TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Liisbeth Laasik 211977IVGM

DATA GOVERNANCE IN THE ESTONIAN GOVERNMENT SECTOR: THE MOTIVES AND STRATEGIES FOR IMPLEMENTATION

Master's thesis

Supervisor: Innar Liiv
PhD

Tallinn 2023

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Liisbeth Laasik 211977IVGM

ANDMEHALDUS EESTI VALITSUSSEKTORIS: RAKENDAMISE MOTIIVID JA STRATEEGIAD

Magistritöö

Juhendaja: Innar Liiv
PhD

Tallinn 2023

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Liisbeth Laasik

06.05.2023

AKNOWLEDGEMENTS

I am sincerely grateful to my research supervisor Innar Liiv, for his guidance and valuable feedback throughout the thesis writing process.

I would like to thank Kuldar Aas for his expertise and collaboration, Ants Sild for his valuable suggestions and questions, and Kairi Rais for her support and introducing me to the topic.

I would also like to extend my gratitude to Sadok Ben Yahia, for his expertise and feedback during the early stages of the thesis.

Finally, I would like to express my gratitude to all the participants who contributed to the study by sharing valuable insights and experience during interviews. Thank you Evelyn Liivamägi, Viola Murd, Kaupo Heinma, Ulla Saar, Liina Põld, Mariin Ratnik, Olavi Seisonen, Margit Martinson, Alvar Pihlapuu, Reimo Tarkiainen, Katre Seema, Liisi Lillipuu, Karl Viilmann, Sirli Niibo-Tamm, Aivi Saar, Adele Johanson and Hendrik Valgemäe.

Abstract

Increasing volumes of data are putting pressure on the public sector to systematically focus on data governance to ensure data quality, security and accessibility, as well as to enhance the maximum value of data throughout its life cycle. The principles of data reuse and interoperability require a unified and coordinated approach to data governance from the Estonian government sector, as a result of which several institutions need a change in mindset, practices and processes.

The aim of this thesis is to identify and describe the value that data governance brings to the Estonian government sector and map strategies on how to implement data governance framework, processes and activities in the organization. In the course of 15 interviews, it became clear that the Estonian government institutions are in the early stages of implementing data governance practices and therefore finding the right motivation and reason for implementing changes is essential. The findings indicate 14 points of view that demonstrate the value of data governance as well as the motives for implementing its framework, processes and activities in government organizations. The thesis also outlines the steps that need to be followed to successfully implement data governance in a government organization, while drawing attention to the existing shortcomings that can be addressed by national policy makers.

This study contributes to increasing the Estonian public sector data governance knowledge and supports the successful implementation of new data governance practices, which are in line with guidelines of the European Commission (EC).

Keywords: data governance, data management, public sector, government sector

This thesis is written in English and is 66 pages long, including 5 chapters, 3 figures and 2 tables.

Annotatsioon

ANDMEHALDUS EESTI VALITSUSSEKTORIS: RAKENDAMISE MOTIIVID JA STRATEEGIAD

Kasvavad andmemahud avaldavad survet avalikule sektorile süstemaatiliselt keskenduda andmehaldusele, et tagada andmete kvaliteet, turvalisus ja kättesaadavus ning suurendada andmete maksimaalset väärtust kogu nende elutsükli jooksul. Andmete taaskasutuse ja koostalitlusvõime põhimõtted nõuavad Eesti valitsussektorilt ühtset ja koordineeritud lähenemist andmete haldusele, mille tulemusena vajavad mitmed asutused mõtteviisi, praktikate ja protsesside muutust.

Käesoleva lõputöö eesmärk on tuvastada ja kirjeldada väärtust, mida andmehaldus Eesti valitsussektorile toob, ning kaardistada strateegiad, kuidas andmehaldusraamistikku, protsesse ja tegevusi organisatsioonis juurutada. 15 intervjuu käigus selgus, et Eesti valitsusasutused on andmehalduspraktikate juurutamise algusjärgus ning seetõttu on muudatuste elluviimiseks õige motivatsiooni ja põhjuse leidmine hädavajalik. Tulemused toovad esile 14 seisukohta, mis näitavad andmehalduse väärtust ning selle raamistiku, protsesside ja tegevuste rakendamise motiive valitsusorganisatsioonides. Lõputöös tuuakse välja ka sammud, mida järgida andmehalduse edukaks juurutamiseks valitsusasutustes, juhtides samas tähelepanu olemasolevatele puudujääkidele, mida riiklikud poliitikakujundajad saaksid lahendada.

Käesolev uuring aitab kaasa Eesti avaliku sektori andmehaldusalaste teadmiste suurendamisele ning toetab uute andmehaldustavade edukat rakendamist, mis on kooskõlas Euroopa Komisjoni suunistega.

Märksõnad: andmehaldus, avalik sektor, valitussektor

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 66 leheküljel, 5 peatükki, 3 joonist, 2 tabelit.

List of abbreviations and terms

AI	Artificial intelligence
DDDM	Data-driven decision making
EC	European Commission
EU	European Union
IT	Information technology
MFA	Ministry of Foreign Affairs of Estonia
MKM	Ministry of Economic Affairs and Communications of Estonia
ML	Machine learning
OOP	Once-Only Principle
RQ	Research question
SQ	Sub-question

Table of contents

List of figures

List of tables

1 Introduction

Data plays an important role in the society; it is like fuel that keeps the system running. It has the potential to give meaningful insights into the past, present and future. The data itself is of little value; it needs context to turn into useful information [1]. In an age where the rapid development of information technology (IT) significantly affects society and its functioning, data is produced, processed and used more than ever, which creates new challenges in handling large volumes of data and the need for new processes, frameworks and policies. Data governance, like any governance, is the process of taking control and authority over data and overseeing that it is managed accordingly [2].

The work of government sector, such as policy-making and public service provision, influences the whole society, which makes it of great interest that data collected by the public sector is governed and used for the benefit of the society. Whether it is security, quality or privacy issues, it is expected that the data is handled ethically and properly. Data governance with processes, frameworks and policies plays an important role in its successful implementation.

1.1 Research problem and purpose

Estonia's Digital Agenda for 2030 highlights data governance and reuse of data as one direction to ensure the development and sustainability of digital government [3]. The national strategy brings focus on increasing interoperability in the Estonian public sector, which becomes possible when more emphasis is placed on data and its governance.

In 2022, the Ministry of Economic Affairs and Communications of Estonia (MKM) conducted a data strategy survey among the government institutions of Estonia to map the current situation in various data areas, which included the assessment of data governance and its current coordination in the organizations. 37 out of 55 organizations participated in the survey, and the findings are a valuable input for national coordination and development in the field of data governance, management, open data and other data developments like artificial intelligence (AI) initiatives. The results of the survey

highlighted that many of the government organizations have room to improve the internal processes of the institution to develop frameworks, processes and activities related to data governance [3]. According to the Ministry of Foreign Affairs of Estonia (MFA), the lack of data governance practices lies both in understanding the additional value that data governance brings to a government organization and in the knowledge of how to start implementing data governance.

The aim of the thesis is to identify and describe the value that data governance practices bring to the Estonian government sector and map strategies how to implement data governance practices or programs in the government area. The research will contribute to increasing the knowledge on data governance, especially in the Estonian government sector and brings into focus its practical implementation.

1.2 Research questions

The thesis focuses on two main areas: identifying the value that data governance brings to a government institution and finding the strategies to get started with data governance. Researching these areas will support Estonian government sector to improve its data governance capabilities, which is aligned with national and EC directions.

To achieve this objective, the author proposes the following Research Questions (RQ) and Sub-Questions (SQ):

RQ1. What is the value of data governance in the Estonian government sector?

- SQ1. What motivates Estonian government institutions to implement data governance practises?

RQ2. What are the strategies for implementing data governance practices in a government organization?

- SQ1. What kind of organizational changes are needed to improve the state of data governance?

1.3 Thesis outline

The thesis is divided into five chapters. The first chapter gives an overview of the research purpose, problem and provides the research questions from which the thesis is based. The second chapter introduces the topic of data governance from a theoretical perspective and provides concepts, frameworks and theories that set the literary foundation for the thesis. The following chapter describes the methodological approach, introducing the used qualitative method for data collection and data analysis. The fourth chapter focuses on presenting and analysing the collected data and drawing conclusions based on it as well as bringing out the research limitations. The fifth and the final chapter summarizes the thesis and brings out the most important findings and future directions.

2 Theoretical framework

This chapter gives an overview of the existing concepts, frameworks and theories related to data governance on which the thesis is based. In addition, the perspectives of European Union and Estonia are described, which create relevant context for data governance due to the aim of the thesis.

2.1 Data governance principles and framework

To comprehend data governance, the concept of data must first be understood. Piece of data separately looked has little value, it needs context to be understood and giving it context is what transforms data to information [1]. Data is also seen as a representation, which means that it gives context to other things than themselves, whether it is object, concepts or even other data [4]. The term has received more attention with the popularization of IT, which enables to collect, process and produce large amounts of data in digital form.

Data can be divided into two different groups: structured and unstructured data. It is said that about 80% of data in an organization is unstructured and it is not easily analysable like structured data. Examples of unstructured information are emails, memorandums, videos, audio etc. Structured data, like data in databases can be easily accessed, processed and analysed due to technological automation tools available [5]. This is important to distinguish since data governance could be applied to all kinds of data, but the approach on structured and unstructured data may be different due to the nature of data and the goal of data governance.

There is no one right definition of data governance, the term is defined in numerous ways depending on the source. The Data Governance Institute defines data governance as ‘the exercise of decision-making and authority for data-related matters’ [6]. Ladley [7] elaborates that ‘data governance is the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of

information assets'. While Dama International [2] describes data governance 'as the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets'. The key takeaway of these definitions would be, that data governance is about taking control and responsibility of the data assets in possession by creating rules, procedures and structure for managing the data.

The definitions of data governance and data management are often intertwined, as they are related concepts but have different objectives. Data management represents the practical operational activities and processes of managing data as an asset and enhancing the value of it throughout data lifecycle [2]. It is said that there should be separation of responsibilities between those who oversee the management of data and those who govern it [7]. From an organizational perspective, both data management and governance are important data to be treated as a valuable asset.

Even though data governance can be defined in one sentence then the definition leaves room for interpretation. Data is everywhere around us and organizations as well as people are consuming, producing and using it daily. It can be hard to grasp what is meant by 'controlling the data' since data governance covers such broad areas and concepts.

Brous, Janssen, and Vilminko-Heikkinen [8] carried out a systematic review of literature to gather and highlight some of the most important principles of data governance, which gives a good overview of the areas that data governance concerns (Figure 1). The research findings are divided into four main principles: organization, alignment, compliance, and common understanding.

Under 'organization' the listed principles support the understanding that data governance has an organizational dimension, and the data governance principles should align with the organizational goals, activities, structure, and hierarchy. The 'alignment' sub-principles bring out that data governance should be intertwined in the organization's everyday activities whether its business processes, strategies or policies and procedures. The 'compliance' principles highlight the importance of controlling and defined authority in effective data governance which fosters accountability in the organization. 'Common understanding' brings focus on creating a general understanding of what is data governance and why organization should need it and its' sub-principles highlight the importance of standardization [8].

