

Koskinen, I., Alasuutari, P. & Peltonen, T. (2005) *Laadulliset menetelmät kauppatieteissä*. Gummerus Kirjapaino Oy, Jyväskylä.

Kähkönen, A. K. (2014) Conducting a case study in supply management. *Operations and Supply Chain Management: An International Journal* 4, 1, 31-41.

Lalic, B., Gracanin, D., Tasic, N. & Simeunović, N. (2022) Process Mining in Public Procurement in Croatia. *Proceedings on 18th International Conference on Industrial Systems - IS'20*, 473-480.

Langseth, M. & Haddara, M. (2021) Exploring data analytics adoption in public procurement: The case of Norway.

Linneberg, M. & Korsgaard S. (2019) Coding qualitative data: a synthesis guiding the novice. *Qualitative Research Journal* 19, 3, 259-270.

Mahanti, R. (2021) *Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools*. Springer Nature Business and Management (R0) eBooks 2021 English/International.

McKevitt, D. & Davis. P. (2015) How to interact, when and with whom? SMEs and public procurement. *Public Money & Management* 35, 1, 79-86.

Mensah, C. & Tuo, G. (2013) Evaluation of Procurement Processes and its Operational Performance in the Public Sector of Ghana: A Case Study of Komfo Anokye Teaching Hospital and Kumasi Polytechnic. *European Journal of Business and Management* 5, 29, 121-132.

Merhi, M. I & Bregu, K. (2020) Effective and efficient usage of big data analytics in public sector. *Transforming Government: People, Process and Policy* 14, 4, 605-622.

Ministry of Economic Affairs and Employment of Finland. (2021a) Public procurement is regulated [Web page]. [Referenced 30.9.2023]. Available: <https://tem.fi/en/public-procurement>

Ministry of Economic Affairs and Employment of Finland. (2021b) Publication of contracts and thresholds [Web page]. [Accessed 30.9.2023]. Available: <https://tem.fi/en/publication-of-contracts-and-thresholds>

Ministry of Economic Affairs and Employment in Finland. (2022) EU and national thresholds. [Web page]. [Accessed 17.12.2023]. Available: <https://tem.fi/en/eu-and-national-thresholds>

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.

Mynarz, J. (2014) Integration of public procurement data using linked data. *Journal of Systems Integration* 4, 19-31.

Nieminen, S. (2016) *Hyvä hankinta - parempi bisnes*. Helsinki : Talentum Pro.

OECD. (2019) *Productivity in Public Procurement. A Case Study of Finland: Measuring the Efficiency and Effectiveness of Public Procurement*. [online document]. [Accessed 4.12.2023]. Available: <https://www.oecd.org/gov/public-procurement/publications/productivity-public-procurement.pdf>

Patton, M. (2015) *Qualitative Research & Evaluation Methods*. SAGE Publications, Inc.

Pawar, P. & Paluri, R. (2022) *Big Data Analytics in Logistics and Supply Chain Management: A Review of Literature*. *Vision (New Delhi, India)*, 2022, 1-20.

Pekkala, E., Pohjonen, M., Huikko, K. & Ukkola, M. (2022) *Hankintojen kilpailuttaminen ja sopimusehdot*. 9th edition. Helsinki, Tietosanoma.

Rafati, L. & Poels, G. (2015) *Towards Model-Based Strategic Sourcing*. *Global Sourcing*, 29-51.

Rejeb, A., Süle, E. & Keogh, J., G. (2018) Exploring new technologies in procurement. *Transport & Logistics: the International Journal* 18, 45, 76-86.

Salo, I. (2013) *Big data: Tiedon vallankumous*. Jyväskylä: Docendo.

Sedkaoui, S. (2018) Data analytics and big data. Hoboken, New Jersey : ISTE Ltd/John Wiley and Sons Inc.

Seiner, R. (2014) Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success. Technics Publications, LLC.

