Government data-driven decision-making (DDDM) framework implementation. Test case: crisis management

Deliverable 1.3: Evaluation of alternative to-be scenarios

Technical Support Instrument

Supporting reforms in 27 Member States

Funded by the European Union

Republic of Estonia
Government Office

This document was produced with the financial assistance of the European Union. Its content is the sole responsibility of the author(s). The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

The project is funded by the European Union via the Technical Support Instrument, managed by the European Commission Directorate-General for Structural Reform Support.

This report has been delivered in June 2022, under the EC Contract No. REFORM/SC2021/076. It has been delivered as part of the project "Government data-driven decision-making (DDDM) framework implementation. Test case: crisis management".

Directorate-General for Structural Reform Support
REFORM@ec.europa.eu
+32 2 299 11 11 (Commission switchboard)
European Commission
Rue de la Loi 170 / Wetstraat 170
1049 Brussels, Belgium

Executive summary

Scope of the Project

This report has been developed within the Project carried out by PricewaterhouseCoopers EU Services EESV (hereinafter – PwC) on behalf of the DG REFORM, according to the specific contract No. REFORM/SC2021/076 (21EE02), signed on October 14, 2021. The report covers the items required in the Request for Service (RfS).

This report covers Outcome 1 of this Project – **Government data-driven decision-making**. Separate reports are issued for Outcome 2 and 3 which all combined make up the complete package of deliverables.

Purpose of the Project and Report

The report has been drafted for the purpose to describe the vision, functionalities, and implementation alternatives of the Government data-driven decision-making (DDDM) system, the foundation of the vision and the government memorandum to-be business process.

The report introduces and describes the DDDM system functionalities in two categories:

1) Main Functionalities of the Technological Solution,
2) Support Functionalities of the Technological Solution.

The report presents and analyses the pros and cons of three different implementation alternatives of the visioned DDDM system.

Key findings

The objective of the vision is to reach fully automated and data-driven end-to-end decision-making process, where the Government Memorandum is generated by intelligent technical solution. The visioned DDDM system is based on four pillars: Data, Automation, Intelligence, and Reliability.

DDDM system consists of five core components: Preparation of Memorandum; Data Search; Data Processing; Data Analysis; Visualisation of Data Analysis Results.

Three different scenarios can be used to achieve the objectives of the DDDM system, which are described in the report. The organisations included in the scenarios are Statistics Estonia, Government Office, and Ministry of Justice, which have various competences and support organisation to develop and administer the visioned functionalities.

The following scenarios were introduced:

- Distributed System – in distributed system scenario the core components of the DDDM system are distributed between Statistics Estonia, Government Office, and Ministry of Justice, where each of the organisations are responsible for one or several components, they have experience or developments that could be reused for DDDM system.
- Government Office System – in this scenario the development and administration of all core components of DDDM system are in the hands of one authority, tasks are better coordinated, and system is located closer to the Government.
- Statistics Estonia System – in this scenario the existing developments and technologies of Statistics Estonia are fully utilised, all core components of DDDM system are in the hands of one authority and this scenario would be the fastest to start to achieve the DDDM goals.

There will be several innovations in the development of the DDDM system. The biggest leap if we look at the current situation in Estonia is the mixing of structured and unstructured data analysis in the same system. There are no such solutions today in use where the user can analyse microdata, aggregated data, and so-called unstructured data together.

Recommendations

Recommendations regarding the DDDM system development are:

a. Collaboration – introduce the established vision of the DDDM system to stakeholders and potential users to invite their ideas and collect feedback;

b. Communication of expected benefits – explain to stakeholders and users what the benefits are of using the DDDM system;

c. Validation – validate the envisioned system functionalities with potential DDDM systems;

d. Prioritisation – prioritise the functionalities and the development tasks;

e. Data sources specification – specify the data sources used in the DDDM system to understand their legal statuses and special restrictions and explore how to address these;

f. Dividing the system development into parts – approach the DDDM system development step by step, by focusing on developing the highest priority components;

g. Test the system with users – involve the system users in development process from the very beginning to test the system's functionalities on ongoing basis;

h. Analyse the opportunities of deploying AI-based solutions – when AI solutions become available, it is important to test them and assess their use possibilities in DDDM system.

Lühikokkuvõte

Aruande eesmärk ja ulatus

Aruanne on koostatud Euroopa Komisjoni struktuurireformide toe peadirektoriaadi (DG REFORM) tellimusel PricewaterhouseCoopers EU Services EESV (edaspidi PwC) poolt läbiviidud Projekti raames vastavalt 14. oktoobril 2021. aastal allkirjastatud lepingule nr REFORM/SC2021/076 (21EE02). Aruande koostamisel on lähtutud Projekti lähteülesandes esitatud nõuetest.

Aruandes kajastatakse ainult Projekti esimese tulemiga piiritletud teemasid – **andmepõhise otsustusprotsessi edendamine**. Eraldi aruanded väljastatakse Projekti teise ja kolmanda tulemi kohta, mis kokku moodustavad lepingus ettenähtud väljundid.

Käesolev aruanne on koostatud eesmärgiga anda ülevaade valitsuse andmepõhise otsustussüsteemi (DDDM) visioonist, funktsionaalsustest ja rakendusalternatiividest, visiooni põhialustest ja Vabariigi Valitsuse memorandumi koostamise protsessi tulevikuvaatest. Aruandes tutvustatakse ja kirjeldatakse DDDM süsteemi funktsioone kahes kategoorias:

1) tehnoloogilise lahenduse põhifunktsioonid,
2) tehnoloogilise lahenduse abifunktsioonid.

Aruandes esitletakse ja analüüsitakse DDDM-süsteemi kolme erineva rakendusalternatiivi plusse ja miinuseid.

Tähelepanekud

Visiooni eesmärk on jõuda täielikult automatiseeritud andmepõhise otsustusprotsessini, kus Vabariigi Valitsuse memorandum koostatakse intelligentse tehnilise lahenduse abil. DDDM süsteemi visioon põhineb neljal sambal: andmed, automatiseeritus, intelligentsus ja usaldusväärsus.

DDDM süsteem koosneb viiest põhikomponendist: memorandumi koostamine; andmete otsing; andmetöötlus; andmete analüüs; andmeanalüüsi tulemuste visualiseerimine.

Aruandes kirjeldatud DDDM-süsteemi eesmärkide saavutamiseks on välja pakutud ja kirjeldatud kolme erinevat stsenaariumi. Stsenaariumidesse kaasatud organisatsioonid on Statistikaamet, Riigikantselei ja Justiitsministeerium, kellel on erinevad pädevused ja organisatsioon visioonis kirjeldatud funktsionaalsuste arendamiseks ja haldamiseks.

Aruandes tutvustatakse järgmisi stsenaariume:

- Hajutatud süsteem – hajutatud süsteemi stsenaariumi korral on DDDM süsteemi põhikomponendid jaotatud Statistikaameti, Riigikantselei ja Justiitsministeeriumi vahel, kus iga organisatsioon vastutab ühe või mitme komponendi eest, kuna neil on kogemusi või varasemalt tehtud arendusi, mida on võimalik DDDM-süsteemi huvides uuesti kasutada.
- Riigikantselei süsteem – antud stsenaariumi korral on DDDM süsteemi kõigi põhikomponentide arendus ja haldamine ühe asutuse käes, ülesanded on paremini koordineeritud ja süsteem asub valitsusele lähemal.
- Statistikaameti süsteem – antud stsenaariumi puhul on Statistikaameti senised arendused ja tehnoloogiad täielikult ära kasutatud, kõik DDDM süsteemi põhikomponendid on ühe asutuse käes ja see stsenaarium võimaldaks kõige kiiremini DDDM süsteemi eesmärke täitma asuda.

DDDM süsteemi arendamisel on ette näha mitmete innovaatiliste lahendustuste kasutuselevõttu. Suurim hüpe, kui vaadata Eesti hetkeolukorda, on struktureeritud ja struktureerimata andmeanalüüsi kooskasutamine ja kombineerimine samas süsteemis. Selliseid lahendusi, kus kasutaja saaks koos analüüsida mikroandmeid, koondandmeid ja nn struktureerimata andmeid, täna kasutusel ei ole.

Soovitused ja head praktikad

Soovitused DDDM-süsteemi arendamiseks on järgmised:

- Koostöö – tutvustada sidusrühmadele ja potentsiaalsetele kasutajatele DDDM süsteemi visiooni, koguda neilt tagasisidet ja täiendavaid ideid;

- Kasutegurite kommunikatsioon – selgitada huvirühmadele ja kasutajatele, milline kasu tekib neile DDDM-i kasutamisest ja kuidas see nende igapäevast tööd saab lihtsustada;
- Valideerimine – kavandatud süsteemi funktsionaalsuste valideerimine DDDM-süsteemi potentsiaalsete kasutajatega;
- Prioriteetide seadmine – funktsionaalsuste ja arendusülesannete tähtsuse järjekorda seadmine;
- Andmeallikate spetsifikatsioon – DDDM-süsteemis kasutatavate andmeallikate täpsustamine, et mõista nende õiguslikku staatust ja eripiiranguid, millega on vaja süsteemi arendamisel arvestada;
- Süsteemi arenduse osadeks jagamine – soovitus läheneda DDDM süsteemi arendamisele samm-sammult etappide lõikes, keskendudes kõrgeima prioriteediga funktsionaalsuste arendamisele esmajärjekorras;
- Süsteemi testimine kasutajatega – süsteemi tulevaste kasutajate kaasamine arendusprotsessi algusest peale, et jooksvalt testida süsteemi toimivust;
- AI-põhiste lahenduste juurutamise võimaluste analüüsimine – erinevate tehisintellektilahenduste turule tulemisel on oluline neid testida ja hinnata nende kasutuselevõtu võimalusi DDDM süsteemis.

Table of Contents

1. Introduction

1.1 Scope of the Report

1.1.1 Purpose of the Report

The report has been drafted for the purpose described in the vision of data-driven decision-making system, to propose alternative solutions and functionalities of the system to improve the current DDDM process.

The report includes a gap analysis and proposes a recommendation to the Beneficiary of which system functionalities to include in the future DDDM system as well as outlines the recommended implementation alternative.

The approach and results of the topics are described in respective paragraphs.

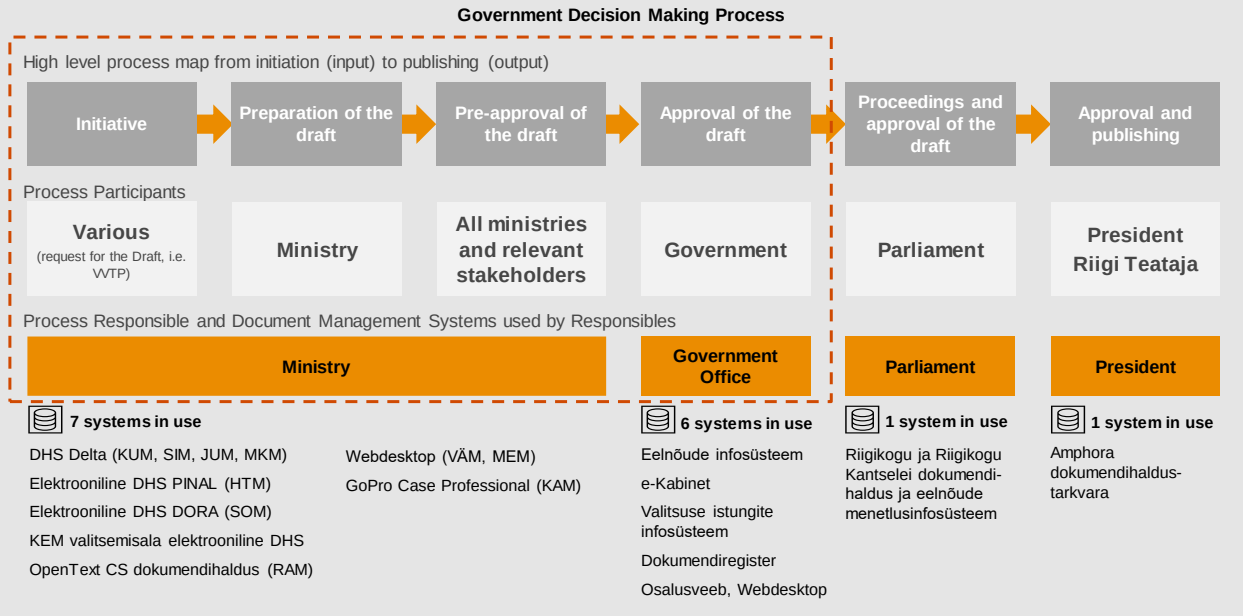
This report covers only Outcome 1 – Government data-driven decision-making framework implementation. Separate reports are issued for Outcome 2 and 3 which all combined make up the complete package of deliverables.

1.1.2 Scope of Outcome 1

Decision-making process in general involves number of institutions (Ministry, Government Office, Parliament, President) as described on Figure 1. As there are a lot of different legal types of Documents and decisions in Estonia (described in Deliverable 1.1.), the level, extent and course of the decision-making process varies.

The Project Scope approved in Deliverable 1.1. **covers the areas of responsibility of Ministries and Government Office** as shown on Figure 1.

Figure 1. Scope of the Project by Institutions in Outcome 1



It was acknowledged that certain types of Documents are handed over to the Parliament for proceedings and approvals as well as Legal Drafts go to the President for announcement and publishing at Riigi Teataja, but considering the purpose of the Project, **the working process and practices at the Parliament and President are not covered**.

In short, the Project Scope covers the following:

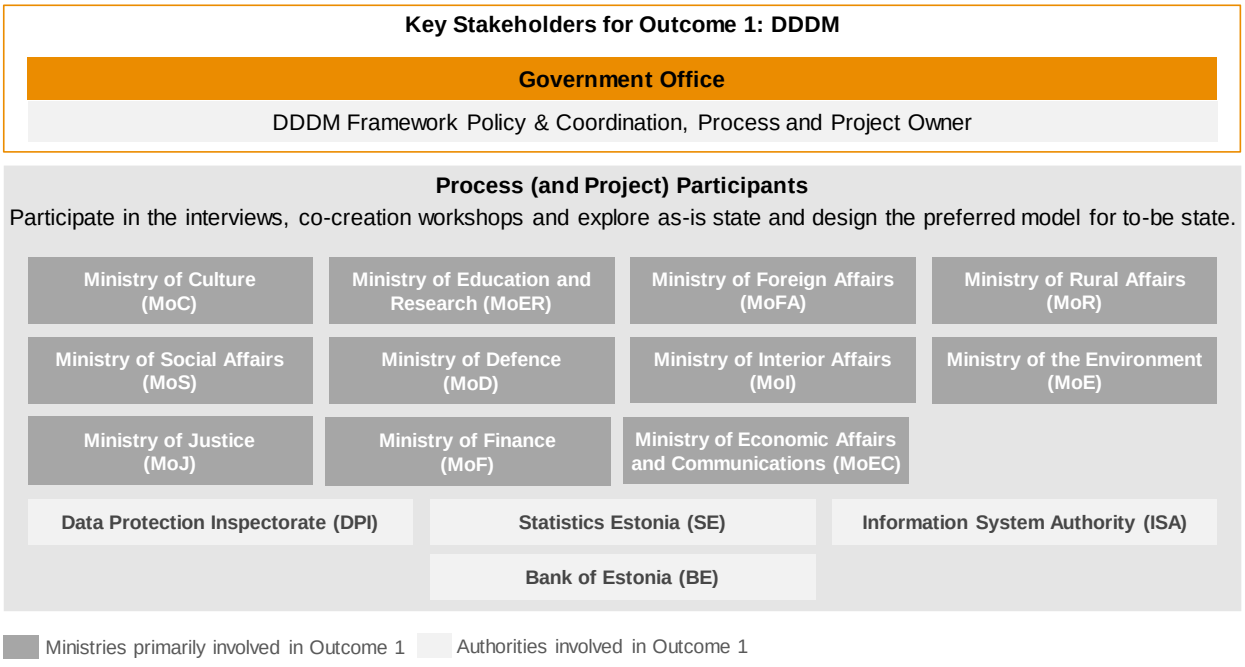
Table 1. Project Scope

Area	Description
1. Institutions	Process Responsible: • Ministries • Government Office
2. Document Type	• Government Memorandum
3. Process	• End-to-End process of Government Memorandum • End-to-end describes a process that takes the process from beginning to end and delivers a complete output for Government decision-making
4. Data and Technology	• Data and Technology used in the process of Government Memorandum
5. People	• Participants and decision-makers such as public servants and/or third parties (i.e., subject matter experts) involved in the process of Government Memorandum

1.1.3 Project Stakeholders for Outcome 1

To conduct an effective stakeholder engagement, we have identified the following key stakeholders and process participants for the Outcomes 1 (Figure 2) who are participating in the Project work.