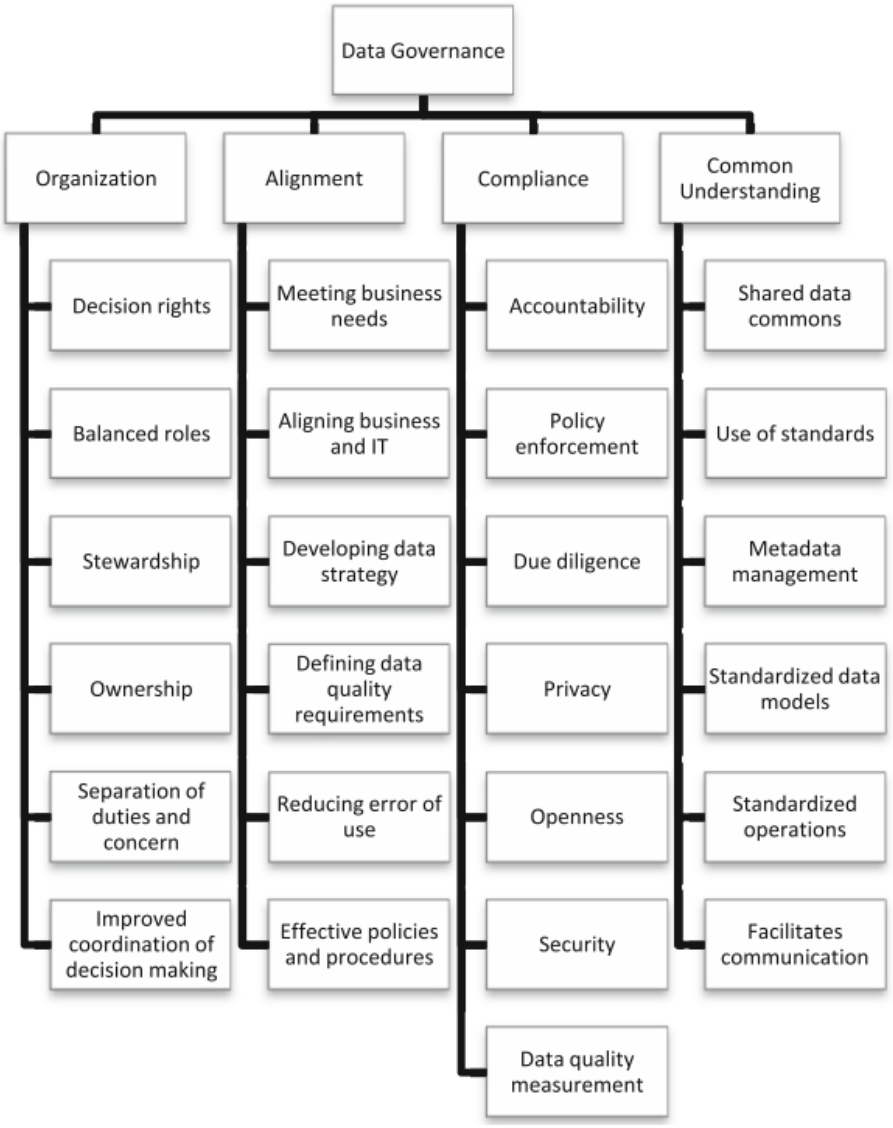

Figure 1. List of key principles of data governance [7].

The long list of principles highlights two important aspects. Firstly, data governance covers many different areas of expertise that are arduous to tackle all at once. Secondly, since the list is so long, then actually implementing all these principles is a continuous multifaceted process which requires cooperation and willingness from all parties in the organization, since the principles have an impact on the whole organizations' workflow.

Building on the principles of data governance Abraham, Schneider, and Brocke [9] identified the main components of data governance to form a framework which would allow to approach data governance in a structured way (Figure 2).

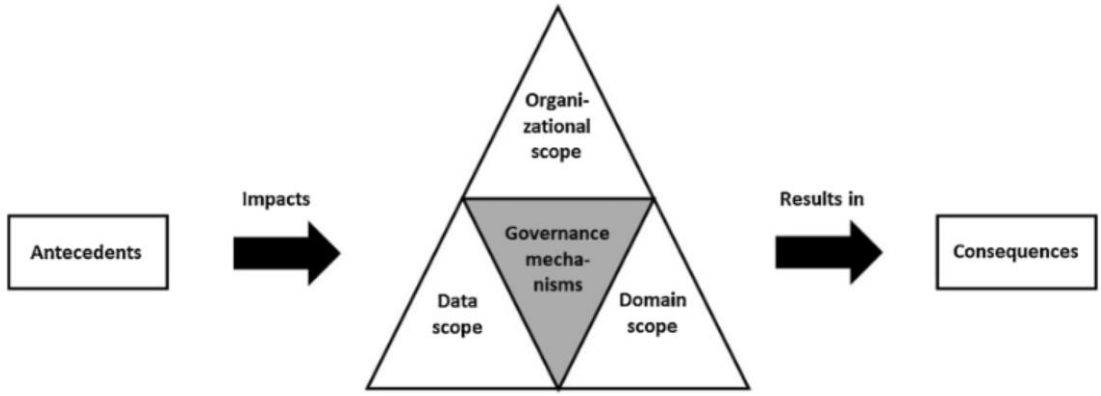

Figure 2. Conceptual framework for data governance [8].

According to the framework, there are antecedents that impact the adoption of data governance, which precede before the implementation [10]. They are categorized as internal antecedents, like organizational strategy and culture, IT architecture, decision making authority, management support and external antecedents, like legal and regulatory requirements, industry, and country. The antecedents impact data governance implementation in an organization [9].

The main building block of data governance is governance mechanisms, which are divided between structural, procedural, and relational mechanisms. Data governance roles and responsibilities as well as placement of decision-making authorities fall under structural mechanisms. Procedural mechanisms are in place to ensure that data itself is managed properly, according to agreed standards, policies, processes etc. [11]. While relational mechanisms facilitate communication, training, and coordination of decision-making in the organization between parties. The data governance framework is supported by an organizational scope, which aims to determine whether data governance is project-/organizational-level or inter-organizational. The data scope highlights to need to focus on the type of data that data governance programs impact, like structured or unstructured data, big data etc. [12]. The next building block in the framework is domain scope, which like data scope is set on choosing the focus areas for the data governance program. The main domains are data quality; data security; data architecture; data lifecycle; meta data; and data storage and infrastructure [13].

The last building block of the framework is consequences of data governance, which are either intermediate performance effects that immediately show the positive effect of data

governance or risk management, as data governance activities can also prevent risks from happening [10]. This data governance framework, which is based on comprehensive literature review, gives a structured conceptual overview on how to approach designing and implementing data governance programs and activities in organizations.

2.2 The value of data governance

Data governance initiatives and new practices require resources, whether human or financial, that support the changes in the organization. This means the added value of data governance must be well demonstrated and explained to the top management as well as the rest of the organization. Describing the value can turn out to be quite difficult, since data governance covers a broad area of concepts that can be linked to all kinds of information in the organization. Ladley [7] brings out three tangible benefits of data governance from the private sector perspective. These are 1) increased efficiency, 2) growth of direct business drivers (e.g., revenue, market share) and 3) reduced risk. Looking at these benefits from a public sector perspective, efficiency and risk management are the most applicable.

Efficiency can be expressed in the ability to respond to external changes and threats faster and more effectively. These external changes may be natural disaster, new technologies, public health security etc. [14]. The value of data governance will show when in these circumstances the public administration has adequate and quality information available to make decisions in the best interest of the country and its' citizens. Efficiency may also express in improved processes inside the organization, e.g., managing development projects more efficiently which rely on available data, its' accessibility and quality, as well as the ability to respond better to regulatory compliance [2].

Reduced risk is also one of the ways to show benefit in data governance since its focus is keeping data safe and of high quality according to the rules and regulations. Risks associated with poor data governance include the misuse of private data and therefor violation of privacy, making incorrect decision which are based on inaccurate data, data breaches and losing data if it is not stored and managed properly [7]. If these risks materialize, the consequences can be fines, lawsuits, financial loss, and damage to the organization's reputation. Costs related to non-governance of data or poor data quality can lead to unnecessary use of resources, which should be taken into account when

explaining the advantages of data management and highlighted as a savings opportunity [7]. It is widely known that data can be a valuable asset, which is why news of various data breaches is quite common and concerns both well-known companies and the public sector.

With growing amounts of data in the public sector there are also expectations that the data would be used in the best possible way that would support honest and transparent decision making. Evidence-based decision making represents an objective process, where gathering and analyzing data is with high importance that would lay basis for trustworthy body of facts, rather than individuals' own interest or personal experiences [15]. Evidence-based decision-making leads to evidence-based policy, which essentially means that policy design and implementation is supported by the best available evidence from research. This approach is usually applied in addition to other factors influencing policy making, such as resources, personal experiences, judgments, traditions, etc. [16], but the difference is determined by how much evidence is considered among other factors.

The downside of evidence-based decision making is that even the information derived from data can be biased if it is gathered or analyzed with pre-determined bias whether it is done consciously or not. In order to avoid it, the data and evidence must be applicable to the context and in high quality. In addition, the policy makers should have access to competences which could evaluate the technical or issue bias in the evidence [17]. Therefore, even if evidence-based decision making is a growing trend that brings accountability and transparency to the public administration then it must be done consciously, bearing the possible risks in mind.

Another term associated with evidence-base decision making is data-driven decision making (DDDM), meaning that decisions are based on the analyses of the data, rather than purely personal experience and intuition [18]. DDDM builds on the traditional decision-making process by expanding the data usage in additional ways. This results in employees and decision-makers needing also additional competences for adopting DDDM, like data literacy, critical thinking, data analytical skills, and innovativeness [19]. Meaning that for adopting more evidence-based and data-driven decision-making processes, the people at its core need to adapt as well.

2.3 Implementing data governance

Data governance is expressed in practical activities that need to be implemented in an organization for anything to change. Alhassan, Sammon, and Daly [20] conducted a literature review of activities reported in scientific and practice-oriented publications, for data governance. Their extensive research summarized the main activities in different areas of government for five decision domains (Figure 3).

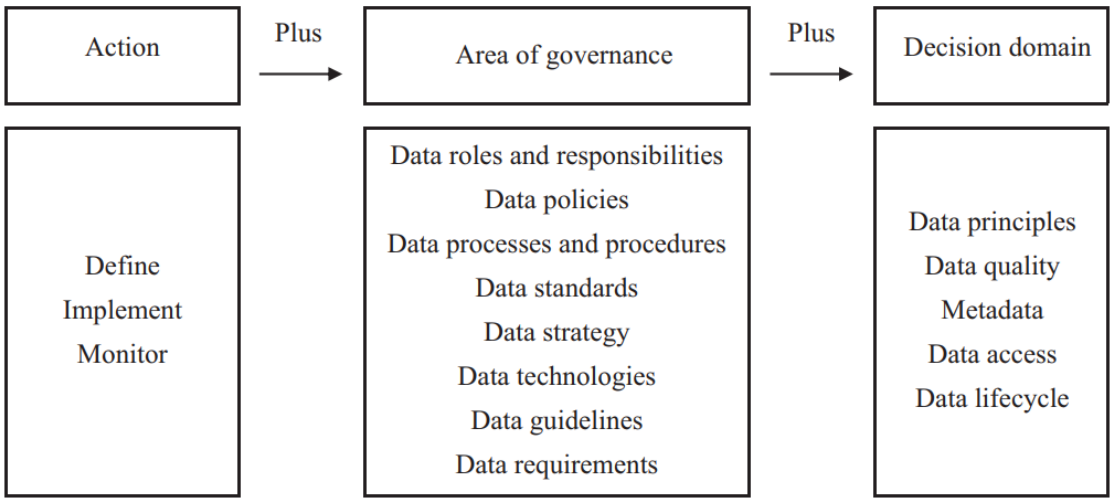

Figure 3. Overview of data governance activities [20]

The three described actions, ‘define’, ‘implement’ and ‘monitor’, can relate to eight different areas of government (roles and responsibility, policies, processes and procedures, standards, strategy, technology, guidelines, and requirements) that apply to five decision domains [20]. The referred decision domains (data principles, data quality, metadata, data access and data lifecycle) originate from Khatri and Brown [13] research that adds an area of focus to the data governance actions. Implementing these activities does not only need resources but also a strategic approach, since they broadly affect the organization and its work processes.

2.3.1 Roles and responsibilities

Data governance initiatives start with people who lead the change in the organization. The specific roles and number of people who work with data governance can differ depending on the size of the organization and the goal, but according to literature, there

are some roles and responsibilities that are ‘must be’ when implementing data governance.