Shaikh, R. & Sasikumar, M. (2015) Data Classification for achieving Security in cloud computing. Procedia Computer Science 45, 493 – 498.

Shao, Benjamin B. M., St. Louis, R.D., Corral, K. & Li, Z. (2022) Best Practices for Leveraging Data Analytics in Procurement. MIS quarterly executive 21, 2, 131-142.

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

Tan, M. & Lee, W-L. (2015) Evaluation and Improvement of Procurement Process with Data Analytics. International journal of advanced computer science & applications 6, 8, 70-80.

Tankard, C. (2015) Data classification– the foundation of information security. Network security 5, 8–11.

Thompson, N., Ravindran, R. & Nicosia, S. (2015) Government data does not mean data governance: Lessons learned from a public sector application audit. Government information quarterly 32, 3, 316-322.

Tuomi, J. & Sarajärvi, A. (2009) Laadullinen tutkimus ja sisällönanalyysi. Kustannusosakeyhtiö Tammi.

Työ- ja elinkeinoministeriö. (2021) Julkiset hankinnat yhteiskunnan tärkeiden kehitystavoitteiden toteutuksessa. [online document]. [Accessed 28.1.2024]. Available: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/163164/TEM_2021_34.pdf?sequence=1&isAllowed=y

Valtiovarainministeriö. (2020) Suomen julkisten hankintojen tilannekuva. [online document]. [Accessed 4.12.2023]. Available: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/162171/VM_2020_25.pdf?sequence=1&isAllowed=y

van Ooijen, C., Ubaldi, B., & Welby, B. (2019), A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance. OECD Working Papers on Public Governance 33, 1-59.

Van Weele, A. (2014) Purchasing and Supply Chain Management: Analysis, Strategy, Planning and Practice. Melbourne, Cengage Learning Australia Limited.

Voss, C., Tsiriktsis, N. & Frohlich, M. (2002) Case research in operations management. International journal of operations & production management. 22, 2, 195–219.

Wang, G., Gunasekaran, A., Ngai, E. W.T. & Papadopoulos, T. (2016) Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International journal of production economics 176, 1, 98-110.

Weber, R. (2011) Basic Content Analysis. SAGE Publications, Inc.

World Bank. (2022) Using Data Analytics in Public Procurement: Operational Options and a Guiding Framework. [online document]. [Accessed 7.12.2023]. Available: <http://hdl.handle.net/10986/37467>

Öhman, M., Arvidsson, A., Jonsson, P. & Kaipia, R. (2021) A knowledge-based view of analytics capability in purchasing and supply management. International journal of physical distribution & logistics management 51, 9, 937-957.

Appendix

Appendix 1. The interview invite (translated from Finnish to English).

Hi, I'm doing a study about procurement data management for VTT. The purpose is to map:

1. what kind of data is generated in each phase of the procurement process

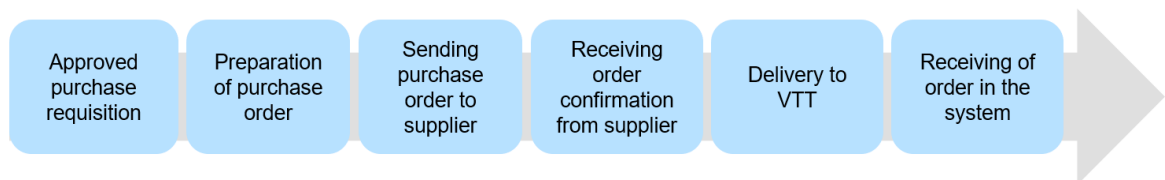
2. what this data is used for

3. in which system the data is

4. who owns the data and who knows more about the data

For the purpose of this work, I will be interviewing members of the procurement team one by one during the summer and early autumn, would this time be suitable for the interview?

Appendix 2. Purchasing/post-tender process description for interview.



Appendix 3. Tendering/pre-tender and tender process description for interview.