Figure 2. Outcomes 1: Key Stakeholders and Project Participants

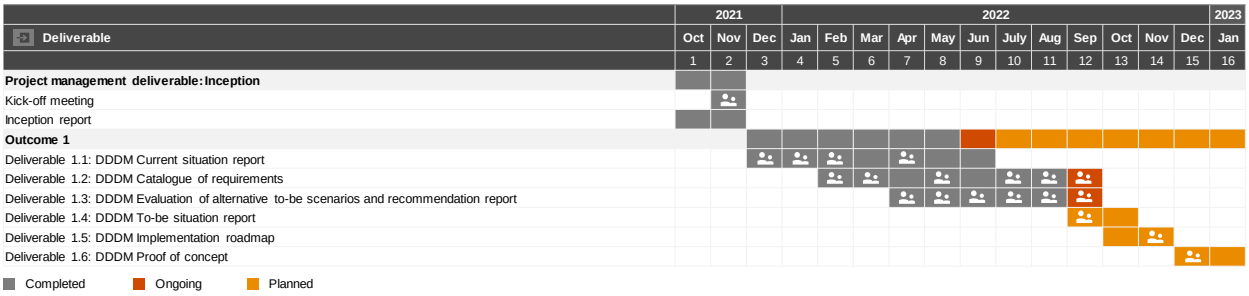


1.2 Project Timeline

1.2.1 Timeline

Figure 3 gives a high-level overview of the project activities and timeline. The activities of the third deliverable took place from April 2022 to September 2022 and were in progress at the time of issuing the draft version in early September 2022.

Figure 3. Project Activities and Timeline



By the time of issuing the final report in September 2022, all planned tasks and activities have been completed.

2. DDDM System

2.1 Vision

2.1.1 Introduction of the Vision

The vision of the Government Office of Estonia is to develop a **highly ambitious technical solution** that supports the Government's data-driven decision-making process by providing fully automated analytical overviews of various subject matters and decision proposals for discussion at the Government's weekly cabinet meetings. The outcome of the vision will replace and innovate the current approach to creating the Government Memorandum.

The objective of the vision is to reach fully automated and data-driven end-to-end decision-making process, where the Government Memorandum is generated by intelligent technical solution. The input to the process is provided by the official who briefly describes the problem or a topic that will need analytical insight. This problem statement is an input to the automated decision-making process.

As visioned, the future solution should be able to synthesise the given problem statement, search and find all relevant data on topics described in the problem statement, produce data visualisations that include and combine all relevant data available on the topic, describe and analyse the subject matter at hand, and provide decision proposals for the Government to discuss at the weekly cabinet meetings. Decision proposals generated by the technical solution are the output of the automated process.

The reason for the vision is to:

- innovate the Government Memorandum process,
- use more data throughout the decision-making process and
- accelerate the data analytics process and provide decision proposals faster to the Government.

2.1.2 Foundation of the Vision

The vision is based on four pillars (see Figure 4) that together create the foundation for the fully automated and data-driven end-to-end decision-making process.

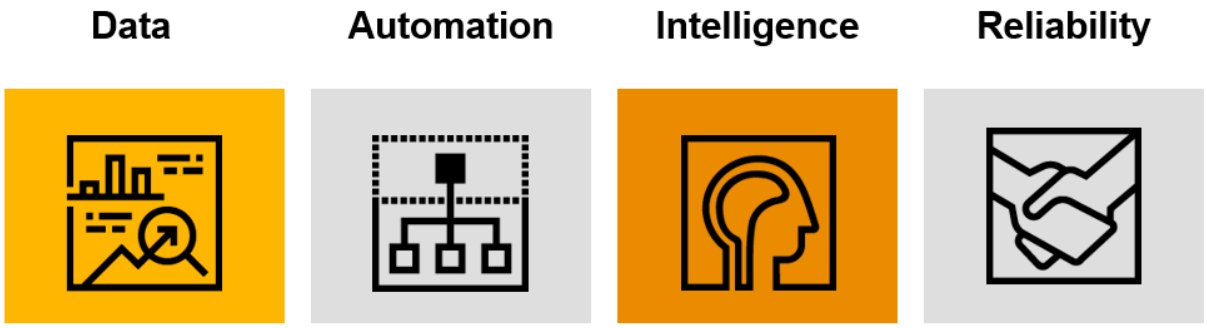


Figure 4. Foundation pillars of the Vision

Data pillar represents two key principles. First principle is that data quality will be improved throughout the process, because errors in datasets will become visible on data visualisations and once errors are identified, they can be continuously solved. Second principle is that by applying the vision, sufficient amount of data will be used during preparation of the decision proposal.

Automation pillar represents two key principles. First principle is that data acquisition and data transformation is automated to an extent that is possible at given time. Second principle is that the steps of the process that can be automated, will be automated, but at least in first phases of implementing the technical solution, human and manual interaction will be present in the decision-making process.

Intelligence pillar represents five key principles. First principle is that the currently available opportunities of machine learning will be considered and taken into account when developing the technical solution. Second principle is that the future opportunities of the machine learning will be applied to the solution when desired machine learning functionalities become available to use in the future. Third principle is that the solution will have modern and innovative functionalities that meet and exceed the needs of decision-making process participants. Fourth principle is that the solution will provide proper and accurate analytical outputs and recommendations to the decision-making process participants. Fifth principle is that the user interface of the solution will be intuitive to use, user-friendly and it will provide positive overall user experience.

Reliability pillar represents two key principles. First principle is transparency that will be enhanced due to the direct links to data sources that will enable users to directly go to the data source and review the data used in the decision-making process. Second principle is traceability which will mean that the data sources are correctly referenced and when needed, could be traced back to the original source (data provenance), and data transformation steps are explained and can be repeated when for example, data is updated, and analysis requires an update as well.

2.1.3 Government Memorandum TO-BE Business Process

To reach the vision of fully automated data-driven decision-making process, **several to-be business processes** of **different maturity levels** are necessary to be developed and applied along the way of moving towards the ambition. Once initial modules and functionalities are developed, only then it is suggested to move on to development of modules and functionalities that are part of the next maturity level, the second to-be business process.

The final ambition of the vision is to reach fully automated to-be process, meaning the end-to-end process of creating Government Memorandum is done by highly intelligent technical solution.

To-be vision of the Government Memorandum business process consists of eight high-level process steps, where **steps 3-7 are aimed to be fully automated** and delivered by the intelligent technical solution. Eight high-level process steps of the Government Memorandum business process (see Figure 5) are following:

1. Request for creating the Government Memorandum
2. Describing the problem or topic
3. Searching for data
4. Selecting relevant data
5. Creating data visualisations
6. Describing and analysing the situation
7. Proposing the decision based on analysis
8. Discussing the situation and decision proposal at the Government's cabinet meeting.

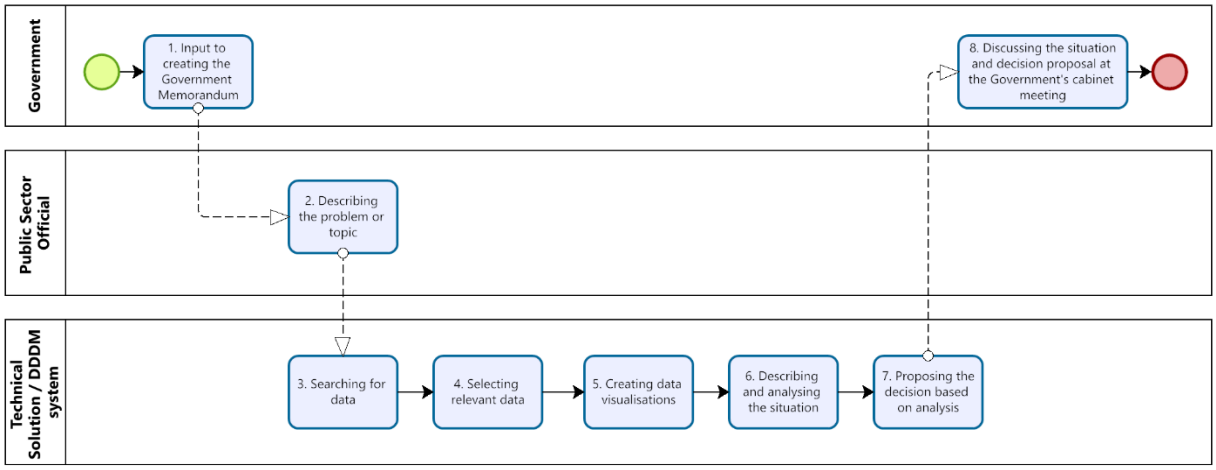


Figure 5. Final Ambition of the Vision: high-level TO-BE business process of the fully automated creation of the Government Memorandum

In the following chapters, three phases – maturity levels – of the to-be business processes are outlined. These phases can be considered as three phases of developing and implementing the technical solution. Each phase of the to-be process is focused on developing and solving one key area of the desired future state.

Key areas of the to-be processes are:

1. Data Access,
2. Automation of Data Transformation,
3. Automation of Analysis.

Table 2. Comparison of the TO-BE business process maturity levels

		Maturity level					
		Phase 1. Data Access		Phase 2. Automation of Data Transformation		Phase 3. Automation of Analysis	
	Government Memorandum business process	Human	Machine	Human	Machine	Human	Machine
Identified need or problem	1. Request for creating the Government Memorandum	√		√		√	
Input	2. Describing the problem or topic	√		√		√	
	3. Searching for data		√		√		√
	4. Selecting relevant data	√		√			√
Output	5. Creating data visualisations	√			√		√
	6. Describing and analysing the situation	√		√			√
	7. Proposing the decision based on analysis	√		√			√
Decision	8. Discussing the situation and decision proposal at the Government's cabinet meeting.	√		√		√	

Maturity level of the process and the technical solution will increase step by step (phase by phase) and once the third key area is solved, the final ambition of the vision of fully automating the decision-process can be achieved.

2.1.3.1 First phase of the TO-BE Business Process

The first phase of the to-be process is focused on solving the data access matter, so that the desired data search functionality becomes available for the Government Memorandum creators.

To enable the upgrade of the as-is process to desired to-be process (see Figure 6), in the first phase, user interface with desired functionalities need to be developed that allow the user (creator of the Government Memorandum) to complete the end-to-end process of Government Memorandum creation in one environment (web-based user interface).

In the first phase, steps 1-2 and 4-8 will be completed by the officials and most of the work on compiling the Government Memorandum will be manual, as the automated functionalities for data visualisation and analytical interpretation of data and the situation are not developed yet. Step 3 – searching for data – will be completed by the machine, the technical solution. In first phase, all necessary activities should be taken to build a functioning search engine, so the machine can search for data from pre-defined data sources from Estonian public sector authorities.

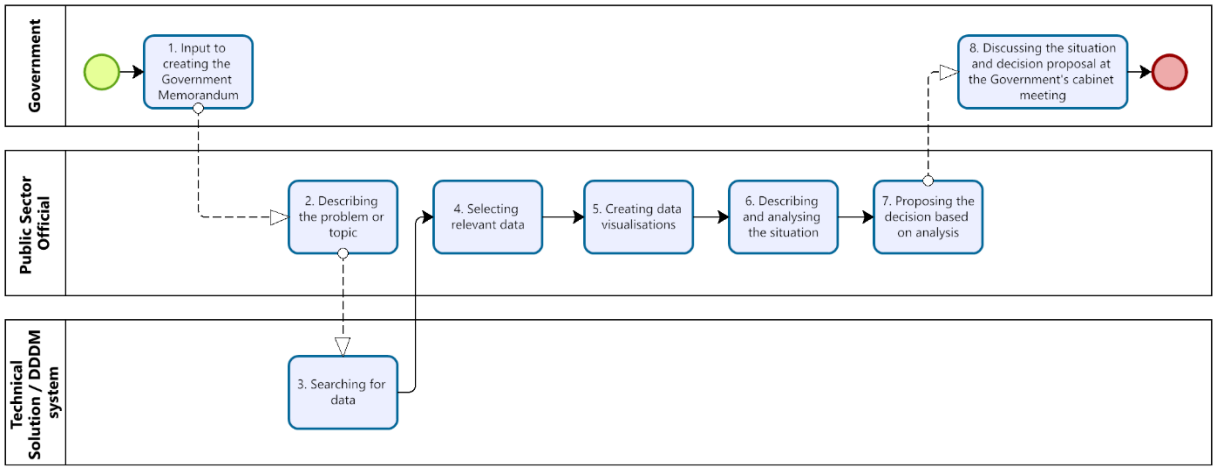


Figure 6. Phase 1 of the TO-BE business process

The aim of the first phase is to solve and minimise the number of current underlying problems regarding identification of relevant data sources, access to those data sources, data protection, data quality, data governance and other data related matters as described in the Deliverable 1.1 of this project.

An example of current underlying problems that need solving are

- Not all data is structured,
- Not all data sets are available due to technical or legal limitations,
- Poor overall data quality across public sector institutions,
- Data is protected and cannot be accessed via search function (data availability classes may not be sufficient),
- Due to large number of datasets and registers used by government sector (from both public and private data sources), the know-how of availability, location and accessibility of specific and relevant data is limited in civil service

Prerequisites to achieving the first phase are

- Officials across public sector are trained and onboarded to start using the new solution for creating Government Memorandums,
- Defining the scope of data sets to be included in the search functionality,
- Systematic practices, principles and approaches to data governance are in place, so that data across public sector can be integrated to the search function,
- Data sets have meta data that enable the search within data sets,

- Data volumes are roughly defined, and technical solution has enough capacity for sufficient functioning.

When the prerequisites are fulfilled, the search functionality can be developed and made available for testing and using, to create Government Memorandums by using more potentially relevant data and support the process of data-driven decision-making at Government's cabinet meetings.

The goal of the first phase is to save the officials' time and provide more comprehensive overview on searching for data, the time they would otherwise spend on identifying various data sources, writing requests to accessing the data because often data sources have limited access and data protection rules in place. When there is finally access to data sets, then analysing the data content itself, to understand whether the data set includes specific data that would be useful for the specific Government Memorandum to be created.

Current estimated time for identifying, collecting, and selecting relevant data is 2-4 weeks of work per one memorandum and this time can be narrowed to 1 day or less per memorandum by providing the data search functionality to the officials who are compiling the Government Memorandum documents.

So, in one year, when there are approximately 150 memorandums created and each memorandum takes on average 3 weeks to get data (plus additional time to complete all other steps of the memorandum process), then a total working time of 450 weeks / 2 250 business days to get only the data can be narrowed down to approximately 150 working days or less by having the search available. The effort varies depend on what is the topic of the analysis. The new system has features and data that were not available before, and the user can perform analyses that previously would have required a very large effort or were impossible.

2.1.3.2 Second phase of the TO-BE business process

The second phase of the to-be process is focused on solving the **automation of data transformation** to build the functionality of creating dynamic data visualisations and basic dynamic descriptions of the situation, based on the visualisations.

When in the first phase of the to-be process, the official is selecting data and building data visualisations manually, then in the second phase, the technical solution needs to be developed to a higher intelligence level, where the machine itself is able to visualise the selected data into suitable visualisations.

The aim is to have the machine so smart that it can understand the data structure, find the best type of visuals for the specific data structure, and create one or several visualisation alternatives of the data, of which the official can then pick the most suitable one for using in the memorandum.

In second phase (see Figure 7), steps 1-2, 4 and 6-8 will be completed by the officials and approximately half of the work on compiling the Government Memorandum will be manual, as the automated functionality for the analytical interpretation of the situation is not developed yet. Steps 3 and 5 – creating dynamic data visualisations – will be completed by the machine, the technical solution.

In the second phase, all necessary activities should be taken to build the base for creating the data visualisations automatically by the machine. Here, one of the critical success factors of building this functionality is the input from the officials/analysts who are creating (complex/custom) visualisations by using scripts. The aim is to build a repository of applications (collection of scripts, program codes, algorithms) that can be used to teach the machine and eventually have the machine to use these various data visualisation applications automatically to create complex/custom visuals.

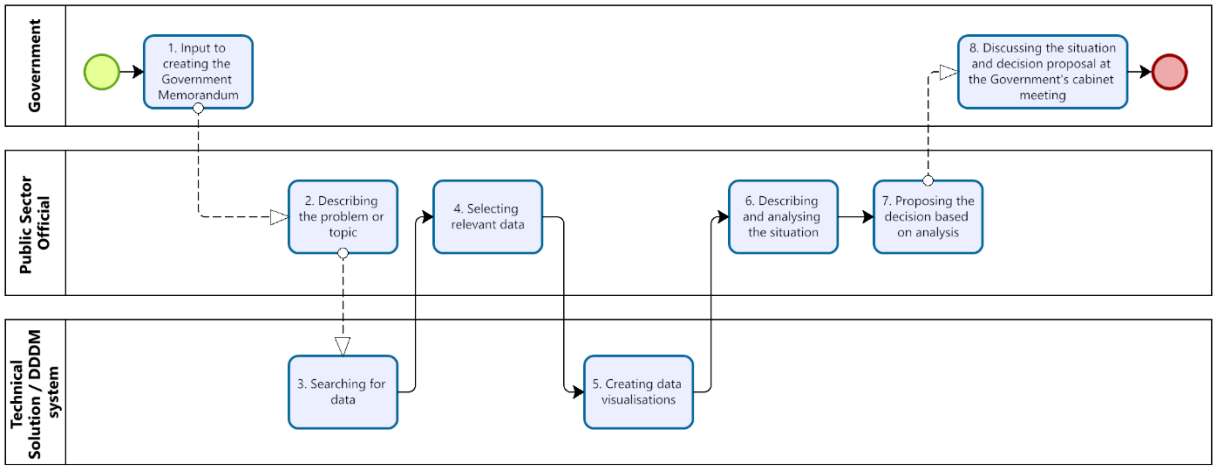


Figure 7. Phase 2 of the TO-BE business process

The aim of the second phase is to solve and minimise the number of current underlying problems regarding data transformation to be able to automate data transformation tasks and the creation of dynamic data visualisations.