Firstly, there’s a need for an executive sponsor in the organization, ideally a member from the top management, who has authority and decision-making rights and provides the strategic steering for the data governance program in the organization [21]. From hierarchical viewpoint, the next important role is chief data steward/officer, sometimes referred as data governance manager who is responsible for the running of data governance program in the organization, which includes documentation, communication, and enforcement [22]. In the organizational structure, the chief data steward could place in the middle management [23]. From the operational level, data governance needs business and technical data stewards. Business data stewards are the ones who know the meaning of the data, what the data should represent, and the business rules associated with the data [22]. They create a bridge between the business and technical side and ensure that one understands and supports the other. The technical stewards work with specific IT systems or applications where the data is held, processed, or analyzed. They have the knowledge on how the data is created, stored and what kind of changes it goes through in IT systems [22]. These four roles 1) executive sponsor; 2) chief data steward, 3) business data steward and 4) technical data steward, are necessary in the organization for starting a successful data governance program. These roles don’t have to be covered by a whole job position but can be fulfilled partly.

2.3.2 Strategies for implementation

For data governance to become a normal everyday activity, the organization needs to change, whether it is the organizational structure, processes, polices or competences. Even though change is a normal process in an organization, it still faces resistance at some level. This is no different for data governance initiatives [24].

This is why the implementation process of new initiatives should be done strategically, to face less resistance during the process. A. Cave [25] explored in her doctoral study the strategies which would support implementing data governance practices in an organization. Firstly, having boards or committees in place is a way to make sure that data governance practices align with the organizational objectives [25]. They provide oversight and are crucial in creating data governance policies and procedures early on which take into account the winder landscape of the organization [26]. Secondly, strategic

communication is necessary, which includes having defined roles in place and educating, training the people in the organization. In addition, compliance with laws and regulations is a strategic way to start implementing new data governance practices, whether it is coming from international agencies, local government, or other institutions whose regulations need to be followed. Fourthly, gaining stakeholder buy-in is necessary, which can relate to having an oversight board/committee with great involvement, but also it relates to piloting new practices in smaller groups or units, which would help people adapt smoother to changes. The last emerged theme from the study was benchmarking and standardization, which in practice means comparing to and learning from other institutions from similar fields. This requires effort in collaboration but could result in better interoperability and development on a wider scale [25]. These five strategies could help organizations implement data governance programs with more ease and efficiency.

The implementation process of data governance practices can also be looked at from an organizational change perspective, when considering top-down and bottom-up approaches. Top-down illustrating that change is initiated by the top management [27] and bottom-up meaning that need for change comes from middle management [28]. To successfully implement data governance practices with top-down approach, it is necessary to have clear objectives in place that link data governance to other organizational activities, e.g., having a strategy in place which highlights the need for better data governance. This would transfer the importance of the topic to all the departments and units, which activities should proceed from the overall strategy [29]. A top-down approach means that the top-management sets the focus, and the organization follows.

The bottom-up approach focuses on the existing processes and activities in the organization and understanding how data governance relates there. By analyzing through some of the critical processes in the organization, the goal is to define the roles and responsibilities linked to it as well as the rules and principles that should be followed. Analysis, with the aim to achieve efficient and high-quality processes, helps to determine which aspects of data governance are incomplete. Whether it is missing competences and roles or insufficient policies and procedures in place [29]. The bottom-up approach is more tailored for improving some of the aspects related to specific processes but has obstacles implementing new data governance practices on a wider scale in the organization.

2.4 Big data and artificial intelligence

Haenlein and Kaplan [30] have defined AI as ‘a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation’. The High-Level Expert Group on Artificial Intelligence appointed by EC [31] have defined AI as ‘systems that display intelligent behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals’. Both definitions emphasize important aspects of AI, such as data interpretation, goal orientation, and adaptability to partial autonomy. Due to the increase in computing power and big data, AI has become more accessible and practicable than ever before [30]. Big data can be described as large amounts of different type of data collected in different formats [32] which also affects the public sector data governance programs and policy [33]. Machine learning (ML) as a branch of AI can have practical implementations in data governance, e.g., using it for organizing, standardizing and comparison of data with criteria, which are activities done for increasing data quality [34]. Additionally, in finding similarities and patterns in data and learn from this process automatically [35]. The quality of AI systems in the field of data governance and beyond relies heavily on the quality of the data inserted, which often needed to be cleaned, labeled, or tagged before the input to AI systems [36]. AI solutions, based on big data or not, are only as good as the data they are based on. It creates a bridge between data governance and AI solutions, as one depends on the other.

2.5 Context of European Union and Estonia

The EC is the responsible body in the European Union (EU) for proposing, managing, and enforcing new legislation across the EU [37]. The laws, polices, and strategies cover a wide range of topics, including data-related matters. The European Strategy for Data is valid from 2020 and sets the direction on creating a single market for data in the EU to ensure the competitiveness and sovereignty of the Union. From the strategy derives the Data Governance Act, which is in force since May 2022 and the Data Act which is currently in proposal status [38]. The Data Governance Act, which relates the most to the thesis topic, has a goal to make more data available in the EU by creating processes and structures that promote data sharing and re-using for available data that is held by the public sector. This applies to data that can be made accessible by categorizing it as open data as well as data that is private or protected but could have a valuable impact if data

sharing is facilitated in a controlled and ethical way. In practice this means that the Data Governance Act sets measures and mechanisms in place for [39]:

1. safe data sharing in the EU for different sensitivity type of data,
2. simplified ways for citizens and businesses to share their data,
3. ensuring trustworthiness of data sharing intermediaries,
4. enabling finding and using the right data for different purposes.

Enforcing these measures should support the goal of making more data available in the EU as well as contributing to making the EU a single digital market for data. The responsibility falls to the EU member states to adjust their legal environment, processes, and technologies according to the EU legislation. Currently, with the Data Governance Act there is a 15-month grace period, meaning that the regulation will become in force in September 2023 and in 2025 EC will carry out an evaluation to see the results and outcomes of the Act [40]. The EU legislation regarding data highlights the importance of data sharing and re-using, especially in the context of public administrations. To reach these objectives, the public sector organizations must first have a good understanding of the data their holding and how it is governed, otherwise the trustworthiness, security and simplicity is not met in practice even if technology would support it.

In 2017 the EC launched an initiative called the Once-Only Principle (OOP) with the aim to reduce EU citizens and businesses obligation of data submission to the public administration. If a data is already submitted to a public sector organization once then it should not be asked again, but it is expected that the public administration exchanges the data if it is necessary. The goal of this project is not only to change the way how EU member states collect and exchange data in their countries, but also to simplify data exchange cross-borders [41]. So far EU member states have initiated pilot projects to implement the principle, but the cross-borders implementation is still quite limited. The OOP contributes to creating a single digital market for data in the EU [42]. The OOP guides the way data should be collected and shared across the EU and affects every member state’s public administration, as well as data governance principles and practices in place.

In Estonia, data governance is coordinated by the Statistics of Estonia, which is regulated by the Public Information Act and the Official Statistics Act [43]. They work in close

collaboration with different public sector institutions, including MKM who has a specialized unit dealing with the field of data and its' development programs, including data governance initiatives. Since the development of Estonian public sector data governance practices requires the cooperation of many organizations, the activities and desired results are described in the national data governance action plan. The latest version currently available is from 2021 till 2022. The focus areas of the action plan are [44]:

- crowing the competences of data stewards,
- developing data re-use opportunities and activities,
- improving data quality principles,
- regulation of data related legislation and developing data description tools.

The named focus areas and actions based on them have an impact on the entire public administration of Estonia. The development of the data governance action plan and other necessary guidelines is carried out in cooperation with various public sector organizations in Estonia. For this purpose, the policy makers launched a data governance steering group in 2022, which includes representatives from the field of data from various government institutions. In addition, a data panel has been launched by MKM, to coordinate the use of data in national ongoing and planned projects to find and create more areas for cooperation [3].

The action plan is supported by different guidelines that are applicable to all public administration in Estonia [45]:

1. **Data governance framework**, which is an extensive document published in 2020 with the purpose to describe how data governance could be organized in an organization, by describing the roles, responsibilities, processes, and activities.
2. **Data description guidelines**, which draw together principles to be followed when describing data and the corresponding roles and responsibilities. Data description standards are also attached, which is a list of mandatory and recommended description elements for the public administration.
3. **Data quality guidelines**, which describe the criteria and principles to follow when managing, setting rules, and measuring data quality.

One central tool offered for better data governance is an application called RIHAKE, which purpose is to make data more accessible and reusable. The main functions include describing databases and classifiers, creating dictionaries, and sharing necessary information to other systems [46].

In general, the basic knowledge and tools regarding data governance are described and available for the Estonian public administration, the question is the involvement of different institutions and the implementation of new practices.

3 Methodology

To identify and describe the value that data governance has on the Estonian government organizations and what kind of strategies are used to implement new practices, the qualitative research method was adopted as the methodological approach. The purpose of qualitative research is to gain an in-depth understanding of the concepts and behaviours. It is focused on exploring the 'how' and 'why' of situations and systems by collecting and analysing data that is non-numeric [47]. Considering the aim and research questions of the thesis and the interpretation that data governance is about taking control and authority over data, the qualitative method was considered as the best approach.

3.1 Data collection methods

The perceived value and importance of data governance is especially a topic where qualitative method is necessary for gaining a profound understanding, therefore 15 in-depth interviews were conducted with different government representatives. This method of data collection is relevant when the research is focused on exploring rather than confirming [48]. The interviews were semi-structured leaving room for follow-up questions based on participants answers, changing the order of the questions if necessary and overall flexibility.

The 15 interviews took place from March 10 to April 10 and the interviews lasted from 40min to 1h10min. Most interviews were one-on-one, except for two cases, data governance representatives of Ministry of Social Affairs and MKM, in which case a group interview was conducted, since two interviewees participated in one interview. The interviews took place in Estonian since all the participants were Estonian speaking, therefore the transcripts and initial analyses was also done in Estonian language. The questionnaire, analyses and results were all translated to English.

From the 15 conducted interviews, 5 took place face-to-face, 9 online and 1 in hybrid format. The participant choice of environment was considered and if the choice was left open, then the author preferred face-to-face interviews for better capturing verbal and

non-verbal signs. All the interviews were recorded with participants consent and it was asked from each participant whether their name mentioned in the thesis is acceptable to them or not. Since all the participants gave verbal consent then interviewees names and titles are non-confidential in the thesis (Appendix 1, Table 2).

3.1.1 Sample selection

The author conducted semi-structured interviews with 8 different administrative areas of Estonian government. The target group of interviewees was divided into two following categories:

1. A representative of the administrative area, who is the head of a policy field at a ministry, in the position of undersecretary. A field where it is known that there is a lot of valuable data or where a large impact of data is already felt.
2. A representative of the administrative area from the support unit side, someone who is coordinating data governance in an area of government, e.g., data governance manager, advisor or an undersecretary of a ministry under whom data governance coordination falls.

From the first category, interviews were conducted with the undersecretaries of 6 different ministries and from the second category, the author conducted 9 interviews with 11 data governance representatives. The participants for the interviews were selected by purposive sampling, meaning that they were chosen by predefined aspects determined by author's own knowledge, literature review and anecdotal information from experts involved in the field of data governance [49]. An expert, whose expertise and knowledge were considered when identifying data governance experts in the Estonian government sector, was Kuldar Aas, the Data Governance Programme Lead at MKM. Since he is one of the organisers of the national data steering group, involving data governance, that brings together data governance experts from different public sector institutions, his expertise was valuable addition to participant selection process.

After careful selection of the participants, an introductory email was sent out to all the representatives, asking them to take part in the research. This resulted in a total of 15 interviews where representatives indicated a desire to participate, and 8 cases where the participant could not be contacted, or the time was not suitable, or an interview was not

preferred. Interview questionnaire was not shared beforehand, however if the participant requested it, then the questionnaire was sent (5 requests).

3.1.2 Developing the interview guide and questionnaire

For conducting interviews, an interview guide was developed (Appendix 1), which set the structure for the interview plan, on-set protocol and interview main topics and questions. The author composed two different set of interview questions, since there were two interview target groups. Policy field representatives had 11 questions and data governance representatives 12 questions. The questions were mostly open-ended questions and often continued with follow-up questions depending on the participants' answer. Most of them were also opinion questions, that guided the participant to express their thoughts about the topic, or knowledge questions, that directed the participants to share factual information [49]. Due to the nature of semi-structured interview and open-ended questions, the natural flow of the interview was different with each participant, resulting in sometimes changing the order of the questions or rephrasing them.