An example of current underlying problems that need solving are

- There is no clear identification system in place for identifying/describing data objects, therefore data sets from several different sources cannot be combined with each other.

Prerequisites to achieving the second phase are

- There is a system in place for identifying/describing data objects, so that data sets can be combined with each other,
- There is a system in place for classification of data objects, so that data can be visualised.

The goal of the second phase is to provide the officials dynamic data visualisations and basic dynamic descriptions of the situation, that are done automatically, based on the relevant data the official has selected.

Currently, no dynamic data visualisations are used to support the analytical content of the Government Memorandum, as the data analysis is provided in static written documents. Therefore, having functionality of dynamic data visualisations, which are moreover done by the machine, will highly upgrade the process of creating Government Memorandums as well as save significant amount of time the officials will otherwise spend on manually working on creating the visualisations.

The amount of time that is spent on visualising the data is significant, because the overall skills of working with quantitative data, especially skills of data transformation and visualisation are considered low or very low across public sector. Of course, there are exceptions of highly skilled officials and analysts who are familiar to this type of work with data, but the number of such specialists is limited.

2.1.3.3 Third phase of the TO-BE business process

The third phase of the to-be process is focused on solving the **automation of analysis**, to building the functionality of describing and analysing the situation automatically by the machine.

In the third phase (see Figure 8), steps 1-2, 4 and 8 will be completed by the officials and these steps basically cover the task of giving an input to the technical solution to be solved, selecting relevant data, and reviewing the output, which is the Government Memorandum including dynamic data visualisations, highly intellectual analytical description, and interpretation of the situation together with decision proposals. Steps 3 and 5-7 will be completed by the technical solution, which are the most complex and in the current process the most time-consuming tasks.

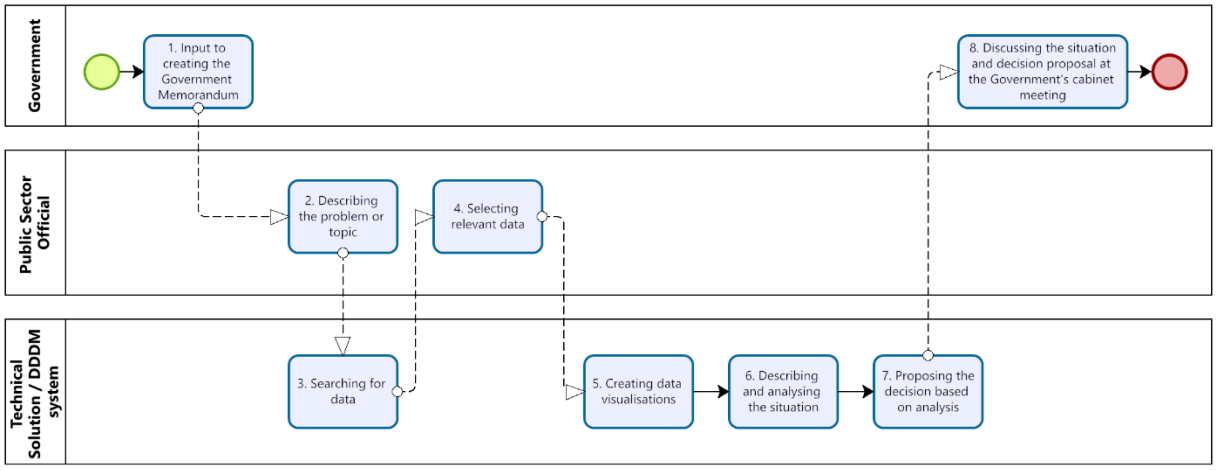


Figure 8. Phase 3 of the TO-BE business process

The aim of the third phase and the vision is to build a fully automated process, where the most intellectually demanding, complex and complicated tasks of interpreting the situation and proposing decisions for Government's discussion, are solved by the highly intellectual technical solution.

An example of current underlying problems that need solving are

- There is no such highly intelligent AI system available yet in the technological world that would be able to deliver highly analytical data and situation interpretation, as well as decision proposals by itself in Estonian language.

Prerequisites to achieving the third phase are

- Successful completion of phases 1 and 2,
- AI solutions are available in the world, that would enable the development of desired automated functions,
- Fully automated solutions are available or can be developed in Estonian language.

When the prerequisites are fulfilled, the automated situation interpretation functionality can be developed and made available for testing and using, to create Government Memorandums in a more efficient, data-driven, and technically savvy way, with minimal time and intellectual effort required from the officials, to support the process of data-driven decision-making at Government's cabinet meetings.

2.1.4 Government Memorandum 2.0

DDDM system shall produce **the Government Memorandum 2.0 output** (see Figure 9) that includes a compact overview of memorandum components, which could be clicked to discover more details and visuals could be zoomed in or out to analyse the data content in greater detail.

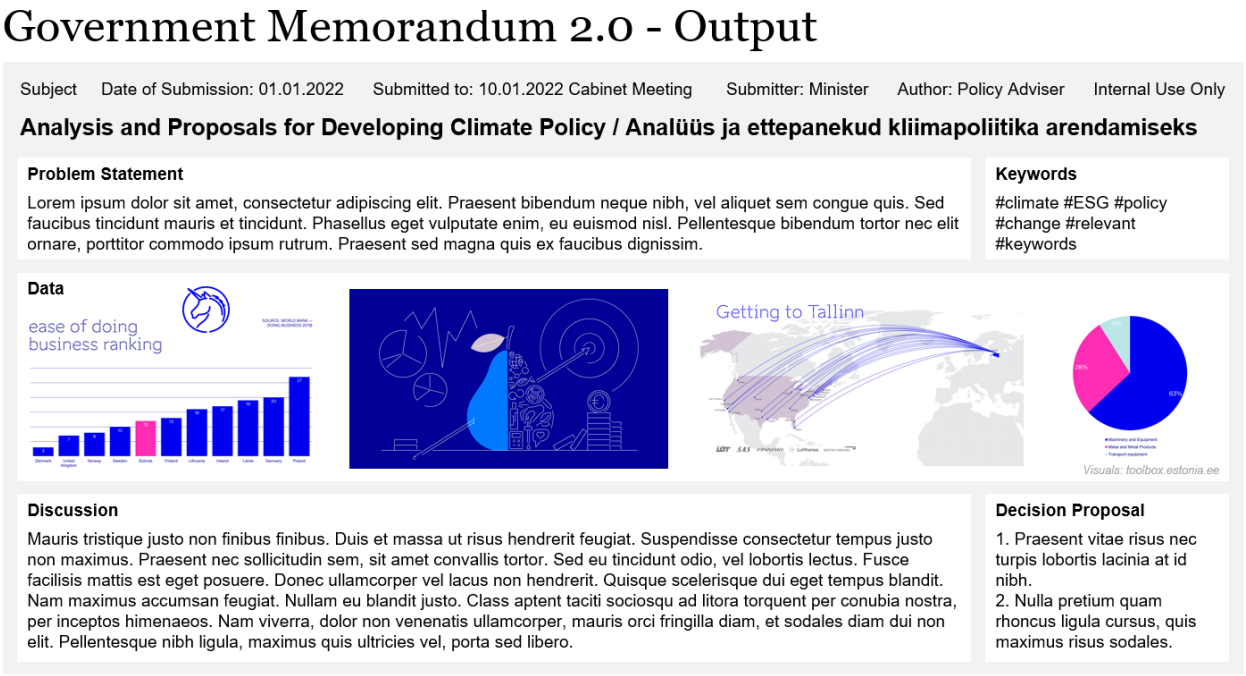


Figure 9. Government Memorandum 2.0 - Output

The process of creating the Government Memorandum 2.0 is described in Appendix 4.3 by introducing example screenshots of potential user interface. The structure of the Government Memorandum (i.e., title, problem statement, data, discussion, decision proposal) follows the current practice in Estonia as described in Deliverable 1.1. (current situation analysis).

2.1.5 Users

There are two distinguished group of users of the DDDM System:

1. Government Memorandum **Author**
2. Government Memorandum **Viewer**

Authors or creators of the Government Memorandums can edit and iterate the memorandum content and would have the access to data and perform analytics, while the viewers can access the memorandum and view and/or leave comments to the authors who could then consider modifying the content before finalising and submitting the memorandum to the government.

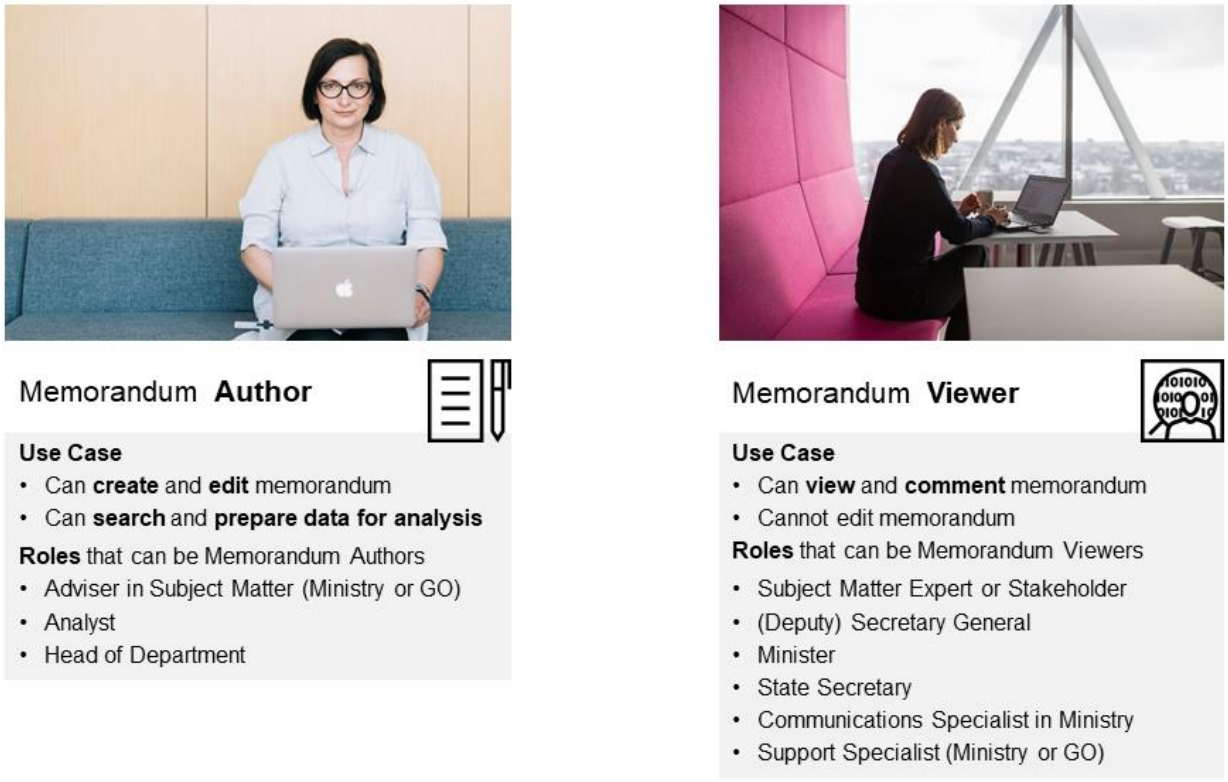


Figure 10. Government Memorandum 2.0 – Users/Collaborators

User stories and needs are considered throughout the design and development of the DDDM System. Furthermore, users form different level of authorities are invited to the development and validation process of the system.

2.2 Gap Analysis

Gap Analysis of the Government Memorandum AS-IS and TO-BE Business Processes

To better understand what the gaps are between the current (AS-IS) situation and the future (TO-BE) state of Government Memorandum business processes, a comparative analysis was conducted. The analysis compared the participants and the milestones of both processes to identify the gaps.

AS-IS business process

The AS-IS business process consists of 8 milestones, where each milestone includes one to several tasks to completed. AS-IS process milestones/stages are:

A. Initiation of the Government Memorandum preparation,

B. Adjustment of the Government Memorandum content,

C. Coordination of meetings and collaboration,

D. Data collection and analysis,

E. Preparation of the analysis of the Government Memorandum,

F. Collection of feedback to the Government Memorandum,

G. Coordination of approvals and signing the Government Memorandum,

H. Discussion of the Government Memorandum at the cabinet meeting.

In the AS-IS process of preparing the Government Memorandum, there are at least 14 roles and/or institutions participating who are giving their input/insights into the memorandum, conducting various tasks on making the memorandum document or giving their feedback to the initial version of the memorandum.

Collaboration, collecting insights in the first stages and collecting feedback and suggestions to amending the decision proposals is the collateral part of preparing the memorandum for the Government for discussion and decision-making.

TO-BE business process

The high-level TO-BE business process consists of 5 milestones, where each milestone includes one to several tasks to be completed. In TO-BE process, most of the visioned tasks are aimed to be completed by the technical solution - DDDM system.

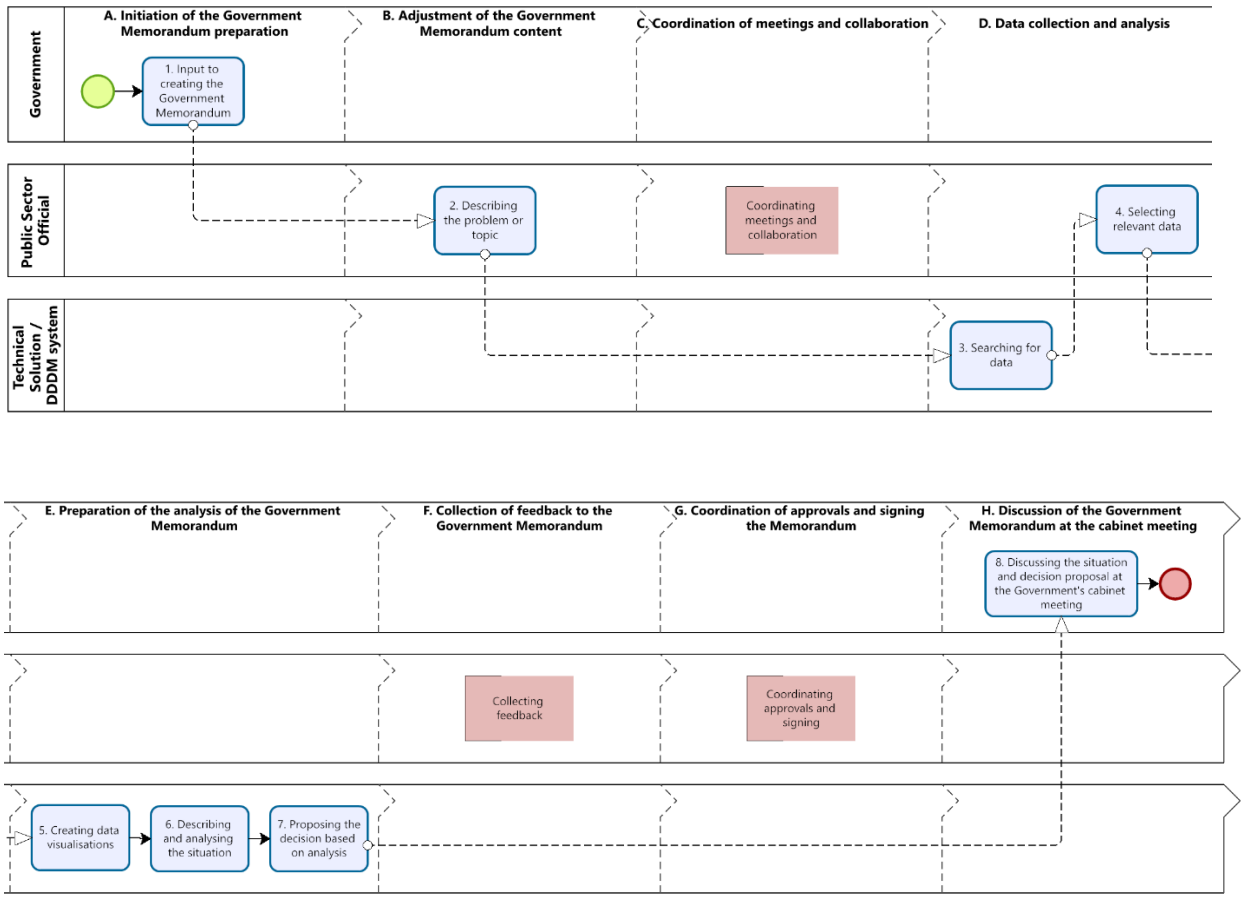


Figure 11. High-level Government Memorandum TO-BE business process

Figure 11 includes red comment boxes on milestones C, F and G, which illustrate that current tasks that are present at the AS-IS process, are not explicitly visioned to be part of the TO-BE process milestones:

A. Initiation of the Government Memorandum preparation,

B. Adjustment of Government Memorandum content,

C. **Coordination of meetings and collaboration** (this milestone's tasks are missing in the visioned TO-BE process),

D. Data collection and analysis,

E. Preparation of the analysis of Government Memorandum,

F. **Collection of feedback to Government Memorandum** (this milestone's tasks are missing in the visioned TO-BE process),

G. **Coordination of approvals and signing the Government Memorandum** (this milestone's tasks are missing in the visioned TO-BE process),

H. Discussion of the Government Memorandum at the cabinet meeting.

In the TO-BE process of preparing the Government Memorandum, there are 3 roles envisioned to be part of the process:

1. Government, who gives an input to start the memorandum preparation process (i.e. request for the work) and receives an output of analysis and decision proposals for discussion,
2. Public sector official / subject matter adviser, who prepares the memorandum supported by the DDDM system (also referred to as the Technical Solution),

3. DDDM system, that searches for data, creates data visualisations, analyses data and provides an analytical description/interpretation of the situation, and generates decision proposals.

Currently there are no collaboration activities in the TO-BE process visioned with subject matter stakeholders, although those activities are considered necessary parts of the process as explained by the process owners during the current situation mapping meetings. Because the decisions proposed to the Government should be aligned and agreed on with the stakeholders (subject matter organisations, ministries influenced by the future decision, others) before the decision is done in the Government. Collaboration during the creation (preparation) of the memorandum is important for communication between stakeholders as well as collecting qualitative data that can only be acquired through meetings and detailed discussions.

The following Table 3 gives an overview of the differences/gaps between the AS-IS and the TO-BE process milestones and outlines whether tasks within these milestones are automated or done manually.

Table 3. Comparison of the AS-IS and the TO-BE process milestones, including the gaps/differences of how those milestones are completed

Milestones of the AS-IS process		**Milestones of the TO-BE process desired in the vision**		**Gaps/Differences in the process milestones**
Tasks within this milestone are done manually	*Tasks within this milestone are automated*	*Tasks within this milestone are done manually*	*Tasks within this milestone are automated*	
Initiation of the Government Memorandum preparation	-	Initiation of the Government Memorandum preparation	-	In both processes the steps of this milestone will be done manually
Adjustment of Government Memorandum content	-	Adjustment of Government Memorandum content	-	In both processes the steps of this milestone will be done manually
Coordination of meetings and collaboration	-	N/A	N/A	Collaboration is yet to be visioned in fully automated decision-making process
Data collection and analysis	-	-	Data collection and analysis	In to-be process the steps of this milestone will be automated
Preparation of the analysis of Government Memorandum	-	-	Preparation of the analysis of Government Memorandum	In to-be process the steps of this milestone will be automated
Collection of feedback to Government Memorandum	-	N/A	Collection of feedback and comments	In to-be process the steps of this milestone will be automated
Coordination of approvals and signing the Government Memorandum	-	N/A	N/A	Coordination of approvals and signing the Government Memorandum is yet to be visioned in fully automated decision-making process
Discussion of the Government Memorandum at the cabinet meeting	-	Discussion of the Government Memorandum at the cabinet meeting	-	In both processes the steps of this milestone will be done manually

To conclude, the main differences between the AS-IS and TO-BE vision of the process are the milestones and tasks that include collaboration and collecting feedback.

Gaps between the AS-IS and TO-BE situations

In May 2022, a seminar was held between the Government Office and Deputy Secretary General of Ministries, to introduce and discuss the vision, as well as collect feedback to the idea of creating automated DDDM system.

During the meeting, Deputy Secretary Generals outlined and discussed several key gaps that exist between the AS-IS and TO-BE situations of data-driven decision-making process.

The gaps are following:

- Process as Communication Tool – the users of the memorandums are besides the government considered to be other ministries, and memorandum process together with the analytical output is a means of communication and information exchange between stakeholders to be informed on various policy matters and this information exchange should not be minimised/discarded.
- Poor Data Quality – data quality is poor; data is not standardised nor machine-readable and therefore the currently existing data cannot be used yet in the DDDM system.
- Limited Access to Data – data protection regulations restrict and limit the access to data.
- Limited Competences and Skills to work with Data Analytics Tools – public sector officials / subject matter advisers in ministries have limited overall competencies, skills, and interest in working with data analytics and business intelligence tools such as Tableau, MS Power BI or R.
- Lack of Guidelines and Methodologies to improve the Analytical Content of AS-IS Memorandums – there are no formal guidelines, methodologies or structured check-lists in place to help memorandum owners create the memorandum in a better, more analytical and data-driven way, Deputy Secretary Generals emphasised that as long as there are no rules or structured requirements for creating “a good memorandum” in place, the quality of memorandums’ analytical content will fluctuate because all memorandum creators are doing their best and this varies because people have different levels of data analytics and data visualisation competencies and experience.
- Uneven access to Data Analytics Tools – since the use of tools and technologies is decided case by case in public sector institutions, not all public sector officials can access the data analytics tools when they happen to work in an organisation where the tools are not available.
- Lack of Technologies – technological advancements are not there yet to be able to build/develop the desired AI-assisted DDDM system.
- Legal Framework – developing new information system requires the development of legal documents that regulate the use of the system.
- Unstructured data – there is no automated solution to unstructured data usage. It is needed to develop a tool for doing analysis on unstructured data, e.g., legal texts, web content etc.

As the DDDM system is visioned to be AI-based intelligent technical solution, it is important to highlight here the challenges, gaps and solutions mapped during the creation of Estonia’s National Artificial Intelligence Strategy for 2022-2023, as those challenges apply to the development of DDDM system as well.

Table 4. Challenges and solutions mapped during the creation of Estonia's National Artificial Intelligence Strategy for 2022-2023[1]

Challenges
1. Expectations from the business side of the application of the AI solutions cannot be met with today's level of technology;
2. There is a lack of IT support, resources, and commitment to projects;
3. Public authorities do not have sufficient support for the development, deployment, and subsequent management of the AI solutions;
4. There is a lack of overview of measures and initiatives to support implementation;
5. Organisations lack the competence and experience to assess the quality of data and how to pre-process data;
6. There is a lack of detailed overview of the solutions, tools, and co-developments that have been made, and the same solutions are being developed;
7. There is a lack of legal certainty about the processing of data, which delays project implementation;
8. Not being able to see how to implement the AI in the existing information systems and business processes.

AI strategy pinpoints important data-related obstacles that need overcoming and solving before AI-based solutions can be built:

„*The lack and insufficient quality of the data are major obstacles to the launch and development of the AI project. Data therefore plays an important role in the development of the AI. For projects to succeed, it is necessary to ensure that data is available in a machine-readable format*."

To close the gap and move step by step towards the TO-BE vision, alternatives of the functionalities of the technical solution / DDDM system are described in following chapters. A combination of those functionalities can formulate the base for future DDDM system.

[1] Ministry of Economic Affairs and Communications, 2021, Estonia's National Artificial Intelligence Strategy or Kratt Strategy for 2022–2023

2.3 DDDM System Functionality

The DDDM system as an automated data-driven system for supporting the Estonian Government decisions must have both an automated backend system and a modern web-based user interface. It's reasonable to use an existing system as a user interface, or as an extension of the DDDM system, e.g., KOOS[2].

All activities related to data-driven decision making must be automated as much as possible. The automation is possible in areas where we have enough micro data and digitalized knowledge. The digitalized knowledge is a fuel for the Artificial Intelligence (AI). The knowledge can grow using machine learning on **high-quality data**. A goal is fully automated system where the government can use **descriptive, predictive, and prescriptive analytics** for speeding up the high-quality decision-making processes.

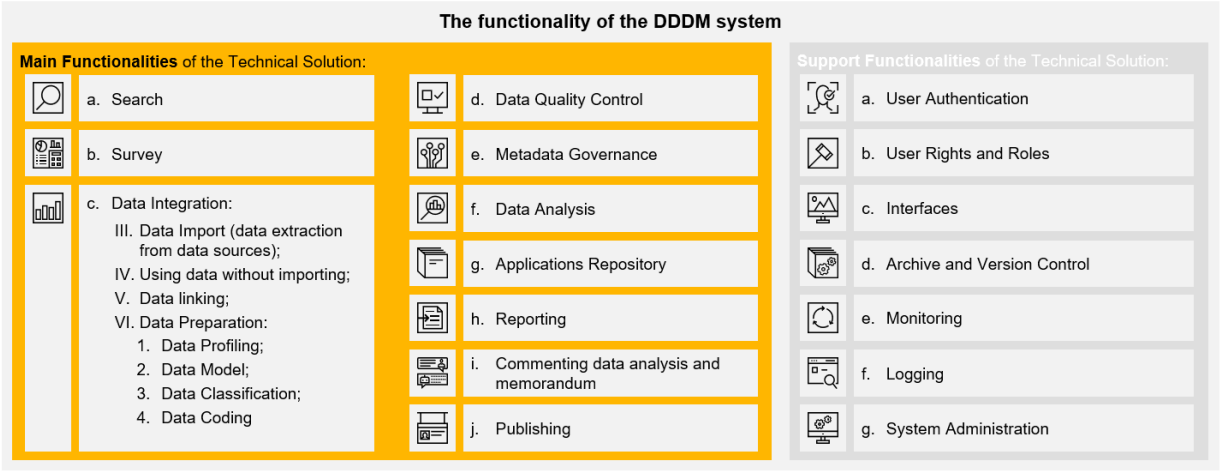


Figure 12. Government Memorandum 2.0 – DDDM System Architecture

As we see in the figure the DDDM system functionality is divided into two parts:

1) Main functionality.
2) Support functionality.

In this figure are presented main functionalities related directly to the data analysis. The DDDM system has two sides. The first is the composing of memorandum texts and the second is data analysis subsystem. The simplest way to do the data analysis is shown in the following figure.

[22] https://www.just.ee/oigusloome-arendamine/riigi-koosloome-keskkond

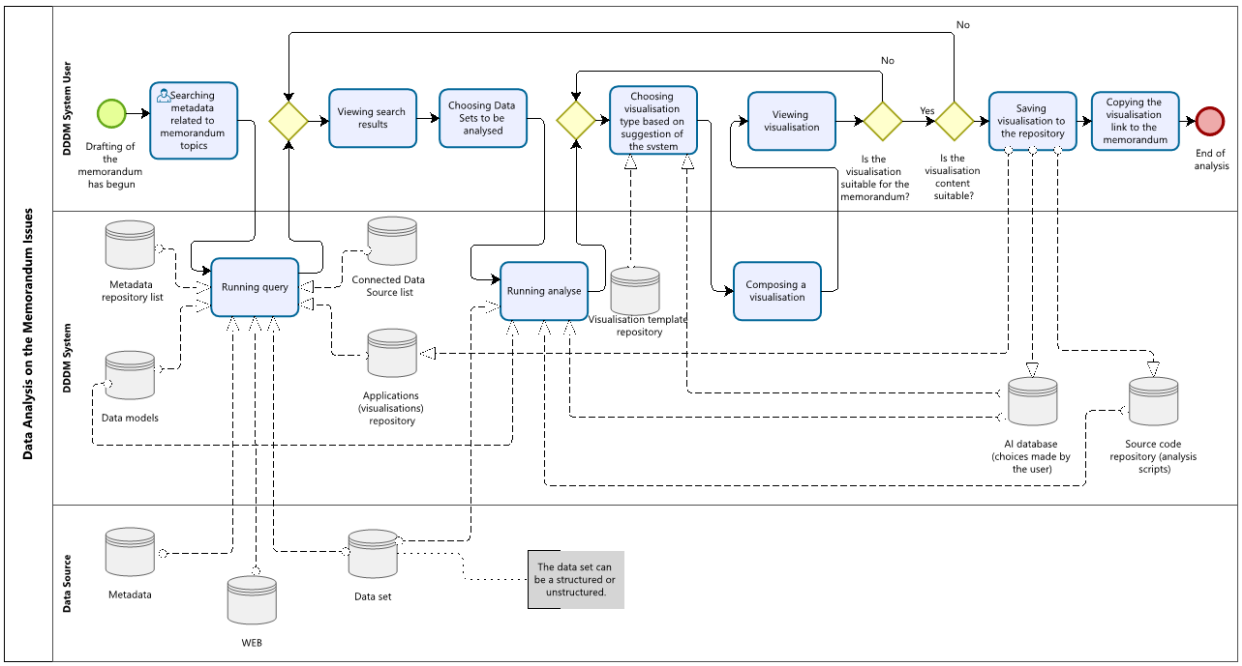


Figure 14. The Simplest Data Analysis Workflow in the DDDM system

There are more different workflows in the system. One of the most important on data analysis is data modelling that is closely related to the data integration. Data models are description of how the data source data can be linked, and what are main columns or parts of the data to be used. Also, the data quality checking and data profiling workflow must exist to guarantee the quality of the output from DDDM data analysis subsystem. The following chapters describe what are functionalities that need to support the DDDM system analysis process in the future.

2.3.1 Main Functionalities of the Technical Solution

2.3.1.1 Data categories and features

The DDDM system must support the data usage on different data categories.

Data can be divided into categories as follows. **Main categories** are:

1) Data.
2) Metadata.

The second level categories are:

1) Structured data.
2) Unstructured data.

Both the data and the metadata can be structured or unstructured.

The third level categories are:

1) Aggregated data.
2) Microdata.

Data features are:

1) The data is provided with metadata. Also, many data sources can provide partial metadata, or the metadata that is not up to date.
2) A data set can be utilized in the server of a data owner (e.g., on the SQL server), or the data must be downloaded and used on premises.
3) The data can be cleansed or not. The not cleansed data can be called raw data.
4) The data can be classified or not.
5) The data can be coded or not.

6) The data format can be machine-readable, or non-machine readable.
7) The data structure information can be machine-readable (e.g., database DDL script with descriptions of data types, lengths, mandatory attributes, and relationships) is provided, or non-machine readable (e.g., the data is in CSV file that has no structure definition.

2.3.1.2 Data protection and IT security

An essential question on the DDDM system is the data protection. Without data protection features the system cannot exist. There are different levels of data security measures:

1. Legal level.
2. Organisational level.
3. Technical level.

Data usage must be legalised. The DDDM system must have a law-based, or agreement-based permission to use the data. An option to resolve the legalisation of the data processing is to use the researcher's data processing and analysis environment of Statistics Estonia. Statistics Estonia has legal environments already configured. The second option is to design and put-up environment where the prevention infringement of individual rights is resolved with technical measures, e.g., data encryption or shared computing.

The DDDM system must have an IT-organisation having a capability to monitor data usage and system for authorizing the user.

There are also known technical measures on how to protect the data:

1. Data encryption.
2. Shared computing.
3. Data anonymization.

Several requirements from the point of view of IT security must be fulfilled:

1. Reliable user authentication solution must be implemented in the system.
2. User authorization must be systematic and auditable.
3. Actions in the system must be logged and logs must be analysable in real time.
4. Estonian Information Security Standard[3] must be followed by the user and owner of the DDDM system.
5. The DDDM system must have passed security testing, e.g., using OWASP Application Security Verification Standard[4].

2.3.1.3 Search

The data search functionality is the first thing for supporting the user of the DDDM efficiently.