Both interview question sets had three main themes under which the questions were distributed. For the undersecretary of a policy field, the questionnaire themes were:

1. Goals and focus of the field, most important data.
2. Data use cases, data-driven decision-making, data quality and reliability.
3. Data related problems, improvement areas and perspective.

For the representative of the data governance coordinating support unit, the questionnaire themes were:

1. The importance of data and its governing
2. Current state of data governance and established processes.
3. Data governance perspective.

From the author's point of view, the interviews were conducted in a safe environment that allowed the participants to freely share their opinions and knowledge about the topic.

3.2 Data analyses methods

All 15 conducted interviews were first transcribed for analysis, using available Estonian speech recognition application, which automatically transcribed Estonian speech into text [50]. After automatic transcription, the author went through all the recordings and transcriptions, and corrected wording and typos. Since the period of collecting data lasted for a month, then the author started ongoing analyses during the interviews period for noting down emerging thematic patterns in the collected data [51]. All the thematic patterns were documented since they reflect ideas and meanings, which are relevant for analysis. All together 12 themes emerged from the interviews, which enabled a thorough thematic analysis to be done:

1. The importance of data
2. Most important use cases of data
3. Quality and trustworthiness of data
4. The importance of data governance
5. The benefit and added value of data governance
6. Existing processes, procedures and activities related to data governance
7. Problems and challenges related to data governance
8. Using data in the decision-making process
9. Roles and structural units involved
10. Competences and people
11. Keeping up to date with data governance
12. Perspective of data governance

The interviews were conducted with two different target group representatives, an undersecretary of a policy field and a representative of data governance coordinating support unit, to get different viewpoints to importance of data and its governance. This sample selection gave the chance, at same cases, to thematically analyse answers by different target group representative and point out the similarities and differences.

4 Research results and findings

This chapter focuses on analysing and interpreting collected data from the interviews and forming conclusions and recommendations based on the research problem and research questions.

4.1 Results of in-depth interviews

The aim of the interviews was to gain an in-depth understanding if the Estonian government institutions knowingly engage in data governance activities and to what extent; what kind of value and importance is perceived in data and data governance; what are the existing frameworks, structures, processes supporting it; and what are the challenges, problems and future perspective of the topic. This chapter focuses on presenting data collected from the interviews.

The importance of data

All the representatives from the data governance supporting unit agreed that there is an understanding in the respective area of government that data has value. This knowledge has increased in the last years, during which time several government institutions, where the interviewees work, have created new positions aimed at dealing with data in a targeted manner. Even though data value is recognized then the mindset of treating data as an asset has not really been implemented, meaning that data has not been interpreted into financial value and managed accordingly. Rather there is an understanding that data plays an important role in providing quality services and it supports the decision-making process. The fact that data can provide valuable information is recognized till the top management and in many cases, it is the management who carries the principle that decisions need to be evidence-based and data is like foundation in the public sector, meaning that good data produces good results.

The undersecretaries representing the policy field in the ministries also highlighted the importance of data in their respective field. All were able to name many important data sets that support them in either daily work, policy making or decision-making process.

All interviewees mentioned that the value of data for the policy side only exists when it is placed in the right context, which means that data without meaning does not support them, and a data analytics layer is needed to place the data in the right context depending on the need.

Most important use cases of data

To get the interviewees in the mindset of how the existing data is used in their field, it was asked of them to introduce some of the most important use cases of data. Even though the interviewees approached this question very differently, some started explaining their everyday activities, others approached from strategic point of view, all the answers had common lines and similarities. By both data governance representatives and policy makers, the most common data use cases highlighted were fulfilling the objectives of the institution or a specific field, which directly relates to another recurring answer, which is setting metrics in place for these objectives and monitoring the fulfilment of these metrics. In addition, policymaking using data was cited as a key use case, while the undersecretaries brought out that policy impact evaluation as one of the most difficult challenges. Day-to-day operational work and strategic planning, including budget decisions, were also mentioned as an important data use case.

Data governance representatives additionally brought out making data publicly accessible as open data an important use case and using data for research and innovation, either in the public sector or by academia or private sector.

Quality and trustworthiness of data

From all interviewees were asked questions regarding data quality and it was a topic where everyone could participate and give examples from real life, some in more detail than others. The representatives of the data governance unit had different assessments of the current situation of data quality in the administrative area. 4 out of 9 interviewees assessed that the current data quality is generally good, 1 mentioned that it is rather bad, and 4 others were not able to assess it. Only 2 interviewees mentioned that they have had some kind of quality control assessment done in recent years that gave them an understanding of the current situation. Others have not evaluated their data quality, whether it is because lacking proper tools, resources or importance. It was also pointed out that the problems with data quality are mostly solved reactively and it has not evolved

to a stage where it is monitored proactively. For solving data quality problems proactively, it was mentioned that it needs to be integrated to the IT systems, profiling tools that would measure and monitor data automatically and inform about inconsistencies. Some of the interviewees who brought out these automatic tools said that the public sector is just not there yet in its data governance evolution to use them, but it is only matter of time. One data governance representative also pointed out that data quality is a matter of requirements that are set when collecting data and if during requirements determination phase the future use cases of data are not foreseen and considered then in retrospect it is not fair to say that the data quality is bad. Rather it is a problem of lack of foresight and analyses. 2 different representatives also mentioned that data quality is this years' focus, meaning that most of their activities and resources are dedicated to this domain of data governance.

The representatives from the policy side also evaluated data quality differently. 3 out of 6 interviewees admitted that there are problems with data quality and brought different examples, e.g. the time between data collection and until their ready for use is too long, so the data does not represent the current situation anymore, or it is known that data is underreported, so it does not show the full picture of a situation, or the data is not uniformly collected and presented and therefore not comparable. The 3 other representatives mentioned that it seems to them that data quality is not the problem, but there may be exceptions when delving into details. 2 undersecretaries also pointed out that in their respective field, the problem is other institutions data quality and changing or influencing that is much more difficult, although necessary because it impacts their respective fields' work and output quality.

Data governance and policy representatives similarly pointed out that data quality issues arise with the use of data analytics tools. The analytical layer of data is good place to validate if the current data governance practices support the policy needs or not, from which it can be concluded that strong data analytics could contribute to better data management and governance in the organizations.

The importance of data governance

Questions 2 and 3 aimed at the data governance representatives focused on understanding how important data governance and its activities are considered in the governing area. 8

interviewees out of 9 mentioned that data governance is either currently mentioned in the organization's strategic document or in the organization's priorities, or a similar strategic document is currently in the making that sets data governance as a relevant topic. Another way to express the importance of an issue is if management understands and supports it, which was brought up in 4 out of 9 interviews. In some cases, a member of the management has knowingly taken the role of data governance sponsor, other cases the role has unconsciously formed for a member of the management.

In 3 interviews it was also brought out that data governance is a supporting activity, and it may not be popular or prominent, but it is necessary. It is necessary for the development and provision of high-quality services, for the preparation of statistical overviews, and for effective work processes and outputs. This kind change of mindset is difficult to implement in an organization because data governance does not result in cutting a ribbon somewhere, but rather the day-to-day work, activities and services of the organization function as planned. It was mentioned that the national direction and initiative in the field of data governance, from MKM and Statistics Estonia has helped with the knowledge reaching the institutions and driving the change.

The benefit and added value of data governance

Data governance representatives evaluated what are the benefits and additional value that data governance brings to a government sector organization. The three most named benefits were 1) growing effectiveness in the workplace, 2) ability to cross-use data, and 3) control and overview of the data, which makes easier to detect and solve problems. Additionally, data governance is directly linked to data quality which therefore reflects on the decisions that are made based on data. Data descriptions and dictionaries were also pointed out as necessary part of data governance, that helps to create a common data language and has seen to benefit communication with different partners, whether it is the policy makers or IT system developers. The added value of data governance activities pointed out, was the benefit of less duplicating work inside the organization as well as in general in the country, which is necessary to achieve the OOP. Cost saving was also mentioned in two cases, but with a remark that getting the data governance program up and running would at first need additional resources, people and funding, but in the long run the investment will start paying off. Increased transparency and better quality of IT systems and services were also mentioned as benefits that data governance increases.

From the representatives of the policy side was asked, what added value could data bring in their field. Their answers mostly focused on getting a better overview and control of their field, that could help measure the impact of their work and policies as well as support resource allocation process. The added value pointed out also included future prognosis models that in some cases have already proven its value and could help prepare better in the long term. In addition, the value of less duplicating work and data collection was pointed out by some of the undersecretaries from the point of view that this would leave more time for the citizens, companies as well as the public sector, to do more meaningful work and the avoidance of unnecessary costs.

Existing processes, procedures and activities related to data governance

Questions 8 and 9 aimed at the data governance representatives focused on describing the existing framework in place for data governance in the respective area of government. In 4 cases it was mentioned that there is an existing action plan, cooperation agreement or procedure that describes data governance processes, roles and/or activities. In 4 other cases similar framework is currently in the making and in 1 case no such framework has been implemented. It was pointed out that the starting point for creating this kind of document has been the Estonian Public Information Act and the guidelines from MKM and Statistics of Estonia. It was highlighted that the vision and general policy in the field of data governance should come from these institutions to ensure consistent development across the public sector.

In 4 interviews, participants also mentioned that they have created working groups that include data governance topic and coordination across the area of government, where representatives from the ministries and the authorities from the governing area are present. The working group ensures coordination and communication across the area of government and gives the opportunity to provide immediate feedback on existing processes and modify and improve them as needed. Working groups are there to increase accountability and work as self-monitoring systems.

As many interviewees mentioned that they are in the process of creating frameworks, guides and action plans, or have only recently gone through this process, they could elaborate on the steps they took to get started. A common feature that could be pointed out was that external support was used, either from MKM or from another partner whose

data governance competence was procured. The external partner was used to map and evaluate current data sets in the governing area and identify weaknesses and improvement points. In some cases, the external consultation partner also helped to understand and draft what should the future data governance system look like and what resources and activities are needed to make it happen. If an external expert was not used, a competent person was hired, who still did the mapping and evaluation of the databases and worked from there. Interviews with various departments of the organization were also mentioned to understand the challenges and problems that new and improved data governance processes could improve. In addition, 3 interviewees agreed that elementary description of data was one of the first steps that was taken to improve data governance. In some cases, basic monitoring systems were also established for IT applications to help identify and resolve the most pressing issues, including data quality issues.

One of the challenges brought out in 3 different interviews was creating a unified understanding when and how data governance should intervene in different organizational processes, especially in IT development. Since data governance is a horizontal topic that has many contact points across the organization, including service development, then it needs to be included in early stages. In 2 cases, some kind of guidelines for data governance inclusion in IT development were created, but more implementation and testing is needed to understand and prove its value. In one interview was pointed out that the EU structural funding for IT developments is clearly with the data management and governance dimension, bringing extra focus that these topics need to consider one another.

In 2 interviews it was also mentioned that the focus of data governance in their field is moving towards more machine-readable data that can be more efficiently processed, used and made available by machines. It was expressed in the interviews that the focus is on governing structured data, which is more manageable and needs less human handiwork to analyse and use it. E.g., in many cases the use of dashboards for visualizing data is very common and in use daily basis, which is made possible when using structured data.

In addition, training the policy side in basic data literacy, data analysis, data management and data governance was mentioned as one topic that has not received much attention so far, but will in the future. Since data governance is horizontal topic that impacts the entire organization and area of government then a few experts are not enough to make it

successful, but a wider understanding and involvement is needed. In some interviews were brought out that data related trainings are also in their agenda for creating better data governance program.

Problems and challenges related to data governance

All interviewees were asked to describe the problems and challenges they see in their respective field related to data governance. For the policy side representatives, the author gave a brief explanation of data governance and what it entails, so they could think more broadly about the topic when answering. The answers of the different target groups were somewhat consistent, resulting in top 5 problems and challenges that were repeated.