[3] https://eits.ria.ee/et/avalehe-menueue/security-measures-v2021/

[4] https://owasp.org/www-project-application-security-verification-standard/

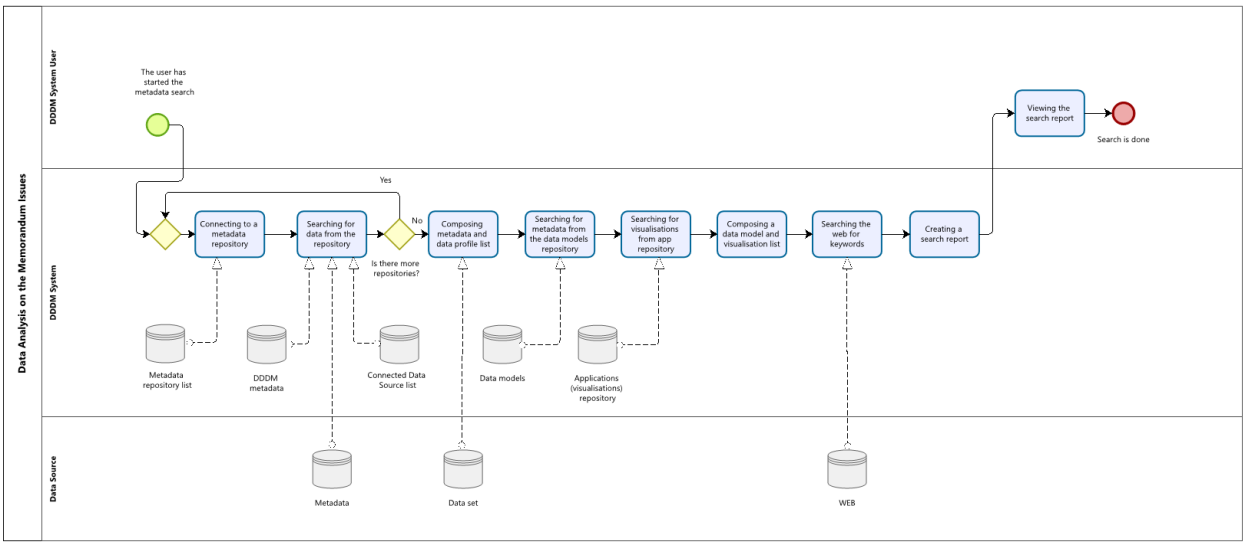


Figure 15. The Data Search Workflow in the DDDM system

The data can be found in numerous data sources. But finding the right one is in most cases impossible due to the lack of accurate meta data. If the data user is not familiar with the data content in a data source, results of a data analysis may lead to wrong decisions in the government. The failure on the decision will come quietly because of invisible misunderstanding of the data profile.

To avoid such problems, it is important to use measures in the search of data as follows:

1. Ensuring meta data (data descriptions) availability and quality.
2. Visualising data profile to explain data scope and content.
3. Displaying example visualisations on the data generated by the DDDM system.
4. Displaying previously performed analyses on the same data.
5. In the case of text data, make available context sensitive search and analysis of the data.
6. Doing checks on the data quality and visualising quality indicators.
7. Label the data with keywords in the data source, if possible, and use label-based search as an additional possibility of finding the data.
8. Making available data steward contacts for getting consultancy on the data content, semantics, and quality.

If the DDDM system can use measures mentioned above the results and outcomes on data driven decisions could be much better than without the DDDM system.

2.3.1.4 Survey

There is no digital data always available for an analysis. In such cases the data user must collect the data through a survey.

To arrange a survey, the system must have use cases as follows:

1. Creating a survey questionnaire.
2. Defining questions.
3. Defining answer variants (lists).
4. Defining and generating sample. Sample is a set of respondents.
5. Searching contact information of respondents, e.g., doing X-road query to Business Register to get contact details of respondents from a certain field of activity.
6. Running survey.

7. Survey result analysis and results visualisation.
8. Downloading the survey answers as a file.

Survey data can be used as an ordinary data table or file. The survey tool is a non-mandatory part of the DDDM system. As a survey tool can be used many software products, e.g., Microsoft Forms or Google Forms etc. It should be noted that common tools do not have the sample generation features and interfaces to Estonian data sets. Also there is no obtaining sample contact details for automatic sending questionnaire to concrete respondents by e-mail. Such features are important to get additional data as fast as possible, and with no additional cost to organize surveys.

2.3.1.5 Data Integration

The main issue in the efficient data consumption today is data availability. There are different data sets that are intended to internal use in different agencies and authorities in Estonia. Legal aspects of publishing the data for other purposes is not a topic of data integration but it must be considered before the data integration work starts. **The data integration part of the DDDM system must resolve technical issues of data integration.** Issues in the integration are mainly related to:

1. Binding (interfacing) data consumer and data owner systems for data exchange.
2. Different data formats and data element types that are not compatible with each other.
3. Different semantics of technically similar data.

Binding, compatibility, and semantics issues must be resolved in the integration process. The integration process can be automatic if the data has standardized API for data exchange, standardized format, and clearly defined machine-readable semantic description the DDDM system can understand without human intervention to the integration process.

The input data can be from a mixture of external or internal sources, and a variety of the collection instruments, including extracts of administrative and other non-statistical data sources. Administrative data or other non-statistical sources of data can substitute for all or some of the data directly collected from survey.

Data integration is a process where the data from different data sources we need for data analysis will be gathered and made ready for the data preparation and modelling. The data integration can be technically done in many ways:

1. Structured data.

- Retrieving micro data from a relational database.
- Retrieving micro data from a web service.
- Uploading micro data from a file (xlsx or csv) to DDDM system.
- Retrieving aggregated data from statistics databases (e.g., Statistics Estonia databases[5], Eurostat databases[6]).
- Retrieving micro data from researcher's environment in Statistics Estonia.
- Retrieving data from spatial databases (maps and spatial objects, e.g., Estonian Land Board databases[7], Inspire databases[8]).

[5] https://andmed.stat.ee/et/stat

[6] https://ec.europa.eu/eurostat/data/database

[7] https://geoportaal.maaamet.ee/

[8] https://geoportaal.ee/

- Using data in the original data source database.

2. Unstructured Data:

- Querying micro data from web sites.
- Retrieving micro data from text documents.
- Using data in the original data source.

The data integration may include:

1. Combining data from multiple sources, as part of the creation of integrated statistics.
2. Combining geospatial data and statistical data or non-statistical data.
3. Matching or record linkage routines, with the aim of linking micro or macro data from different sources.
4. Data fusion - integration followed by reduction or replacement.
5. Prioritising, when two or more sources contain data for the same variable, with potentially different values.
6. Linking micro data, or aggregated data with metadata.

It's important that the data integration method is dependent on legal status of the data user and the data source. Some users are authorized to use aggregated data only and the microdata usage is prohibited.

2.3.1.5.1 Data Import

Data import is a process the DDDM system extracts data from a data source to the analysis database.

To use the data from the data source, the following work must be performed during the development project:

1. **Determine whether the data is available**, or not. Availability can be created either through the publication of open data in the open data portal or through a corresponding web service managed by the data owner itself.
2. **Determine whether the data has descriptions** and whether the content of the real data set corresponds to the data descriptions.
3. **The description** of the data in the database **must be machine-readable**. If it is not machine-readable, it must be made machine-readable. Otherwise, it is not possible to use the database data for data search.
4. Check whether there is a legal basis for using data from the data source. If not, create a legal basis for the use of the data in cooperation with the administrator of the database.

To interface a database, the database must ensure the performance of the following use cases:

1. Publication of database data on the open data portal.

OR

1. Management of rights to read database data.
2. Issuing data from the database through the web service.
3. Automatic monitoring of the performance of the data issuing web service.
4. Real-time technical support for the availability of the web service to ensure availability.

2.3.1.5.2 Using data without importing

The data can also be used without importing. If the data owner can guarantee availability of the data source, it is possible to utilize the data without importing all the data into the DDDM system database. Then the visualisation will do real time queries to the data set. When visualizations are archived, microdata in use in the visualization must be copied into the visualization object (file).

Data usage without importing is reasonable in the case of large-scale (approx. more than 10 GB) data if the data owner's system has sufficient computing capacity. The computing capacity can be on the DDDM side where the network between the data owner's database and the DDDM system uses a high-speed connection. The data owner system must have a database server or similar system with ability to run queries the data consumer system sends. Data usage without import is impossible if the data location is only a data store. An application server is needed. It means also that the data owner system must have an API to receive tasks from the data consumer. The application server makes queries and calculations and sends back an answer for the data consumer question.

2.3.1.5.3 Data linking

Linking of the data is an essential part of multi-source data processing and analysis. **Linking is a data processing task where you need to find keys connecting same objects or subjects in different data sources.** For example, the personal ID code is a linking key for persons in different data sources. The DDDM system can link the data if keys are defined and known in data sources. The key can differ from the technical database key. Content of a key can be more than one column in the database.

The data linking work can be done by the third party, e.g., Statistics Estonia or a service provider who has people, skills, software, hardware, and legal rights for performing the task.

In the long term, **a public classifier should be considered**, which would systematize the semantics of linking data elements. It means that there should be data type definitions for link elements. If the data element has the same specific type as another element in another data set, elements can be used as linking elements. Every data element must have the data type defined in the metadata. If there is no such type definition in the data set, the DDDM system cannot link the data automatically.

2.3.1.6 Data Preparation

Data preparation is a process making data useful for analysis and visualisation.

2.3.1.6.1 Data Profiling

Data profiling is a process where a data user discovers data structure, content and relationships between data sets or data tables. It is necessary for creation of a data model that can be used as a basis of data visualisation.

Data profiling must contain use cases as follows:

1. Data structure discovery from a data file or database.
2. Statistical metadata generation based on data structure and micro data content. Metadata may have different views on the data, e.g., value scope, value frequency, data type, data format, and other characteristics on data columns for better understanding what the data contains.
3. Relationships discovery between data sets and tables using data content.
4. Storing data profile.
5. Displaying data profile.
6. Linking data profile with the data.

Data profiler is a tool for data analysts and modelers, who needs automated metadata generation possibilities for creating better data models more efficiently than today.

Data profiling can be an automatic process also. The DDDM system must contain automatic data profiling features.

2.3.1.6.2 Data Model

Data modelling is needed for compiling useful views on the data to be analysed making data consumer work more efficient. In the data modelling process the DDDM system or data analyst should act on use cases as follows:

1. Choosing data tables or data sets to put into the data model.
2. Autodetecting relationships between data tables in the data model based on the data in tables.

3. Setting relationships between data tables or data sets manually.
4. Setting relationships between data and dimensions.
5. Uploading and linking auxiliary data into the system and linking them with the data.
6. Setting filters on the data. Filtering out incorrect or unnecessary data from the data model.
7. Setting data displaying formats, e.g., date format, currency format or floating-point numbers format.
8. Describing data model metadata (table, column, and relationships descriptions).
9. Displaying data model as an ERD (entity relationship diagram).
10. Generating ERD layout automatically.
11. Tuning the ERD layout manually. It's needed if the number of data tables is bigger than 4-5 tables.
12. Exporting data model metadata as a document.
13. Publishing the data model on the data model repository of DDDM system.

Result of the data modelling is a well-designed and described relational data view for data consumer (or data analyst). The modelled view is a starting point of the data visualisation. **The model must be done always if we have more than one data table to be analysed together**.

For automating the compilation of a data model there must be automated activities as follows:

1. Detecting primary keys in the data set.
2. Detecting possible link columns.
3. Detecting relationships between data tables.
4. Relationship multiplicity detection.
5. Provisioning of model variants to the user.

2.3.1.6.3 Data Classification

In most cases the data must be classified before the system is able to analyse the data. It means that the system should calculate a class code for each data record in the data set based on other code columns or columns containing unstructured data, e.g., text data or picture. Coding routines may assign numeric codes to text values according to a pre-determined statistical classification to facilitate data capture and processing.

If we have a classification dictionary where the system can detect compliance between text and code, the DDDM system can classify a data record based on text data. Classifying pictures is more complicated. To classify pictures without using machine learning before classifying a data set, is impossible in most cases.

2.3.1.6.4 Data Coding

The data from different data sources can be classified with different classifications having the same or similar meaning of the classification. In such cases the DDDM system must do coding based on coding dictionary. Some data columns may have coded value categories on the administrative source of data.

Coding and classifications are important routines to make data set ready for data analysis.

2.3.1.7 Data Quality Control

In addition, data quality management must be checked automatically, the use cases are as follows:

1. Data quality control.
2. Defining data checks.

3. Running data checks.
4. Preparation of the data quality report.
5. Improving data quality according to identified problems.
6. Automatic correction.
7. Manual correction in the source database by the data owner.
8. Issuing a data quality assessment to the data consumer through the online service.
9. Reception service for descriptions of data errors. Through this, the owner of the data can be informed about the problems found in the database.

2.3.1.8 Metadata Governance

The system must use Estonian metadata infrastructure RIHA(KE), whether it is relevant for the database. It is a metadata management system for every database owned by the Estonian public sector agency or authority. Not all databases and data elements are described in the RIHA. RIHA and/or RIHAKE must be linked with DDDM system. The user of the DDDM system must have possibility and capability to use metadata to avoid mistakes on the interpretation of data analyses. It's very complicated to choose the data for an analysis if there are no data descriptions on the data content, e.g., what are data objects in the database or how many variables the data contains, how the data is classified, what are classifier elements, and what are elements meanings etc.

Use cases on the usage of metadata are:

1. Querying metadata.
2. Reading the metadata.
3. Searching keywords from the metadata to find relevant data columns (or variables) in the database.

The DDDM analysis applications repository must have its own metadata governance system. It should also be mentioned that the applications metadata is not the same as the metadata on microdata. The memorandum user must have use cases as follows:

1. Searching an analysis application based on the application description.
2. Commenting and describing analysis applications for further efficient use of the application.

It should also be considered that metadata created in the DDDM system should be transferred back to the RIHA. RIHA is the primary data source of metadata, especially metadata on public registers and data sets belonging to the state information system. If RIHA could be a metadata management solution and an online metadata repository for the DDDM, it would be the best solution from the state's point of view.

2.3.1.9 Data Analysis

The analysis process should be as much automated as possible. Data source can have several dimensions and filtering options. Also, many analysis scripts can be useful. Choosing right ones may be a sophisticated process. If the user can select useful filters, dimensions and scripts, the result of the analysis can be more effective as without user intervention. It's important that first users must make right decisions to teach the AI system in the DDDM.

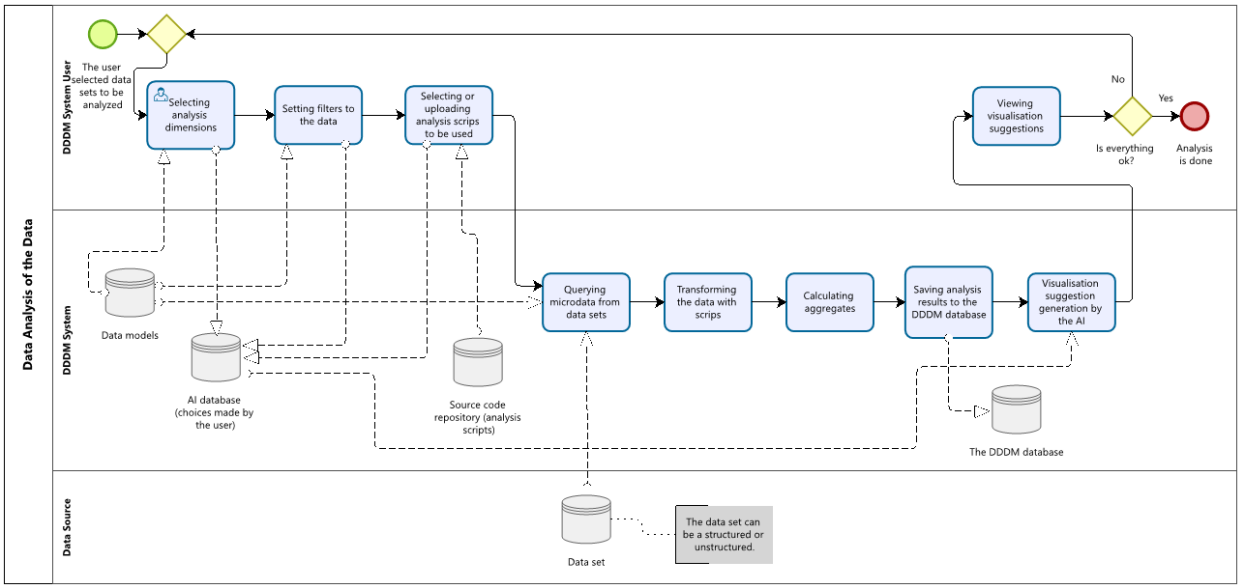


Figure 16. The Detailed Data Analysis Workflow in the DDDM system

In the data analysis subsystem must be implemented use cases as follows:

1. Searching, sorting, and filtering the data in the analysis database (the prepared data model).
2. Choosing prepared data to be analysed.
3. Choosing dimensions for grouping the data in visualisations.
4. Choosing units to be used in the visualisation.
5. Making transformation on data columns, if needed, e.g., setting up generalised dimension having values like “more than 1000” and “less than 1000”.
6. Calculating aggregates.
7. Displaying micro data.
8. Displaying aggregated data.
9. Choosing visualisation types to be used in visualisations.
10. Visualising analysed data:

- Geospatial data visualisation.
- Numerical data visualisation.
- Text data analysis visualisation.

11. WHAT-IF functionality with changeable input parameters.
12. WHAT-IF with prediction features.
13. Describing and commenting visualisations – compile data stories.
14. Publishing visualisations and explanations of visualisations.
15. Gathering feedback on the analysed data and visualisations.

2.3.1.9.1 Geospatial data visualisation

Geospatial data sources are developed in Estonia. There can be found a lot of examples on visualisations and analyses published in the map.

For handling and analysing geospatial data, a special GIS software is needed. Usage of geospatial data can be divided into two general types:

- Space data analysis.
- Utilizing base map (see Figure) and additional map layers for enriching visualisation (see Figure and Figure).

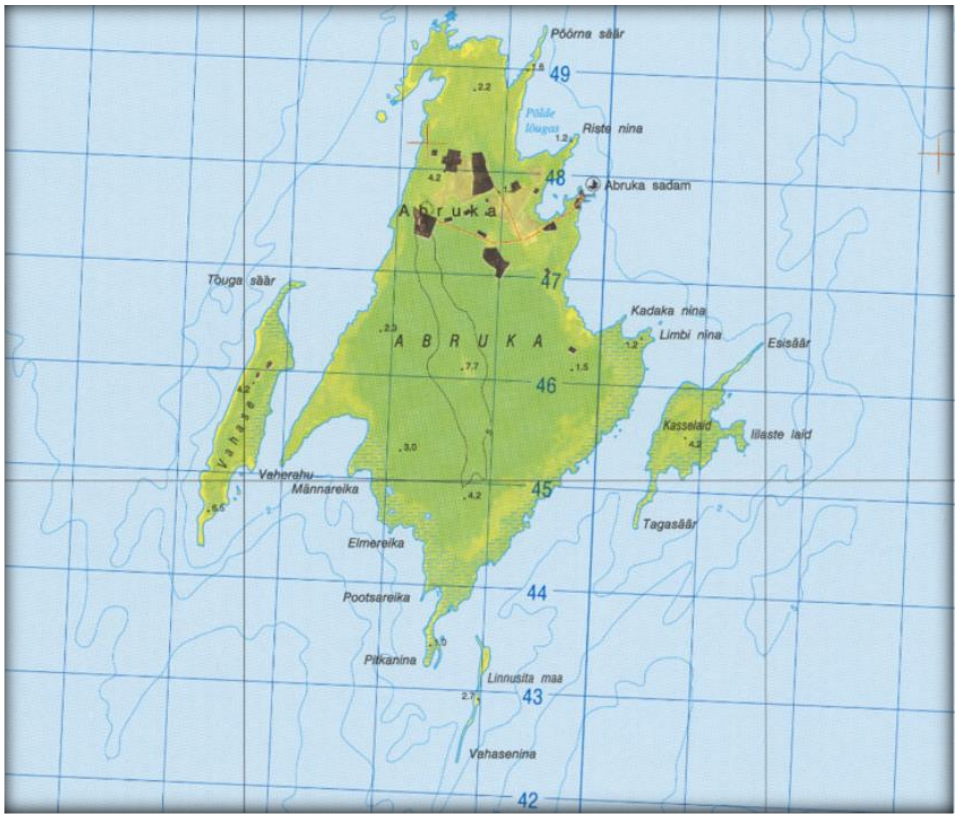


Figure 17. An example of a base map

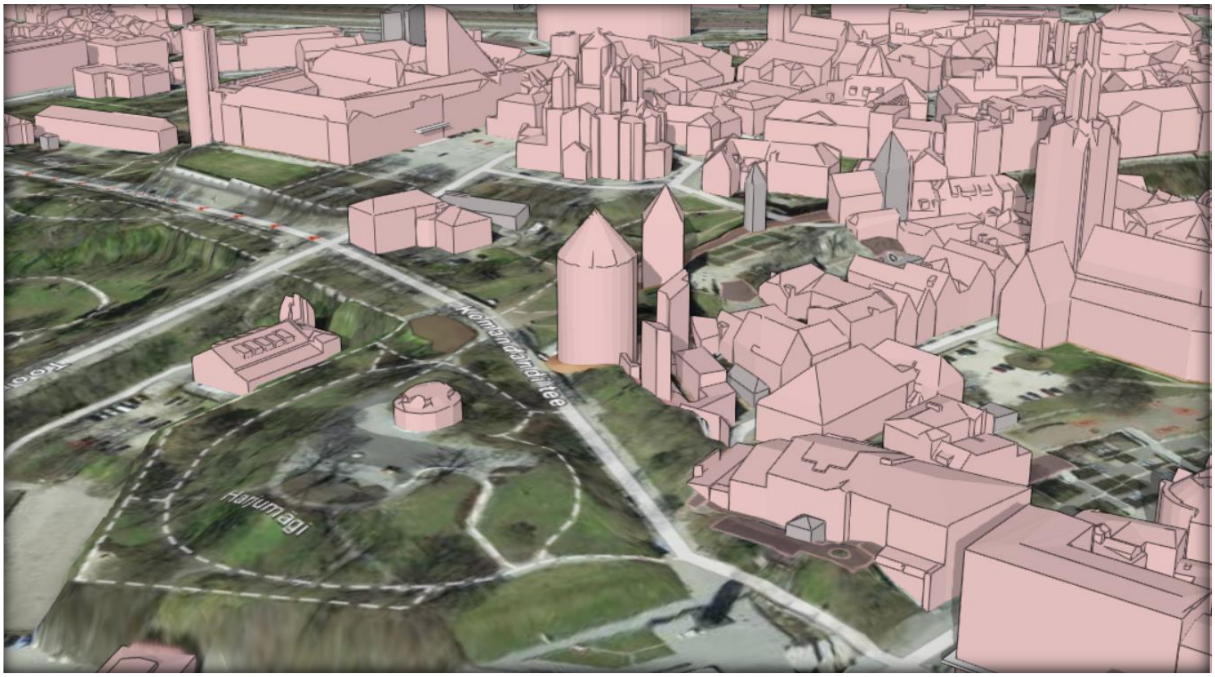


Figure 18. 3D map

Figure 19. Transparent map layers on the base map, ARIB agricultural parcel map layer

All these map types shown in the figures above can be used in the data visualisations on the map. For example, agricultural production in numbers can be visualised with the ARIB field map layer. The visualisation is dynamic, and the user of the visualisation can see general picture and detailed picture in the field level in numbers and as a map. If we add an additional layer, e.g., fertilizer usage, or soil composition, a very informative picture emerges.

Functionalities the DDDM system should have on the analysis and visualisation, are:

1. Analysing spatial data in shape formats in a GIS server, e.g., PostGIS9, ArcGIS10, or similar.
2. Placing map layer in WMS11 format to the visualisation.
3. Placing map layer in WFS12 format to the visualisation.
4. Linking a map layer and a numerical data set, aggregated or micro data.

Spatial data is any type of data that directly or indirectly references a specific geographical area or location. It's important to use spatial data in standardised format, e.g., coordinates in WGS84, or in L-EST97, or a location name must be written correctly to use object automated linking with the map data, or other spatial dimension like roads, or other space shape. The different map layers can be used in the same visualisation, whether layers are in the same coordinate system, or the DDDM system will have a capability to convert different coordinate systems.

The satellite data can also be used for spatial analysis, e.g., to assess the environmental condition in water bodies. The EstHUB[13] is a data source of the satellite data on the Baltic Sea region and Estonian territory.

The Estonian Land Board has a large set[14] of WMS/WFS data services for the user.

The DDDM system must provide different data services (map layers) the Estonian Land Board has prepared for the user automatically dependent on the scale of a numerical data set linked to the geo information. The data set can be linked to the map if the data set contains geographical dimension in the data in the form of geographical coordinates or addresses.

2.3.1.9.2 Numerical data visualisation

Data analysis in the DDDM system must be automated as much as possible, especially on the analysis of numerical data having traditional work process.

The system must have several autodetection features for the automated analysis as follows:

1. Detecting measures in a data set.
2. Detecting dimensions in the data set:
    - Detecting time dimensions.
    - Detecting geographical dimensions.
    - Detecting other dimensions.

---

[9] https://postgis.net/

[10] https://en.wikipedia.org/wiki/ArcGIS

[11] https://www.ogc.org/standards/wms

[12] https://www.ogc.org/standards/wfs

[13] https://geoportaal.maaamet.ee/est/Ruumiandmed/Riiklik-satelliidiandmete-keskus-ESTHub-p443.html

[14] https://geoportaal.maaamet.ee/est/Teenused/WMSWFS-teenused-p65.html

3. Unit columns detection using unit abbreviations classifiers.
4. Detection of what dimensions can be used for data analysis, which ones are most represented in the data set records.
5. How looks like distribution of dimension values in the data set.
6. What are values of the dimension with the highest weight.
7. An index on interestingness: what are more interesting distributions in the data set, an interestingness index must be calculated by the DDDM system.
8. Rating of the data quality in column.

More sophisticated and precise autodetection features can be:

1. Calculating dependency functions based on dimension value distribution data.
2. Running third party analysis (e.g., R scripts).

Recommended set of visualization types for the numerical data analysis in the DDDM system may be:

1. Data table.
2. Matrix.
3. Numerical aggregate's chart.
4. Value frequencies.
5. Line chart with time series of aggregates.
6. Bar chart.
7. Stack chart.
8. Multidimensional bar chart.
9. Sector diagram.
10. KPI chart.
11. Slicer for filtering the data.
12. Map visualisation supporting Estonian base chart.
13. Map visualisation with WFS support for using different map layer, e.g., administrative units,
14. Heatmap visualisation for showing numerical aggregates graphically on the map.

2.3.1.9.3 Text data analysis visualisation

There are several types of text analytics:

1. Word statistics and search.
2. Context sensitive analysis and search.
3. Guided machine learning and analysis on text.
4. Text meaning analysis.
5. Extracting numerical data from text.
6. Dialogue bot:
    - Based on standard questions and answers.
    - Based on guided machine learning questions and answer system.

The representation of the text analysis result can be a frequency table about words, or sentences in the text, or a list of sentences for which information was found, or translation of the text.

2.3.1.9.4 WHAT-IF functionality with changeable input

The DDDM system must have WHAT-IF features to help the government on checking different scenarios a decision may cause.

There can be several ways to run a WHAT-IF analysis. The easiest way to get different results for the same analysis is to change input parameter for an analysis. The DDDM system should have dynamic visualisations with sliders that change input parameters of an analyse. It is the way to give the best user experience.

For example, if an analysis result must show value added tax rate impact on the sector indicators the tax rate should be a slider field on the visualisation. The user could then easily change the value of the field and the content of the visualization would change. If we combine different input parameters in such a way the user can get a powerful tool with very simple user interface. The assumption is that the previous work such as building the data model and creating the visualization has been done before the end use can play with the tool.

2.3.1.9.5 WHAT-IF with prediction features

WHAT-IF functionality can be implemented with prediction features in the same visualisations. Then the user can get a time series for the decision impact.

2.3.1.10 Applications Repository

If an DDDM system user creates visualisations, data sets, or data models that may be useful for other users later, the system should have a repository for storing created things. We can see such repositories in different platforms, e.g., R script repositories[15], Power BI[16], or Tableau[17] example galleries. The DDDM AI features can use script usage statistics to suggest scripts in the concrete situation.

If the DDDM system uses a data model and visualisations repository, the visualisation, or data model, or both together must form an independent object containing data and the model. It means that the data and the data model are in one file and the file contains all necessary component to reproduce a visualisation, or data model.

The repository should have use cases as follows:

1. Repository user and storage management.
2. Uploading data models to the repository.
3. Uploading visualisations to the repository.
4. Checking visualisations and data models source code quality.
5. Searching repository objects in the repository.
6. Downloading repository object by user.

An important topic for the data analysis is a code list and classifications data repository. An example of such repository can be found at Statistics Estonia: https://klassifikaatorid.stat.ee/.

[15] https://cran.r-project.org/web/packages/available_packages_by_date.html

[16] https://community.powerbi.com/t5/Data-Stories-Gallery/bd-p/DataStoriesGallery

[17] https://www.tableau.com/data-insights/dashboard-showcase

2.3.1.11 Reporting

The user may want to use the same report (visualisation) more than once for analysing different time periods in the past and in the future. Standard reports are parts of the DDDM system repository. Reports are data visualisations. It's important to have a possibility to update the data from data source in the report automatically in the DDDM system.

2.3.1.12 Giving feedback on artifacts

It's very important to get feedback from different specialist and experts on data analysis and on visualisations. There must be a feedback section in every visualisation and memorandum in the DDDM system.

2.3.1.13 Publishing

DDDM must have a data analysis results (visualisations) publishing environments, one for authorized user and the second for other public users. It is important to publish visualisation for the public user to get feedback before the government takes the analysis results into account. Essential part of the publishing environment is the feedback subsystem.

The publishing environment must contain functionality as follows:

1. Uploading visualisations containing micro and aggregated data.
2. Displaying dynamic visualisations.
3. Running visualisation as a separate application for integrating them to any web application, e.g., to the government web site, or state internal systems.
4. Downloading visualisations for further development.
5. Deleting visualisations from publishing environment.
6. Authorized micro data download from visualisation.

Integrating visualisation to the memorandum is also a publishing feature. The memorandum and a visualisation must form an independent file that can be a static application or a source file for further development and upgrading tasks.

2.3.2 Support Functionalities of the Technical Solution

Estonia has a well-developed IT infrastructure with interoperability between different systems. It is important to use common systems, e.g., Estonian person's authentication system in the DDDM. Because of that there is a list of supporting functionalities described in this chapter. Without these functionalities the system cannot work, or the data security measures are too week, and the system will be an illegal platform for the user.

2.3.2.1 User Authentication

User authentication must be done with TARA. TARA is a standard user authentication solution in Estonia.[18]

2.3.2.2 User Rights and Roles

A user management should be implemented on the central TARA authentication system. TARA contains identity information for every citizen of Estoia. An independent user management system is not needed in the system. TARA manages user information automatically and gives the user authentication functionality as required.

User rights and roles management system is not a part of the TARA. User authorization must be done in the DDDM (or local LDAP, e.g., Active Directory). The system administrator must assign necessary DDDM roles to the user in the user authorization solution. It's impossible to use the DDDM without an assigned role. A role is an elementary part of user rights in the system. It can be a right to use a data source data, to use a specific system functionality, or add metadata in the system.

2.3.2.3 Interfaces

The Estonian interfacing platform is X-road. This is a secure network of mainly public sector information systems. Many databases are interfaced with X-road via X-road web services. More frequently used data is published on the X-road but it's a small part of the data needed in data analysis for memorandums. It means that there must be more possibilities to transport the data. The DDDM system must have interfacing functionality to any needed data source directly or via standardised format at least as follows:

1. SOAP web services[19].
2. REST web services[20].
3. ODBC[21].
4. JDBC[22].
5. CSV files[23].
6. XLSX files.
7. XML files[24].
8. JSON files[25].
9. PostgreSQL database.

---

[18] https://e-gov.github.io/TARA-Doku/

[19] https://en.wikipedia.org/wiki/SOAP

[20] https://en.wikipedia.org/wiki/Representational_state_transfer

[21] https://en.wikipedia.org/wiki/Open_Database_Connectivity

[22] https://en.wikipedia.org/wiki/Java_Database_Connectivity

[23] https://en.wikipedia.org/wiki/Comma-separated_values

[24] https://www.w3schools.com/xml/xml_whatis.asp

[25] https://www.json.org/json-en.html

10. Oracle database.
11. MS SQL Server database.
12. MS Access files.
13. OData web services[26].
14. Denodo[27].
15. Docx files.
16. Pdf files.

2.3.2.4 Archive and Version Control

The DDDM system must have an archiving system. An archiving object are files containing data models, visualisations, and documents (memorandums) with data visualisations. The system administrator must have possibility to move files to archive and restore from the archive. The archive can be file-based data store in the cloud or on premises.

Version control of documents and visualisations can be organised with the SharePoint or similar product with support of MS Office 365 products that are widely used in public sector.

2.3.2.5 Monitoring

All high availability DDDM system parts must be put under monitoring with ZABBIX[28] software or similar. There must be an organisation unit that is responsible for configuring the monitoring system and unit's specialist must resolve any incident occur to loss of availability.

2.3.2.6 Logging

The system must have event logging system. Every event related to sensitive data that the user initiates must be logged into log files that are separate from the system and are independent, and in common format, e.g., in text format. The system must log event data elements as follows:

1. Which user acted.
2. What the user did.
3. When the user acted.
4. Which functionality was triggered by the user.
5. Where the use came from.
6. What files or data sets the user used.

The system log must be analysable in any text analysis tool that can read structured text data. Text log must be human-readable. It's essential that only authorised persons can access the log. Any change in the log file must be auditable.

---

[26] https://www.odata.org/

[27] https://www.denodo.com/en/denodo-platform/overview

[28] https://www.zabbix.com/

2.3.2.7 System Administration

The DDDM system administration must cover different tasks as follows:

1. Deploying system components.
2. Authorizing users.
3. Configuring system monitoring.
4. Resolving system related incidents.
5. Backup system data and restore it within acceptable time frame.
6. Archiving visualisations, data models, and documents.
7. Configuring interfaces with data sources.
8. Disaster recovery.

There is need to establish a data steward role for helping administrator and users to manage with the data integration from different sources. Data should also be organized and managed after integration to guarantee data availability to the user.

2.4 DDDM System Implementation Alternatives

2.4.1 Introduction to the Scenarios

In principle, **three different scenarios** can be used to achieve the objectives of the **DDDM system**, which are described below. The tables show, by main component, in which environment a specific component is implemented.