The most mentioned problem was lacking competent people in the field of data governance. The labour shortage is most evident during recruitment, when it is even difficult to find people who would apply for these positions. Some interviewees found that people with the necessary skills do not come from the education system but have acquired skills based on the need during their working career. Competition in terms of labour occurs with both the private and public sector, in the private sector it was mentioned that commercial banks have gathered very competent people in this field and the public sector salary is unfortunately not competitive. Resulting that the people who are hired often need a lot of training and time to grow into to the position, which would also explain why external competence has been procured in many cases to get started with data governance programs.

The next most mentioned problem, both by the representatives of data governance and policy, was the lack of knowledge and understanding of the topic in general. The people who are not directly related to data governance activities have a hard time understanding what data governance and management is, why is it necessary, what impact does it have, how it affects them and what is their role. On the one hand, it is a horizontal issue that concerns everyone in the organization, but at the same time, it is not widely recognized or understood what these contact points are. It was mentioned that very practical and real-life examples are needed to understand data governance, for example explaining how not doing data governance and management activities directly affects the services provided and their quality. Additionally, in some cases, the problem is that the policy side sees issues related to data and its management and governance as falling to either IT or some

other support unit, even if the data is essentially supported by the policy side, who knows what is needed from it or how to use it. So, the cooperation between policy side, IT and data governance, also by extension data analytics, is often not well established and needs better practices and processes to function.

The other 4 problems were mentioned less times in the interviews but still came up both from the data governance and policy side representatives. The challenges are:

1) collecting, managing, processing and storing data is expensive. Costs related to data governance are rising and there are situations where questions arise as to whether it is worth it.
2) Redundant and unnecessary datasets are collected and stored. Whether it means doing double work, even though the same data could be obtained from somewhere else or historically some data has been collected for which there is no longer a real need. Although the problem is recognized at some level, it is not easy to renounce this data.
3) Restraining or outdated laws and regulations that limit interoperability between different institutions and makes it easier to duplicate than cooperate. Whether it is the Public Information Act or Personal Data Protection Act.
4) The amount of handy work that is still necessary sets boundaries for a leap in development in data governance practices.

There were also many challenges brought out by representative of one target group and not mentioned by the other. The data governance representatives repeatedly mentioned obstacle was recognizing data governance and management as an important job and making it visible in the organization. Currently it feels like ungrateful background job, that is necessary but not valued. Limited resources were also mentioned, that it is difficult to create new positions who would deal with data governance and financial limitations to implement new monitoring systems. Another obstacle mentioned was the unclear roles of MKM and Statistics Estonia in coordinating national data management and governance. Their roles have changed over time and the expectation is to have a clear management model so that there would be no duplicative activities and time consumption for the public sector institutions. In addition, since implementing data governance practices and programs is quite new in the Estonian public sector, it was mentioned that sharing only theoretical knowledge does little good and there is a need for more practical

and constructive discussion end experience sharing, that not only focuses on the success stories but discusses the challenges and failures as well. Another challenge brought out was the uneven data governance and management capability in the public sector which has a wider influence, since the goal is to be interoperable and to share data in the public sector. Meaning that if one institution is not dealing with the topic like the others and it affects their data quality, accessibility, security and other factors, then all the other institutions and services are affected as well. Which leads to a question brought out in an interview – to what extent should the public sector institutions contribute to these topics? Should everybody have the same expectations or is it fine to nationally prioritise some datasets and their management and governance? Currently it seems the effort given depends on the public sector organizations capability and motivation.

The problems brought out only by the undersecretaries form the policy side were understandably focused on the usage of data. One of the main challenges mentioned (in 4 interviews out of 6) was that data is there but it is not usable in situations where it is necessary, whether it is for legal reasons or the fact that data is collected in so different ways that it is not comparable or relatable. Another problem mentioned as many times was the lack of skills and tools to interpret the data as needed. This relates to the fact that data itself has little value and needs to be enriched to transform into useful information. Undersecretaries brought out that for general trends and statistical overviews the current data is enough and useful, but problems arise when more detail data interpretation is needed. Also evaluating the impact of policies is a challenge, which includes creating metrics that would directly relate to the policies and using data that would support the evaluation. Examples of time criticality were mentioned, where data is needed faster than it possible to collect or transform to a needed form. In addition, in some cases the policy side or decision makers do not trust the data due to data quality problems.

Using data in the decision-making process

Question 7 aimed at the undersecretaries representing the policy side related to using data in the decision-making process. All the interviewees agreed that they are using data in their decision-making, but what varies is the extent of it. Data-based reports, statistics and usage of data dashboards is common, they show the general trends and are monitored quite often, but in terms of decision-making it is not always enough. The most critical point was having the data in the right context in the decision-making process, which is

not easy to achieve. Decisions are more specific than available data, so the data supports providing background context, but in some cases the specific data needed is not available. Some of the undersecretaries also pointed out that there is so much data that is important in their field, that they do not care for random facts, but prefer data with meaning and explanations, which is why data analytical work is valued from their point of view. 4 respondents said for them data-based decision-making is a normal practice, which has evolved over time. In some cases, it is very easily recognizable if the decisions are not evidence-based as it will show in the future outcome.

In 3 interviews was mentioned data quality as a barrier for using and trusting data in the decision-making process. There has been experiences where other public institutions data does not feel reliable or usable or the own data collected is too subjective or does not meet the time-critical requirement. In 2 interviews it was admitted that gut-based or experience-based decision-making is still part of their everyday work, but the barriers have been knowledge and work is ongoing for an evidence-based decision-making process. In addition, politicians were a barrier mentioned in 3 interviews. Political support is necessary for carrying out bigger changes, but in some cases data-based approach does not work with politics. As was mentioned in one interview: *'You have to continue to be very prepared, to argue with good data, and if you come across some political forces that are looking for this kind of topic or ideas or the state budget is better at the moment, then you might be successful with your argumentation.'*

One way to create trust in data and use it in the decision-making process is to use data in prognosis models. In some areas of government data-based prognosis is practiced and it gives the policy side not only an idea what is expected to come, but also certainty that data used in the prognosis is reliable if the expected prognosis comes true.

Roles and structural units involved

As mentioned before, data governance is a horizontal topic that involves a lot of people in an organization, which was also brought out in the interviews. To understand the nature and extent of the roles and structural units involved, question 7, directed to the data governance representatives, was aimed to explore the dynamic in the respective area of government. Firstly, to start with the people who gave the interview from the data governance support unit side, it was clear that there will not be a position called head of

data governance found in every institution. There were only 2 institutions where this specific role was present. Other interviewee roles were data management advisor (3), chief data officer (1), data management expert (2), head of data and analytics department (1) and undersecretary of a support unit in the ministries (2). Considering that their job titles were very different, all of them could discuss data governance in their respective area of government, although with a different level of detail and experience, which was acknowledged limitation. It also varied where the structural unit with the interviewees position and other positions related to data governance were located. Except for data management experts for business and consumer environment at MKM, all the representatives were located in the supporting units in the organization, most commonly units dealing with strategy, innovation and/or development. From structural point of view, the data governance positions were close to data analytics and IT development activities.

Positions from the policy side mentioned that have a tight collaboration with data governance supporting unit, are product owners, service managers, IT system owners, main users and database owners. Again, the government institutions interviewed have different logic on how their workflows and positions related to IT systems have formed, but the similarity is that many of these positions are in policy units. These positions and their inclusion in data governance activities are important since they know the policy side needs, requirements and logic which is crucial not only for IT development, but also for effective data governance processes and activities.

In addition, data analysts or data analytics department was brought out as an important partner, since often the problems with data governance and management become evident in the analysis phase as well as the legal department also plays a role, since they know best the laws that affect data governance. Another important partner is the head of information security, which is a mandatory role for all Estonian state institutions. This position is responsible for the information, data and technological tools needed for processing, which is also an important domain of data governance.

Additional positions mentioned in the interviews were ICT coordinator, open data coordinator, data engineer and document manager, with whom there are points of contact within data governance activities. One of the most important roles mentioned is the sponsor of data governance, which is someone from the top management, who drives the need for data-driven organization including data governance.

Competences and people

Question 9 for the policy side representatives and 8 for the data governance representatives related to competences that are necessary for an effective data governance program. The most repeating answer from the data governance representatives was the lack of awareness and basic knowledge from the policy side on data governance. Since their contribution is necessary at some level then the very basic knowledge of the topic is expected. This relates to mentioned challenge that in some institutions the database or IT system owners' roles and practice has not been fully implemented, which does not support the mentality that policy or business side are the owners of their fields' data. In addition, the need for business analytics competences were mentioned, which could help to include data governance and IT development processes from the beginning of creating a new service. In 3 interviews were mentioned that the competences have significantly grown in the last year, both from the data governance supporting unit and the policy side. Examples were brought with data description developments and implementation with data quality tools and increasing the general understanding of the topic. Digital and data literacy skills were brought out in 2 interviews as a competence to improve and the general mindset that quality data and effective data governance is everybody's responsibility. In one interview was also mentioned that the lack of knowledge of what skills are necessary to have in an organization and how much effort and attention should this topic receive, as barriers to overcome.

For the policy side representatives, the opinions were divided, in 3 cases the undersecretaries found that there are enough people and competences for effective data governance in their experience and in 3 other cases it was brought out what competences are lacking. Mostly were mentioned data quality, accessibility, collecting data in machine-readable formats and data analytics that could be improved if there would be more people working on these topics. It was also pointed out that understanding and creating these new positions and roles for data governance is an evolutionary process for institutions and it needs to evolve step-by-step. 3 undersecretaries also pointed out that finding these people and right competences on the labour market is difficult, which aligns with answers from the data governance representatives.

Keeping up to date with data governance

The representatives from the data governance supporting unit described how do they keep themselves informed of new practices, policies and regulations in the field of data governance. By far the most common answer was that MKM and Statistics Estonia have organized different formats for communications that mostly cover the need for keeping up to date on the topic. There are subject-based mailing lists that people are subscribed to, where important information is passed on and they are also used for feedback for new action plans and guidelines. Also, national data governance steering group is organized where the most important topics are passed on to representatives from different public sector institutions. More detailed topic-based working groups have been gathered for discussions and development of a specific domain of data governance, e.g., data quality. The work that MKM and Statistics Estonia is doing received mostly positive feedback and is appreciated among the government institutions. However, if the public sector institution has not created the positions responsible for data governance or are overwhelmed with exciting work, then there are not enough resources to take part in these different discussions and activities. The theoretical material that has been created in the last years, guidelines and frameworks, was mentioned as too much to go through for some institutions. In 3 interviews was also pointed out that more practical discussions are expected that would give the opportunity to share more real-life examples and experience.

Some ministries who have many institutions in their governing area are also organizing steering groups inside their area of government, to exchange to most important guidelines from the national level and for coordinating data governance in the respective field on more practical level. Trainings were also mentioned as a possibility for keeping up with the newest tools and trends. In some areas of government guidelines and requirements affecting the work of data governance also come from the EU, that must be followed and considered.

Perspective of data governance

The purpose of question 12, addressed to the representatives of data governance supporting unit, was to understand what perspective the interviewees see in data governance in a couple of years. The most mentioned trend (in 8 out of 9 interviews) was AI and machine learning solutions. In terms of data governance, it was mentioned that AI

could help with administrative burden and monotone activities like data descriptions or proactively identifying data quality problems. It is seen that data is produced and collected exponentially, but the manual labour does not grow in the same way, resulting in the need for automated tools. Automation by AI or machine learning technologies could also help with tracking dataflows, from data origin, transformation to outputs, which in general is called data linage. In some public sector institutions basic machine learning solutions are already in place for the policy side, but it was brought out that the expectations are higher. Systems could also better support the decision-making process and national developments like *Bürokratt* could be implemented on a wider scale, including inside the organizations to find information and answers more easily. The current practice seems to be piloting the AI and machine learning solutions, but there is not enough practice on how to sustain these solutions. Even though the interviewees brought out what AI and machine learning could accomplish, then in reality people were quite sceptical, because it is seen that the public sector does not have enough resources to start implementing and successfully sustain these solutions and is not willing to take risks that much.