Statistics Estonia is a state institution whose main task is to provide reliable and objective information about Estonia by collecting data from companies and individuals and capturing them from various databases. The National Statistics Act stipulates that Statistics Estonia, in addition to the obligation to produce national statistics, also provides a data sharing service.

Today, **Statistics Estonia** does not have an information system for the provision of data sharing services, which would enable it to fulfil the task set by law and meet all user expectations. Statistics Estonia has a service for researchers, but today's service environment for researchers is outdated in terms of technology and user-friendliness, and therefore a new development project "**Data reuse environment for research**" (**DRE**) has been initiated. The goals of the new development project largely overlap with the goals of DDDM in terms of data retrieval, processing, analysis, and visualization of analysis results.

The State Chancellery (Government Office) is the main government institution in shaping and implementing policies, supporting the Government of the Republic and the Prime Minister. Among other things, the Government Office plays a leading role in the implementation and dissemination of the data-based decision-making process in ministries and government institutions, from which the DDDM project has also grown.

KOOS, or the state co-creation environment, is an IT solution created in cooperation between the Ministry of Justice, the Government Office, and the Chancellery of Parliament of Estonia, which should in the future gather information related to the law-making process affecting Estonian life in one place and cover the entire creative arc of legislation until it is published in the Riigi Teataja. KOOS has a lot in common in the context of drafting memorandums.

2.4.2 Scenario 1 – Distributed System

The marks in the table (Table 5) mean that the developer and administrator of the given component is the corresponding institution, it has the corresponding task, competence, support organization, budget, etc.

Scenario 1. Distributed system.

Table 5. Scenario 1 – Distributed System

Component	**"DRE System"** (Statistics Estonia)	**"DDDM System"** (Government Office)	**"KOOS System"** (Ministry of Justice and Government Office)
Preparation of memorandum			√
Data Search		√	
Data processing	√		
Data analysis	√		
Visualisation of data analysis results		√	

Pros:

- Statistics Estonia has long-term experience in data processing and analysis.
- Essential elements of law are already covered.
- Data search in the DDDM system, as the search functionality is broader.

- Can reuse KOOS developments.

Cons:

- Synchronization of development between Statistics Estonia, the DDDM system and KOOS can be difficult (e.g., the case if no budget is allocated to one party).
- Building a search system and visualization of analysis results into DDDM is a very complex task.

2.4.3 Scenario 2 – Government Office System

In scenario 2 the whole system is governed and managed by the Government Office as described in Table 6.

Table 6. Scenario 2 – Government Office System

Component	**"DRE System"** (Statistics Estonia)	**"DDDM System"** (Government Office)	**"KOOS System"** (Ministry of Justice and Government Office)
Preparation of memorandum		√	
Data Search		√	
Data processing		√	
Data analysis		√	
Visualisation of data analysis results		√	

Pros:

- Development and administration are in the hands of one authority, better coordinated and closer to the Government.
- It is possible to create a system based on completely new technologies without being stuck in old legacy systems.

Cons:

- The existing experience and technological base of other institutions cannot be fully applied.
- If, without changing the data processing rules, it is possible to process only public data, which, due to their content or generalization, may not enable the achievement of the set goals.
- Duplication of functionalities.
- The Government Office must create the appropriate capacity (personnel).

2.4.4 Scenario 3 – Statistics Estonia System

In scenario 3 the system is mostly based on Statistics Estonia as described in Table 7.

Table 7. Scenario 3 – Statistics Estonia System

Component	**"DRE System"** (Statistics Estonia)	**"DDDM System"** (Government Office)	**"KOOS System"** (Ministry of Justice and Government Office)
Preparation of memorandum			√
Data Search	√		
Data processing	√		
Data analysis	√		
Visualisation of data analysis results	√		

Pros:

- The existing developments and technologies of Statistics Estonia are fully utilized.
- The fastest "start" to achieve DDDM goals.
- KOOS developments are used (the functionality planned by KOOS largely overlaps with the needs of DDDM).
- The legal bases of data processing allow the use of non-public data at the disposal of Statistics Estonia in the analysis phase.

Cons:

- It is used only with the data available in the Statistical Office and the combined use of other data requires the development of the Statistical Office's system. The need to find a compromise between the interests of the interested party and the owner of the environment can harm the flexibility of the system and the achievement of goals.

The inclusion of other data outside the national statistics program can be partly difficult (requires legal regulation and additional developments, including funding).

3. Recommendations

There will be several innovations in the development of the DDDM system. **The biggest leap** if we look at the current situation in Estonia **is the mixing of structured and unstructured data analysis in the same system**. There are no such solutions today in use where the user can analyse microdata, aggregated data, and so-called unstructured data together. **We recommend dividing the development into parts**. Then the user may get quicker a useful tool. It's also important that the development needs feedback from the user makes development plans more precise.

The AI part of the system described above is not known solution today. There may be pitfalls we do not recognize. It's needed to test the idea of using the AI for composing data models, metadata, and recommendations on visualisation types. There is opinion today that only the human can design a data model. There must be taken a deeper analysis on that topic and **test the use of the solution before the solution comes public**.

The recommendation is also to **specify data sources before the DDDM system development**. All data sources have special restrictions, and legal status.

4. Appendices

4.1 List of conducted Interviews

Table 8. List of Interviews conducted

Date	Organisation	Participants, organisation/role
26.05.2022	Several	**Agu Leinfeld**, Datel **Andres Lille**, Tieto **Ave Habakuk**, Innovation Team of the Government Office **Daniel Kotsjuba**, Innovation Team of the Government Office **Indrek Seppo**, Analyst of the e-Residency in Estonia **Mati Mõtte**, Tartu University **Silver Traat**, Texta **Veiko Berendsen**, Statistics Estonia **Dmitri Burnašev**, Deputy Strategy Director of the Government Office **Erik Ernits**, Head of Data of the Government Office
31.05.2022	Government Office and Ministries	**Henry Kattago,** Strategy Director at The Government Office **Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data of the Government Office **Ulla Saar**, Deputy Secretary General, Ministry of Social Affairs **Hanna Vseviov**, Deputy Secretary General, Ministry of Social Affairs **Luukas Kristjan Ilves**, Deputy Secretary General, Ministry of Economics and Communications **Margit Martinson**, Deputy Secretary General, Ministry of Environment **Marko Gorban**, Deputy Secretary General, Ministry of Rural Affairs **Markus Kärner**, Deputy Secretary General, Ministry of Justice **Tarvi Pürn**, Deputy Secretary General, Ministry of Culture **Merilin Piipuu**, Deputy Secretary General, Ministry of Culture **Pärt-Eo Rannap**, Deputy Secretary General, Ministry of Education and Research **Liina Põld**, Deputy Secretary General, Ministry of Education and Research **Raoul Lättemäe**, Deputy Secretary General, Ministry of Finance **Kaur Kajak**, Deputy Secretary General, Ministry of Finance **Evelyn Liivamägi**, Deputy Secretary General, Ministry of Finance **Margit Gross**, Deputy Secretary General, Ministry of Defence **Piret Lilleväli**, Deputy Secretary General, Ministry of the Interior **Olavi Seisonen**, Deputy Secretary General, Ministry of

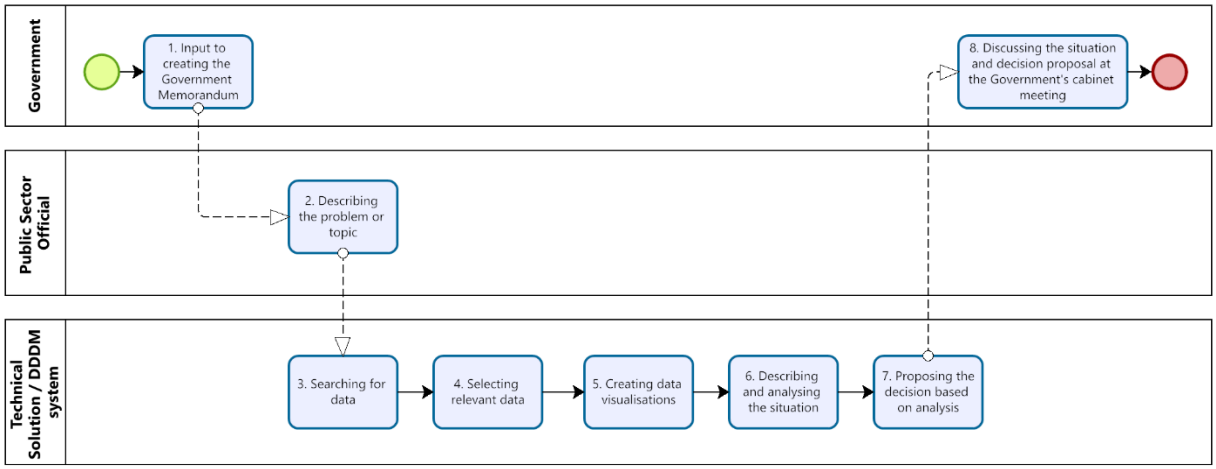
Date	Organisation	Participants, organisation/role
		Foreign Affairs
02.06.2022	Government Office	**Taimar Peterkop**, State Secretary **Henry Kattago,** Strategy Director **Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data + **Management Board of the Government Office**
06.06.2022	Government Office and Ministries	**Taimar Peterkop**, State Secretary **Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data of the Government Office **Maarjo Mändmaa**, Secretary General, Ministry of Social Affairs **Meelis Münt**, Secretary General, Ministry of Environment **Marko Gorban**, Acting Secretary General, Ministry of Rural Affairs **Tõnis Saar**, Secretary General, Ministry of Justice **Tarvi Sits**, Secretary General, Ministry of Culture **Kristi Vinter-Nemvalts**, Secretary General, Ministry of Education and Research **Merike Saks**, Secretary General, Ministry of Finance **Kusti Salm**, Secretary General, Ministry of Defence **Tarmo Miilits**, Secretary General, Ministry of the Interior **Jonatan Vseviov**, Secretary General, Ministry of Foreign Affairs
07.06.2022	Members of the Steering Committee	**Taimar Peterkop**, State Secretary **Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data **Urmet Lee**, Director General of Statistics Estonia **Pille Lehis**, Director General of Data Protection Inspectorate **Ott Velsberg**, Chief Data Officer of Ministry of Economics and Communications
08.06.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data
09.06.2022	OECD	**Jacob Arturo Rivera Perez,** Lead, Data-driven public sector & South-East Asia, Digital Government and Data Unit
09.06.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data
16.06.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data
20.06.2022	Statistics Estonia	**Veiko Berendsen**, Data Governance Expert
21.06.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data
27.06.2022	Government Office	**Erik Ernits**, Head of Data **Ivar Hendla**, Strategy Adviser
28.06.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data
29.06.2022	Texta	**Silver Traat**, AI-powered Text Analytics Expert **Dmitri Burnašev**, Deputy Strategy Director of

Date	Organisation	Participants, organisation/role
		Government Office **Erik Ernits**, Head of Data of Government Office
29.06.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data
01.07.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director **Erik Ernits**, Head of Data **Ivar Hendla**, Strategy Adviser
06.07.2022	Government Office	**Erik Ernits**, Head of Data
12.07.2022	Government Office	**Erik Ernits**, Head of Data
15.07.2022	Government Office	**Erik Ernits**, Head of Data
20.07.2022	AI Watch of European Commission	**Colin Pascal van Noordt**, AI External Expert **Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office **Ivar Hendla**, Strategy Adviser of Government Office
20.07.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office **Ivar Hendla**, Strategy Adviser of Government Office
26.07.2022	Cybernetica	**Baldur Kubo**, High Security and Privacy Requirements Expert **Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office
04.08.2022	Cybernetica	**Baldur Kubo**, High Security and Privacy Requirements Expert **Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office
09.08.2022	Government Office	**Erik Ernits**, Head of Data
15.08.2022	Government Office	**Erik Ernits**, Head of Data
16.08.2022	Government Office	**Erik Ernits**, Head of Data
23.08.2022	Government Office	**Erik Ernits**, Head of Data
26.08.2022	Ministry of Finance, Finland	**Aleksi Kopponen**, Special Advisor, Public Sector ICT **Erik Ernits**, Head of Data of Government Office **Ivar Hendla**, Strategy Adviser of Government Office
30.08.2022	Government Office	**Erik Ernits**, Head of Data
31.08.2022	Government Office	**Erik Ernits**, Head of Data
01.09.2022	Government Office	**Erik Ernits**, Head of Data **Dmitri Burnašev**, Deputy Strategy Director of Government Office
01.09.2022	Several	**Erik Ernits**, Head of Data, Government Office **Dmitri Burnašev**, Deputy Strategy Director of Government Office **Kaja Sõstra**, Statistics Estonia **Veiko Berendsen**, Statistics Estonia **Urmo Parm**, Data Protection Inspectorate

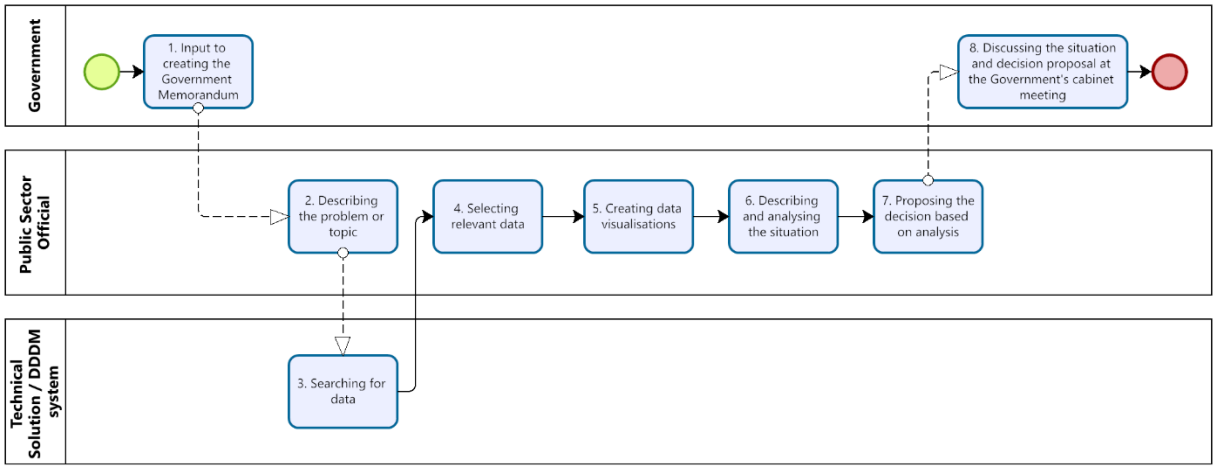
Date	Organisation	Participants, organisation/role
07.09.2022	Prime Minister's Office of Finland, Ministry of Finance, VATT Institute for Economic Research	**Seppo Määttä**, Director General, Prime Minister's Office **Hanna Kivistö**, Chief Specialist, Prime Minister's Office **Olli Kärkkäinen**, Chief Specialist, Ministry of Finance **Mikael Collan**, Director General, VATT Institute for Economic Research **Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office **Ivar Hendla**, Strategy Adviser of Government Office
13.09.2022	Government Office	**Erik Ernits**, Head of Data
14.09.2022	Government Office	**Erik Ernits**, Head of Data
20.09.2022	Government Office	**Erik Ernits**, Head of Data
21.09.2022	Government Office	**Erik Ernits**, Head of Data
26.09.2022	Government Office	**Erik Ernits**, Head of Data
27.09.2022	Government Office	**Erik Ernits**, Head of Data
27.09.2022	Government Department of Public Expenditure and Reform, Ireland	**Barry Lowry**, Chief Information Officer to the Government, Chairman of the Data Governance Board **Bernadette Dempsey**, Head of the Data Governance Unit **Nuala Byrne**, Head of Data Governance Support **Annette Denvir**, Head of Data Policy **Audrey Pender**, Head of Data Analytics **Erik Ernits**, Head of Data of Government Office **Arturo Rivera**, OECD, Lead, Data-driven public sector & South-East Asia, Digital Government and Data Unit
04.10.2022	Government Office	**Erik Ernits**, Head of Data
06.10.2022	Government Office	**Erik Ernits**, Head of Data
10.10.2022	Government Office, OECD	**Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office **Ivar Hendla**, Strategy Adviser of Government Office **Arturo Rivera**, OECD, Lead, Data-driven public sector & South-East Asia, Digital Government and Data Unit
11.10.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office
18.10.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office
25.10.2022	Government Office	**Dmitri Burnašev**, Deputy Strategy Director of Government Office **Erik Ernits**, Head of Data of Government Office