In a few years perspective, it was also seen that the cooperation between state authorities could get better, since the institutions are collectively butting more focus on data governance and management, which includes adding positions, naming clear roles and responsibilities. In 2 interviews was mentioned that data description tool RIHAKE is planned to take in use, while in some cases the focus is on creating glossaries and data dictionaries. Trend is to move towards creating more real-time data, machine-readable data and open data, which comes with its peculiar quality and security questions. In addition, anonymisation and pseudonymization of personal data was mentioned as growing trends which would helpfully allow more usage of this data for different purposes. For experimenting with data, in one interview was mentioned the expectation for data playground or sandbox, where different parties could handle, test new theories and experiment more with data.

In general, most of the representatives of data governance saw that the future would bring more transparency and trust in data, new technologies and data governance practices. The workflows will get more effective, since the focus is on creating maximal value with the data we already have.

Questions 10 and 11, directed to the undersecretaries representing the policy side, gave them the opportunity to think about what could be done more with existing data and what added value could be created. Similarly, to data governance representatives, AI and machine learning solutions were mentioned in all the interviews from the perspective that it could support the decision-making process and search for patterns, which could result in finding things that the policy side has not even considered. In addition, the expectation for different prognosis models was mentioned. In some areas, the models have been adopted and proven, so there is hope for their wider adoption. Better analytical tools, wider selection and use of data dashboards and moving towards creating real-time data was also mentioned as ways of creating more value out of existing data. It was pointed out that in some cases the policy side is not even aware of the possibilities that different use of data could provide, which is why there is no big urge for new solutions. There is also an understanding that all new technological solutions need large investments that are difficult to get in the public sector.

4.2 Discussion

The aim of the thesis was to identify and describe the value of data governance in the context of the Estonian government sector and how to successfully implement its frameworks, processes and activities in the organization. The author proposed two research questions each with one sub-question which now, after the interviews are conducted, can be answered.

The first RQ was ‘What is the value of data governance in the Estonian government sector?’ with a SQ ‘What motivates Estonia’s government institutions to implement data governance practises?’ Although these questions were not asked directly from the interviewees, then the answers to these questions emerged in the light of different questions. An overview of the key findings is gathered in Table 1.

The increase in efficiency is one of the most important benefits that data governance creates for an organization. This relates to easier detection of data-driven problems and solving them, as well as cross-use of data for avoiding duplicating work for the public sector and for the citizens and companies. Data governance activities take control and authority over data to ensure it its suitability and reliability for use, which reflects on decision that are based on the data. Since data is directly likened with organizational work

processes like services offered in the public sector, then the value of data governance is also seen in improved quality and capacity of service provision. In addition, data descriptions, metadata management and glossaries enable easier collaboration with different partners due to understandable and agreed upon common language. Overall, data governance framework implementation helps to increase trust and transparency in data managed by the public sector.

Table 1. An overview of the value and motives of implementing data governance.

The described value of data governance	**Motivation for implementing data governance**
Growing effectiveness in the workplace	Effective work processes and outputs
Easier to detect data-driven problems and solve them	Better overview and control of the data
Less duplicating work because data is cross-used	More time for meaningful work and less administrative burden
Decisions based on data are reliable	Desire to implement automated solutions, e.g., prognosis models, dashboards, AI support for DDDM
Improved quality and capacity of the services offered	Ability to better measure development and impact of the field
Easier to collaborate through a common data language	Cost efficient in the long run
Increased trust and transparency in data	Compliance with laws and regulations

The government sector motivation to start implementing data governance practices can be different for every organization, depends on what is needed and valued from their perspective. It can just be the minimum by just complying with existing laws and regulations or the desire to change work processes more efficient by automated tools, like mechanical quality controls, prognosis models or implementing AI solutions for supporting service provision or DDDM. The motivation can also come from wanting to lighten up the administrative burden or getting a better overview and control over the exciting data. Implementing data governance practices could also help to measure the development of a field more accurately or evaluate the impact of a new policy or measure. Additionally, there is evidence that data governance could be a cost saving for an

organization, even if the Estonian government institutions have not assessed it from financial perspective yet.

The second RQ proposed focused on the implementation process of data governance, 'What are the strategies for implementing data governance practices in a government organization?' with a SQ 'What kind of organizational changes are needed to improve the state of data governance?' It is mentioned in the literature and aligns with the interview answers, that data governance is an ongoing series of activities and processes in the organization that does not have an end date like a project has. If it is decided to start implementing it more systematically and comprehensively, there are some strategic approaches that could bring more success. The following list, with 11 strategic activities, describes how to implement data governance in the organization and what changes are needed for make.

1. **Finding the motivation for implementing data governance**. Without an aim for changing organizational workflows, practices and mindset it is very likely that data governance implementation will not succeed. As mentioned in literature and confirmed by interviewees, top managements' understanding and support is necessary for finding and phrasing the right motivation and vision for data governance program.
2. **Identifying and creating roles and responsibilities that deal with data governance**. To get started with the implementation process, data governance program needs a person who is ready to delve into the topic and lead the change. Depending on aim and scope of the program, then there may be need engage more people over time, but at beginning at least one who is invested is topic is necessary and has the support of the top management.
3. **Familiarizing oneself with existing laws, regulations and guidelines.** Data governance in Estonia is regulated by laws like Public Information Act and Personal Data Protection Act and from the EU side regulation The European Data Governance Act (coming in force in September 2023). In addition, MKM with Statistics of Estonia have developed different guidelines which help to explain and support implementing data governance different practices. These materials could be helpful before creating the framework for a specific organization or area of government.
4. **Mapping the existing data sets.** As mentioned in one interview '*you govern the data you have*', which is why it is necessary to understand what kind of data sets the

organization or area of government is responsible for and for which processes the data is used. To have a clear overview of the data that needs to be governed.

5. **Identifying areas of development that data governance could support.** For getting a clear overview of what needs to change, it is important to understand what currently creates problems and where the bottlenecks are located. Based on the examples given in the interviews, external competence has also been procured, who can help with this process, as well with ideas on how to solve identified problems.
6. **Creating a working or steering group for people involved across the area of government.** Especially when the data governance program transcends the boundaries of a single organization, then coordination across the area of government is necessary for data governance. The purpose of these working groups can be developing new practices, communicating national new guidelines, sharing feedback on existing practices and for creating accountability.
7. **Creating a data governance framework and principles to follow.** The framework sets in place the scope of the data governance program and the policies, processes and principles that support achieving the set goal for data governance in the organization or area of government.
8. **Creating an action plan with monitoring tools.** Framework alone does little good, unless it is followed up by an action plan that involves ways how to monitor the impact of data governance program. The action plan should have a specific timeframe that sets the focus on concrete domains of data governance.
9. **Communication and engagement with different stakeholders.** Since data governance implementation can bring changes on a wider scale in the organization, then the communication is the key, not only to keep people informed, but also to get feedback on the processes and output that it effects.
10. **Identifying the contact points with different processes for collaboration.** As highlighted in many interviews that data governance is a horizontal process, so finding the collaboration points with other processes is important to ensure that the perspective of data governance will intervene in the right time right place, e.g., when developing new services and IT systems.
11. **Monitoring and improving the implementation of the framework, principles and action plan.** A crucial part of data governance activities is monitoring of the progress and adjusting the plans accordingly. This creates a continuous process for data governance, which adapts with changing requirements and circumstances.

The 11 strategic activities brought out could ensure successful implementation of a data governance program in a government sector organization or area of government, according to literature and the interview answers. The order and undivided tracking of the activities listed is recommended but can change depending on organizations' development stage in data governance.

4.3 Recommendations for the policy makers

All the representatives of the data governance support unit mentioned MKM and Statistics Estonia as national policy makers of data governance, whose work has a broad impact on the implementation of practices in this field. Due to this reason, the author lists six recommendations for the policy makers brought out in interviews, that could help make data governance even more successfully acknowledged and implemented than it is today.

1. Create opportunities for a more practitioner-level exchange of data governance and management practices. The public sector institutions have gained experience, lessons and successful practices that would be good to share so that others can benefit from it.
2. For policy makers to have a clear management model so that there would be no duplicative activities and time consumption for the public sector institutions.
3. Creating a unified understanding how data governance should intervene in IT development processes, to tackle the issue that problems with data governance and management appear when IT is already built.
4. In the case of new IT developments in the public sector, the needs of national statistics could be considered before starting the development process. This would result in less duplicative work and save resources.
5. Create interest in data governance and management by talking about the opportunities and practical examples at top management level in the public sector institutions.
6. Distinguish between data governance and data management in Estonian terminology and communicate the difference to Estonian public sector institutions. This would create a better understanding of the substantive difference between these two concepts and would help institutions to better differentiate their activities and goals.

The overall satisfaction with national policy makers, MKM and Statistics Estonia, was very good and their efforts and commitment in the field of data governance is recognized.

Under their coordination, data management and governance could make a big leap in development in the Estonian public sector in the coming years.

4.4 Limitations

The conducted research has its limitations to be considered. Firstly, as mentioned before, the interviewed representatives of the data governance supporting unit had different backgrounds and positions. Even though they all could contribute to this study then their expertise, focus area and knowledge on data governance differed, resulting in responses with varying degrees of detail, which also has an impact on the findings.

The second limitation is the small sample size. Altogether 11 data governance representatives and 6 policy undersecretaries were interviewed, and broad generalizations are limited based on this. In addition, being the sole author of the thesis, there is a potential for research bias in the process of data collection or analysis, which may affect the research findings.

Lastly, in Estonian language, there is no distinction made between data governance and data management, they are both addressed as *andmehaldus*. This also caused some confusion among the interviewees, since in English the concepts and the activities under them are kept separate. To mitigate this, in the beginning of every interview the author mentioned the exact focus of this research.

5 Summary

The aim of the thesis was to identify and describe the value that data governance brings to the Estonian government sector and map strategies how to implement data governance frameworks, processes and activities in the organization. Based on the purpose, the author proposed two research questions each with one sub-question, the answers to which were determined during the research.

In the first section, the thesis provided a review of the literature to give a comprehensive overview of data governance and put it in the context of Estonia and the EU. The principles, framework, activities and main benefits of data governance were introduced, followed up by description of possible implementation methods.

Based on the aim of the thesis and the content of the field, it was decided to apply the qualitative research method. The author conducted 15 interviews with participants from two different target groups: representatives of a ministry's policy field and representatives of the data governance coordinating support unit. In-depth thematic data analysis was used to understand and describe the interviewees' background, experience and understanding of data governance and to highlight 12 emerged key themes that support the fulfilment of the thesis objective.

The first RQ ‘What is the value of data governance in the Estonian government sector?’ with a SQ ‘What motivates Estonian government institutions to implement data governance practises?’ was answered by identifying the already existing benefits and value that data governance has created for the Estonian government sector and by describing the potential benefits that are underway to achieve. Altogether, seven topics emerged for describing the value of data governance: growing effectiveness in the workplace, easier detection and solving of problems, less duplicating work, more reliable decisions, improved quality and capacity of the services offered, easier collaboration through a common data language, and increased trust and transparency in data.

Government institutions' motivations for implementing data governance practices may vary and overlap with the perceived value of data governance. The seven findings that

emerged from the research are: effective work processes and outputs, better overview and control of the data, more time for meaningful work and less administrative burden, desire to implement automated solutions, ability to better measure development and impact of the field, being cost efficient in the long run, and complying with laws and regulations.

Analysing the existing frameworks, processes, resources and considering the obstacles to the successful implementation of data governance in the Estonian government sector, the author was able to answer the second RQ 'What are the strategies for implementing data governance practices in a government organization?' with a SQ 'What kind of organizational changes are needed to improve the state of data governance?' As a result of the research, 11 strategic activities were revealed that could ensure successful implementation of a data governance program in a government organization:

1. Finding the motivation for implementing data governance.
2. Identifying and creating roles and responsibilities that deal with data governance.
3. Familiarizing oneself with existing laws, regulations and guidelines.
4. Mapping the existing data sets.
5. Identifying areas of development that data governance could support.
6. Creating a working or steering group for people involved across the area of government.
7. Creating a data governance framework and principles to follow.
8. Creating an action plan with monitoring tools.
9. Communication and engagement with different stakeholders.
10. Identifying the contact points with different processes for collaboration.
11. Monitoring and improving the implementation of the framework, principles and action plan.

The implementation of data governance practices is a continuous process that develops over time and relies on the outlined strategic activities for systematic and comprehensive approach. Since data governance in Estonia is strongly influenced by the national policy creators, MKM and Statistics Estonia, then the author listed six recommendations for them to consider, when steering the direction of data governance in Estonia:

1. Create opportunities for a more practitioner-level exchange of data governance and management practices.

2. For policy makers to have a clear management model so that there would be no duplicative activities and time consumption for the public sector institutions.
3. Creating a unified understanding how data governance should intervene in IT development processes, to tackle the issue that problems with data governance and management appear when IT is already built.
4. In the case of new IT developments in the public sector, the needs of national statistics could be considered before starting the development process.
5. Create interest in data governance and management by talking about the opportunities and practical examples at top management level in the public sector institutions.
6. Distinguish between data governance and data management in Estonian terminology and communicate the difference to Estonian public sector institutions.

The recommendations that originate from practitioners, would help create a better common understanding of the need for data governance and would support its wider implementation in the Estonian public sector.

In conclusion, the implementation of data governance in Estonian government sector is at an early stage, which creates several challenges for practitioners. However, the author believes that a wider explanation of the value of data governance will help and give momentum to its smoother and broader implementation, that will not only benefit the government institutions in the performance of their duties, but also reflects the quality of services provided to citizens and businesses.

5.1 Future research

Implementation of data governance in the Estonian public sector has potential for future research. In the course of the study, the intervention of data governance in various processes of the organization, such as the development of IT systems and services, emerged. It could be explored what are the potential points of contact for cooperation in the development process, which in the long run could reduce the problems of data governance and management.

In addition, the expansion of research into public sector data governance practices in other EU countries would provide better clarity on the opportunities and obstacles for the development of interoperability and cross-border data exchange.

References

[1] T. C. Redman and A. B. Godfrey, *Data Quality for the Information Age*, 1st ed. USA: Artech House, Inc., 1997.

[2] Dama International, *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Denville, NJ, USA: Technics Publications, LLC, 2017.

[3] Ministry of Economic Affairs and Communications, "Avaandmete mõjuhinnang avalikus sektoris 2022," *Estonian open data portal*, 2023. https://avaandmed.eesti.ee/instructions/avaandmete-mojuhinnang-avalikus-sektoris-2022 (accessed Apr. 25, 2023).

[4] L. Sebastian-Coleman, *Measuring Data Quality for Ongoing Improvement : A Data Quality Assessment Framework*. San Francisco, UNITED STATES: Elsevier Science & Technology, 2013. [Online]. Available: http://ebookcentral.proquest.com/lib/tuee/detail.action?docID=1106491

[5] W. H. H. Inmon and D. Linstedt, *Data Architecture: a Primer for the Data Scientist : Big Data, Data Warehouse and Data Vault*. San Francisco, UNITED STATES: Elsevier Science & Technology, 2014. [Online]. Available: http://ebookcentral.proquest.com/lib/tuee/detail.action?docID=1875436

[6] G. Thomas, "The DGI Data Governance Framework." Accessed: Mar. 06, 2023. [Online]. Available: https://datagovernance.com/the-dgi-data-governance-framework/

[7] J. Ladley, "Data governance : How to design, deploy and sustain an effective data governance program.," *Elsevier Science & Technology*, 2012, Accessed: Feb. 02, 2023. [Online]. Available: https://ebookcentral.proquest.com/lib/nlibee-ebooks/detail.action?docID=1032951

[8] P. Brous, M. Janssen, and R. Vilminko-Heikkinen, "Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles," in *Electronic Government*, Springer International Publishing, 2016, pp. 115–125. doi: https://doi.org/10.1007/978-3-319-44421-5_9.

[9] R. Abraham, J. Schneider, and J. vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *Int J Inf Manage*, vol. 49, pp. 424–438, 2019, doi: https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

[10] P. Tallon, R. Ramirez, and J. Short, "The Information Artifact in IT Governance: Toward a Theory of Information Governance," *Journal of Management Information Systems*, vol. 30, no. 3, pp. 141–177, 2013, [Online]. Available: http://www.jstor.org/stable/43590145

[11] H. Borgman, H. Heier, B. Bahli, and T. Boekamp, "Dotting the I and Crossing (out) the T in IT Governance: New Challenges for Information Governance," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 2016, pp. 4901–4909. doi: 10.1109/HICSS.2016.608.

[12] A. Weller, "Data governance: Supporting datacentric risk management.," *Journal of Securities Operations & Custody*, vol. 1, no. 3, p. 250, 2008.

[13] V. Khatri and C. Brown, "Designing data governance," *Commun. ACM*, vol. 53, pp. 148–152, Jan. 2010, doi: 10.1145/1629175.1629210.

[14] Nascio, “Data Governance - Managing Information As An Enterprise Asset: Part I - An Introduction,” 2008. [Online]. Available: www.nascio.org.

[15] L. Zhang, K. Chen, and J. Zhao, “Evidence-Based Decision-Making for a Public Health Emergency in China: Easier Said Than Done,” *Am Rev Public Adm*, vol. 50, no. 6–7, pp. 720–724, 2020, doi: 10.1177/0275074020942410.

[16] N. Cartwright and J. Hardie, *Evidence-Based Policy: A Practical Guide to Doing It Better*. 2012. doi: 10.1093/acprof:osobl/9780199841608.001.0001.

[17] J. Parkhurst, *The Politics of Evidence (Open Access): From evidence-based policy to the good governance of evidence*. London: London: Routledge, 2017. doi: 10.4324/9781315675008.

[18] F. Provost and T. Fawcett, “Data Science and Its Relationship to Big Data and Data-Driven Decision Making,” *Big Data*, vol. 1, Mar. 2013, doi: 10.1089/big.2013.1508.

[19] J. Dingelstad, R. T. Borst, A. Meijer, H. R. M. Strategisch, U. U. L. E. G. R. U. S. G. P. Matters, and P. management en publieke innovaties, “Hybrid Data Competencies for Municipal Civil Servants: An Empirical Analysis of the Required Competencies for Data-Driven Decision-Making,” *Public Pers Manage*, vol. 51, no. 4, pp. 458–490, 2022, doi: 10.1177/00910260221111744.

[20] I. Alhassan, D. Sammon, and M. Daly, “Data governance activities: a comparison between scientific and practice-oriented literature,” *Journal of Enterprise Information Management*, vol. 31, no. 2, pp. 300–316, 2018, doi: https://doi.org/10.1108/JEIM-01-2017-0007.

[21] K. Weber, “A Model for Data Governance - Organising Accountabilities for Data Quality Management,” in *http://www.alexandria.unisg.ch/Publikationen/67284*, Jan. 2007.

[22] D. Plotkin, *Data Stewardship*. Morgan Kaufmann, 2014. Accessed: Jan. 17, 2023. [Online]. Available: https://learning.oreilly.com/library/view/data-stewardship/9780124103894/?ar=

[23] J. Korhonen, I. Melleri, K. Hiekkanen, and M. Helenius, “Designing Data Governance: An Organizational Perspective,” *The GSTF Journal of Computing*, vol. 2, Jan. 2013, doi: 10.5176/2251-3043_2.4.203.

[24] U. Gupta and S. Cannon, *A Practitioner’s Guide to Data Governance : A Case-Based Approach*. Bingley, UNITED KINGDOM: Emerald Publishing Limited, 2020. [Online]. Available: http://ebookcentral.proquest.com/lib/tuee/detail.action?docID=6238638

[25] A. Cave, “Exploring Strategies for Implementing Data Governance Practices,” Walden University, 2017. [Online]. Available: https://scholarworks.waldenu.edu/dissertations

[26] J. Holmes, “Privacy, Security, and Patient Engagement: The Changing Health Data Governance Landscape,” *EGEMS (Wash DC)*, vol. 4, p. 1261, Mar. 2016, doi: 10.13063/2327-9214.1261.

[27] M. A. Carpenter, M. A. Geletkanycz, and Wm. G. Sanders, “Upper Echelons Research Revisited: Antecedents, Elements, and Consequences of Top Management Team Composition,” *J Manage*, vol. 30, no. 6, pp. 749–778, Dec. 2004, doi: 10.1016/j.jm.2004.06.001.

[28] B. Wooldridge, T. Schmid, and S. W. Floyd, “The Middle Management Perspective on Strategy Process: Contributions, Synthesis, and Future Research,” *J Manage*, vol. 34, no. 6, pp. 1190–1221, Dec. 2008, doi: 10.1177/0149206308324326.

[29] J. Page, “How to Launch a Data Governance Initiative.,” *Business Intelligence Journal*, vol. 16, no. 2, pp. 17–25, Jun. 2011, [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=61965734&site=ehost-live&scope=site

[30] M. Haenlein and A. Kaplan, “A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence,” *Calif Manage Rev*, vol. 61, no. 4, pp. 5–14, Jul. 2019, doi: 10.1177/0008125619864925.

[31] The European Commission’s High-Level Expert Group on Artificial Intelligence, “A definition of AI: Main capabilities and scientific disciplines,” Brussels, Dec. 2018. Accessed: May 04, 2023. [Online]. Available: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

[32] European Commission and Directorate-General for Justice and Consumers, *The EU data protection reform and big data*. Publications Office, 2018. doi: doi/10.2838/190200.

[33] S. Giest and R. Ng, “Big Data Applications in Governance and Policy,” *Politics and Governance*, vol. 6, no. 4, pp. 1–4, Nov. 2018, doi: 10.17645/pag.v6i4.1810.

[34] R. Mahanti, *Data Quality : Dimensions, Measurement, Strategy, Management, and Governance*. Milwaukee, WI, UNITED STATES: ASQ Quality Press, 2019. Accessed: Mar. 22, 2023. [Online]. Available: http://ebookcentral.proquest.com/lib/tuee/detail.action?docID=6262212

[35] A. Berson and L. Dubov, “MDM Services for Entity and Relationships Resolution and Hierarchy Management,” in *Master Data Management And Data Governance*, 2nd ed.McGraw Hill, 2010. Accessed: Apr. 14, 2023. [Online]. Available: https://learning.oreilly.com/library/view/master-data-management/9780071744584/

[36] J. L. Solano, A. Martin, S. de Souza, and T. Linnet, “Governing data and artificial intelligence for all: Models for sustainable and just data governance,” European Parliament, 2022. Accessed: Apr. 25, 2023. [Online]. Available: https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2022)729533

[37] European Commission, “European Commission.” https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/institutions-and-bodies-profiles/european-commission_en (accessed Jan. 26, 2023).

[38] European Commission, “A European Strategy for data.” https://digital-strategy.ec.europa.eu/en/policies/strategy-data (accessed Jan. 26, 2023).

[39] European Commission, “European Data Governance Act.” https://digital-strategy.ec.europa.eu/en/policies/data-governance-act (accessed Jan. 26, 2023).

[40] European Parliament and Council of the European Union, “Regulation (EU) 2022/868 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act),” *Official Journal of the European Union*, Accessed: Jan. 26, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R0868

[41] European Commission, “The Once-Only Principle Project (TOOP).” https://ec.europa.eu/isa2/isa2conf18/once-only-principle-project-toop_en/ (accessed Apr. 15, 2023).

[42] R. Krimmer, “Once-Only,” *TOOP*. https://toop.eu/once-only (accessed Apr. 15, 2023).

[43] Statistics Estonia, “Data governance.” https://www.stat.ee/en/statistics-estonia/data-governance (accessed Jan. 27, 2023).

[44] “Andmehalduse tegevuskava 2021-2022.” https://www.kratid.ee/andmehalduse-tegevuskava (accessed Jan. 27, 2023).

[45] “Andmehalduse juhised.” https://www.kratid.ee/andmehalduse-juhised (accessed Jan. 27, 2023).

[46] Information System Authority, “Administration system for the state information system RIHA and RIHAKE,” 2022. https://www.ria.ee/en/state-information-system/data-based-

governance-and-reuse-data/administration-system-riha-and-rihake (accessed Mar. 04, 2023).

[47] W. A. Edmonds and T. D. Kennedy, *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods*, Second. SAGE Publications, Inc, 2017. doi: https://dx.doi.org/10.4135/9781071802779.

[48] R. Kumar, *Research Methodology: A Step-by-Step Guide for Beginners*, 4th ed. SAGE Publications Ltd, 2024.

[49] N. King and C. Horrocks, *Interviews in Qualitative Research*. SAGE Publications Inc., 2010.

[50] A. Olev and T. Alumäe, "Estonian Speech Recognition and Transcription Editing Service," *Baltic J. Modern Computing*, vol. 10, no. 3, pp. 409–421, 2022, Accessed: Apr. 08, 2023. [Online]. Available: http://bark.phon.ioc.ee/webtrans/

[51] A. Galletta and W. E. Cross, *Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication*. NYU Press, 2013.

Appendices

5.2 Appendix 1 – Interview guide

Interview guide consists of an interview plan, list of interviewees, interview protocol and questions, both in English and Estonian.

1. Interview plan

The goal of the interviews is to gather an extensive overview and understanding of the current data governance situation in the Estonian government sector, which would support the goal of the thesis to identify and describe the value that data governance practices bring to the Estonian government institutions and map strategies to implement data governance frameworks, practices and activities in the organization.

The interviews are conducted with two different target group representatives, an undersecretary of a policy field and a representative of data governance coordinating support unit, to get different viewpoints to importance of data and its governing. The interview participant selection was done with careful consideration, taking into account the specificities of the institutions and their data governance coordination.

Table 2. A list of interviewees, their positions and institutions.

	Public administration institution	**Representative of the policy field** *Name and title*	**A representative of the data governance coordinating support unit** *Name and title*
1	Ministry of Culture	-	Karl Viilmann, Chief Data Officer
2	Ministry of Education and Research	Liina Põld, Undersecretary for General Education and Youth Policy	Katre Seema, Head of Data Governance
3	Ministry of Social Affairs	Ulla Saar, Undersecretary for Labor Affairs	Sirli Niibo-Tamm, Data Management Advisor
			Aivi Saar, former Data Management Advisor and current Data Quality Advisor at The Health and Welfare Information Systems Centre
4	Ministry of the Environment	Kaupo Heinma, Undersecretary for Environmental Management and Foreign Relations	Margit Martinson, Undersecretary for Support Services and Land Policy
5	Ministry of the Interior	Viola Murd, Undersecretary for Rescue and Crisis Management	Liisi Lillipuu, Data Management Advisor
6	Ministry of Economic Affairs and Communications	-	Adele Johanson, Data Management Expert for Business and Consumer Environment
			Hendrik Valgemäe, Data Management Expert for Business and Consumer Environment
			Reimo Tarkiainen, Head of Data and Analytics Department at Estonian Transport Administration
7	Ministry of Finance	Evelyn Liivamägi, Deputy Secretary-General for Financial and Tax Policy	Alvar Pihlapuu, Head of Data Governance at Estonian Tax and Customs Board
8	Ministry of Foreign Affairs	Mariin Ratnik, Undersecretary for Economic and Development Affairs	Olavi Seisonen, Undersecretary for Administrative Affairs

2. Interview protocol

Interview protocol translated to English, which describes the content for both target groups: representative of a policy field and representative of data governance coordinating support unit.

Thank you for agreeing and taking the time to meet with me. My name is Liisbeth Laasik, and I am conducting research for my master's thesis, the aim of which is to create an understanding of what value the Estonian government sector sees in data governance and to map strategies for how to start implementing data governance in the administrative field. The focus of my work is on data governance, not so much on data management.

The interview lasts about one hour, and if it suits you, I will record our conversation so that later in the analysis phase, your thoughts and comments will be remembered correctly.

[Started the recording]

I have prepared 12 / 11 questions and some follow-up questions, divided into three categories:

- *the importance of data and its governing / goals and focus of the field, most important data;*
- *current state of data governance and established processes / data use cases, data-driven decision-making, data quality and reliability;*
- *data governance perspective / data related problems, improvement areas and perspective.*

When answering the questions, please reflect on your own experiences and thoughts as [the position they're in] and from the point of view of the respective administrative area.

Does this interview plan suit you?

[The interview takes place]

Do you have any questions or thoughts about the topic that you want to share?

Thank you for your contribution! I will send you my completed master's thesis in May. Is it okay if I mention and thank you by name in my master's thesis?

Interview protocol in Estonian

Tänan, et olite nõus ja leidsite aja minuga kohtumiseks. Minu nimi on Liisbeth Laasik ja viin enda magistritöö jaoks läbi uurimust, mille eesmärgiks on tekitada arusaam, millist väärtust Eesti valitsussektor näeb andmehalduses ning kaardistada strateegiad, kuidas alustada andmehalduse juurutamist haldusalas. Minu töö fookus on andmehalduse valdkonna juhtimisel ning vähem selle tehnilisel rakendamisel.

Intervjuu kestab umbes üks tund ning kui teile sobib, siis salvestan meie vestluse, et hiljem analüüsi faasis teie mõtted ja kommentaarid korrektselt meeles oleksid.

[Käivitan salvestuse]

Olen ette valmistanud 12 / 11 küsimust ja mõned jätkuküsimused, mis jagunevad kolme kategooriasse:

- *andmete ja nende halduse olulisus / valdkonna eesmärgid ja fookus, olulisemad andmed;*
- *hetkeolukord ja väljakujunenud protsessid / andmete kasutusjuhud, andmepõhine otsustamine, andmete kvaliteet ja usaldusväärsus;*
- *andmehalduse perspektiiv / andmetega seotud probleemid, arengukohad ja perspektiiv.*

Palun, et küsimustele vastates reflekteerite enda kogemusi ja mõtteid … ja selle haldusala vaatenurgast.

Kas selline intervjuu plaan sobib?

[Intervjuu leiab asset]

Kas Teil on endal küsimusi või mõtteid teema kohta, mida jagada soovite?

Aitäh panustastamst! Saadan Teile maikuus enda valminud magistritöö. Kas sobib kui Teid nimeliselt magistitöös mainin ja tänan?

3. Interview questions

Interview questions in English.

Questions for the representative of the policy field:

For introduction, please describe in your own words what your position entails and what your main responsibilities involve.

Goals and focus of the field, most important data

1. What are the main goals and focus of your field at the moment?
2. What information do you need to fulfil your goals, what is the most important data in the field?

Data use cases, data-driven management, data quality and reliability

3. What are the most important use cases for data in your field?
4. How do you measure the development and success of the field?
5. How important is quality and reliable data to your industry?
6. How do you assess the situation in terms of data quality? Is this a concern in your field?

Data related problems, improvement areas and perspective

7. Do you feel that the available data contributes to decisions made in the field?
8. What are the main problems in governing existing data?
9. Do you think that there are enough people and competences to manage and effectively use the data? What is lacking?
10. Do you see how existing field data could be put to better use?
11. What added value could sectoral data provide for your work or the public?

Questions for the representative of the data governance coordinating support unit:

For introduction, please describe in your own words what your position entails and what your main responsibilities involve.

The importance of data and its governing

1. Is the data considered important in your administration? What are the main data use cases?
2. Does data governance play an important role in your administration? What does it mean?
3. Is data management or governance mentioned in one or more of your governance strategies?
4. What is the main benefit and added value that data governance brings to your domain? Do managers and data owners see the benefits and added value of data governance similarly?

State of data governance and established processes

5. How do you assess the quality and reliability of data in the administrative field?
6. In your experience, what are the main problems with data governance?
7. What are the key entities and roles within your administration that are involved in data governance?
8. Do you think that there are enough people and competences to manage and govern data effectively? What is lacking?
9. Does your administration have clear and agreed data governance rules and policies? How do they develop?
10. What mechanisms/processes do you have in place to ensure that data is governed according to agreed rules and policies?

Data governance perspective

11. How does your administration keep up to date with changing data governance regulations and best practices?
12. What new trends or technologies do you think might impact data governance in the future?

Interview questions in Estonian.

Küsimused sisulise valitsemisala juhile:

Sissejuhatuseks, palun kirjeldage enda sõnadega, mis teie ametipositisoon endast kujutab, mis on teie peamised vastutusalad.

Valdkonna eesmärgid ja fookus, olulisemad andmed

1. Mis on hetkel teie valdkondlikud suuremad eesmärgid ja fookus?
2. Millist infot on teil eesmärkide täitmiseks vaja, millised on valdkonna kõige olulisemad andmed?

Andmete kasutusjuhud, andmepõhine juhtimine, andmete kvaliteet ja usaldusväärsus

3. Mis on teie valdkonna andmete kõige olulisemad kasutusjuhud?
4. Kas ja kuidas te mõõdate valdkonna arengut ja edukust?
5. Kui olulised on kvaliteetsed ja usaldusväärsed andmed teie valdkonna jaoks?
6. Kuidas hindad olukorda andmete kvaliteediga? Kas see on murekoht Teie valdkonnas?

Andmetega seotud probleemid, arengukohad ja perspektiiv.

7. Kas tunnete, et olemasolevad andmed aitavad kaasa valdkondlike otsuste tegemisel?
8. Millised on peamised probleemid olemasolevate andmete haldamisel?
9. Kas näete, et andmete haldamiseks ja efektiivseks kasutamiseks on inimesi ja kompetentse piisavalt? Millest on puudus?
10. Kas te näete, et olemasolevad valdkonna andmeid saaks kuidagi veel paremini ära kasutada?
11. Millist lisaväärtust võiksid valdkondlikud andmed teile töö tegemiseks või avalikkusele veel pakkuda?

Küsimused andmehaldust koordineeriva tugiüksuse esindajale:

Sissejuhatuseks, palun kirjeldage enda sõnadega, mis teie ametipositisoon endast kujutab, mis on teie peamised vastutusalad.

Andmete ja nende halduse olulisus

1. Kas andmeid peetakse teie haldusalas oluliseks? Millised on peamised andmete kasutusjuhud?
2. Kas andmehaldus mängib teie haldusalas olulist rolli? Milles see väljendub?
3. Kas ühes või mõnes teie haldusala strateegias on mainitud andmehaldust? Millises? Miks mitte?
4. Milline on peamine kasu ja lisandväärtus, mida andmehaldus teie haldusalasse toob? Kas juhid ja andmete omanikud näevad andmehalduse kasu ja lisandväärtust sarnasena?

Andmehalduse olukord ja väljakujunenud protsessid

5. Kas ja kuidas te hindate andmete kvaliteeti ja usaldusväärsust haldusalas?
6. Millised on teie kogemusest lähtuvalt peamised probleemid andmete haldamisel?
7. Millised on peamised üksused ja rollid teie haldusalas, kes tegelevad andmete haldusega?
8. Kas näete, et andmete efektiivseks haldamiseks on piisavalt inimesi ja kompetentse? Millest on puudus?
9. Kas teie haldusalas on selged ja kokkulepitud andmehalduse reeglid ja poliitika? Kuidas need välja kujunevad?
10. Kas ja millised mehhanismid/protsessid on teil kasutusel, et tagada andmete haldus vastavalt kokkulepitud reeglitele ja poliitikatele?

Andmehalduse perspektiiv

11. Kas ja kuidas hoiab teie haldusala ennast kursis muutuvate andmehaldus regulatsioonide ja parimate praktikatega? Kas lähenemine on pigem reaktiivne või proaktiivne?
12. Millised uued trendid või tehnoloogiad võivad andmehaldust teie arvates tulevikus mõjutada?

5.3 Appendix 2 - Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Liisbeth Laasik

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “Data Governance in the Estonian Government Sector: The Motives and Strategies for Implementation”, supervised by Innar Liiv
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

06.05.2023

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.