4.2 Government Memorandum TO-BE Process

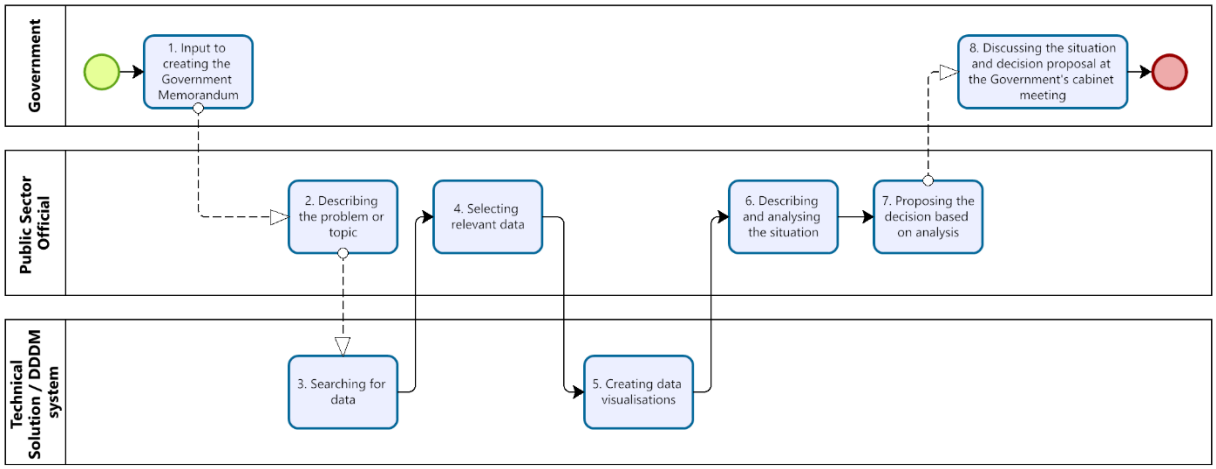
Final Ambition



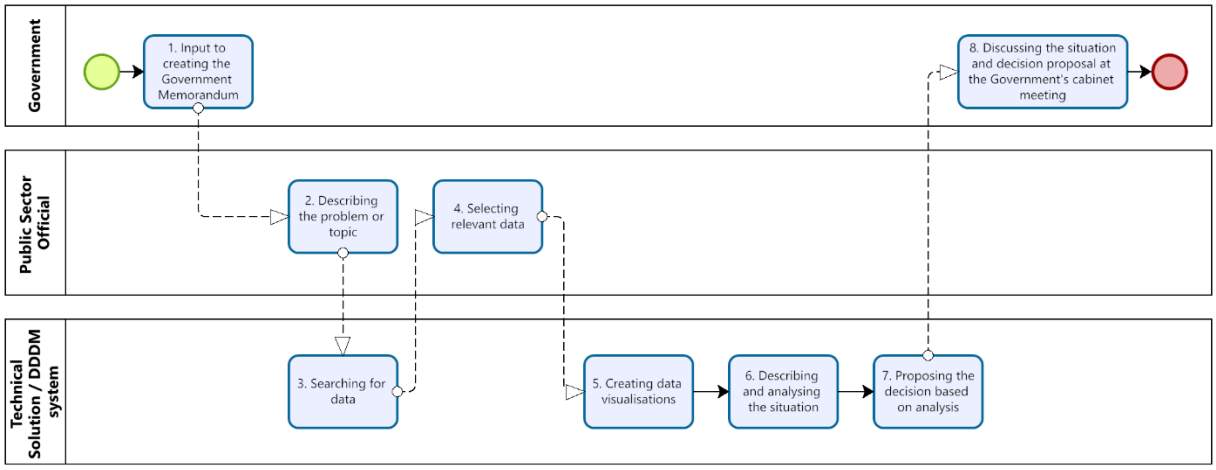
Phase 1 – Data Access (Search)



Phase 2 – Automation of the Data Transformation



Phase 3 – Automation of the Analysis


Government
1. Input to creating the Government Memorandum
8. Discussing the situation and decision proposal at the Government's cabinet meeting
Public Sector Official
2. Describing the problem or topic
4. Selecting relevant data
Technical Solution / DDDM system
3. Searching for data
5. Creating data visualisations
6. Describing and analysing the situation
7. Proposing the decision based on analysis

4.3 Government Memorandum Output and User Interface

The description is supported by the initial visuals of the potential user interface, to illustrate the use of various alternative functionalities of the future DDDM system. As mentioned earlier in the report, one of the key principles of the vision and the future system is that the user interface of the solution will be intuitive to use, user-friendly and it will provide positive overall user experience. This principle was kept in mind when the initial user interface visuals were created.

The aim is that the person working on preparing a new memorandum, can do it on a web-based platform by using dynamic system. Figure 13 illustrates the first step, where new memorandum can be started by filling in the necessary information fields, like the subject of the memorandum and other key fields.

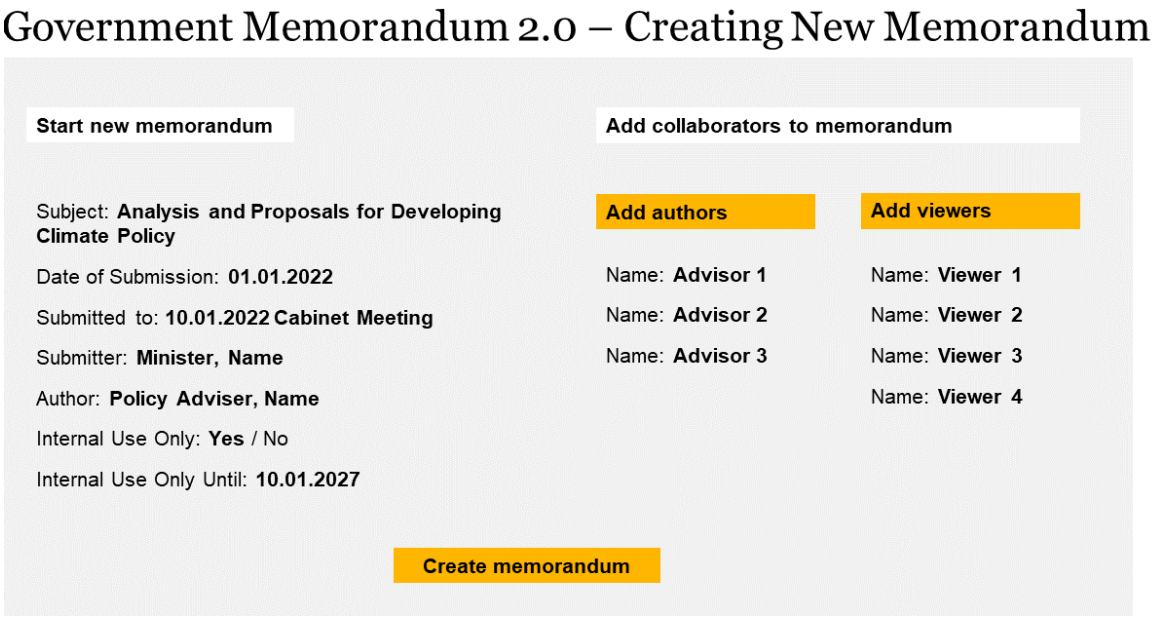


Figure 13. Government Memorandum 2.0 – Creating New Memorandum

Then collaborators (see Figure 14) could be added/connected to the specific memorandum and they can either be co-authors of this specific memorandum who can edit and iterate the memorandum content, or they can be viewers who can access the memorandum content, and view and/or leave comments to the authors who could then consider iterating the content before finalising and submitting the memorandum to government.

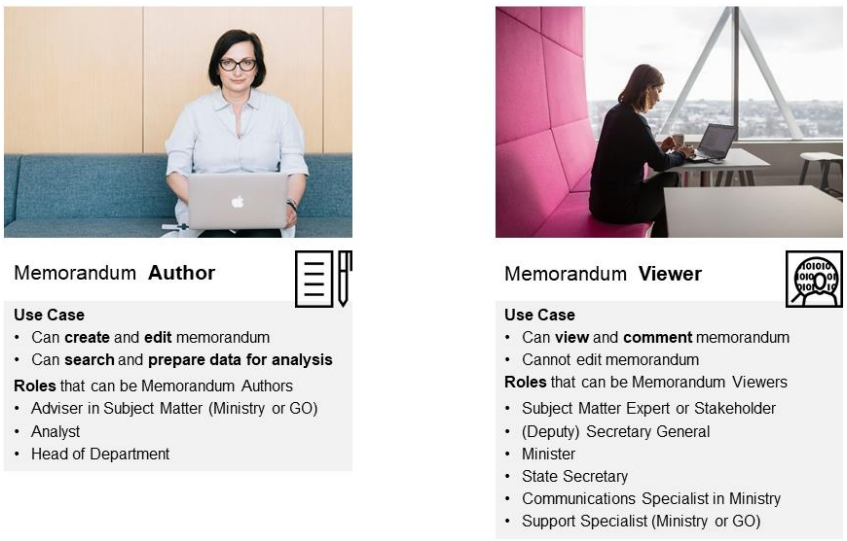


Figure 14. Government Memorandum 2.0 – Users/Collaborators

When the initial information is filled in, and co-authors and/or viewers are added, the button of “Create memorandum” could be clicked. Then the system creates new memorandum and adds in/pre-fills the memorandum’s information on next sections/pages.

The next section (see Figure 15) is focused on describing the background of the subject matter and the problem that needs solving. The aim of this part of the system is to identify relevant keywords about the subject that could be used for searching and finding potentially relevant data.

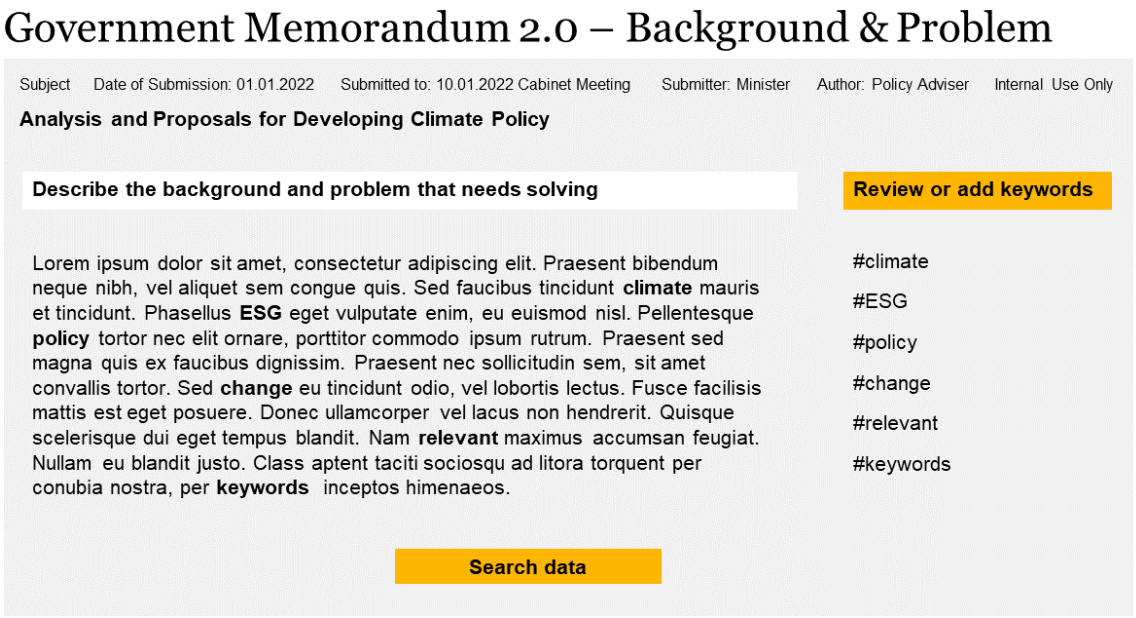


Figure 15. Government Memorandum 2.0 – Background & Problem

When the background and problem are described, and the relevant keywords identified, reviewed, or added, the button of “Search data” could be clicked. Then the system works with data, searches for specific data that is connected to the keywords and displays potentially relevant search results / data sources in the next section. The technical aspects of how data is integrated into the DDDM system and how the search would work, are outlined under the paragraph DDDM System Functionality.

The next section (see Figure 16) displays search results of potentially data sets that are connected to the keywords and could be used for data visualisation and further analysis. The person working in the system

could check the data quality control measures, metadata availability and other relevant information available about the data sets, for example whether the data set has been analysed and visualised before. It is important that each data set has data steward or data manager contacts attached to it, so when a question rises about specific data set or its variables, there is a contact person available to answer them.

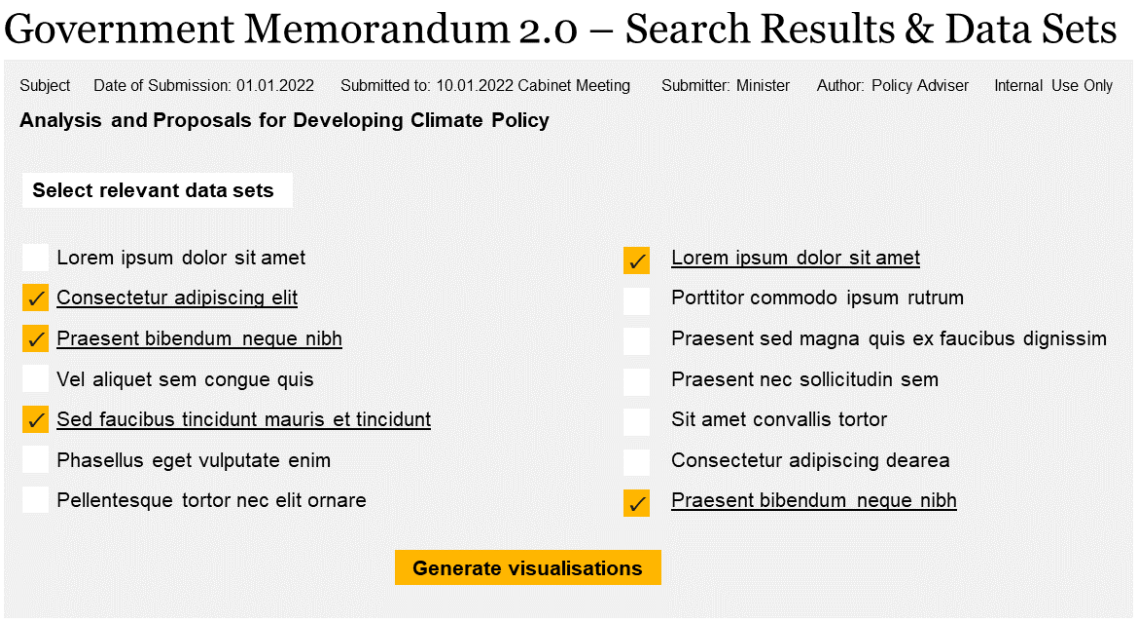


Figure 16. Government Memorandum 2.0 – Search Results & Data Sets

When the relevant data sets are selected, the button of "Generate visualisation" could be clicked. In the DDDM system, data preparation, modelling and analysis is then taking place. The technical aspects of data preparation, data modelling and data analysis are outlined under the paragraph DDDM System Functionality.

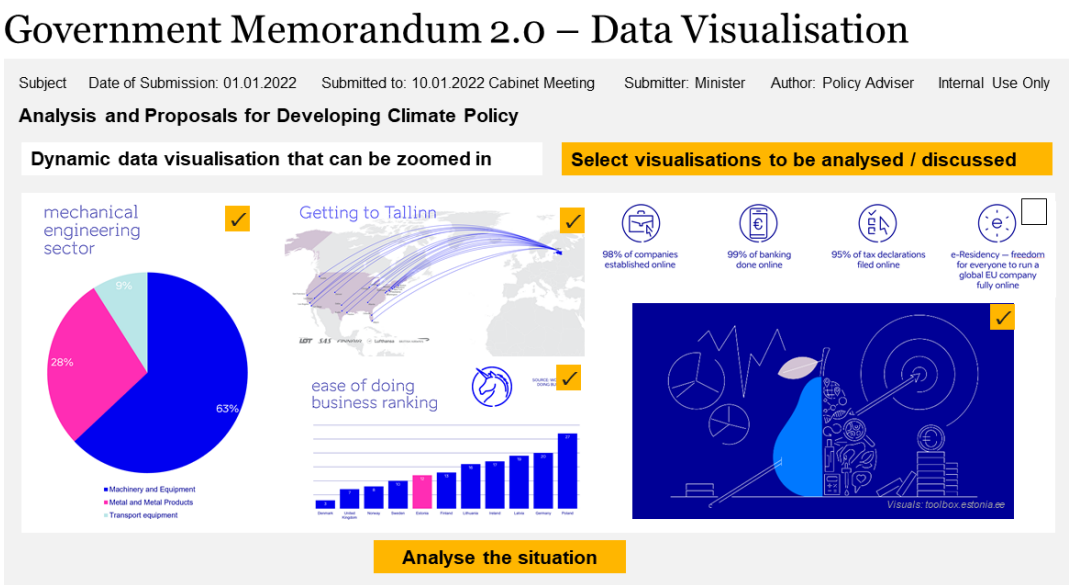


Figure 17. Government Memorandum 2.0 – Data Visualisation

The next section (see Figure 17**Error! Reference source not found.**) shows data visualisations that are generated based on the data from previously selected data sources. The vision is that the system is able to build (custom-made) visuals that are easy to understand, that look modern and display the data in an informative and beautiful way.

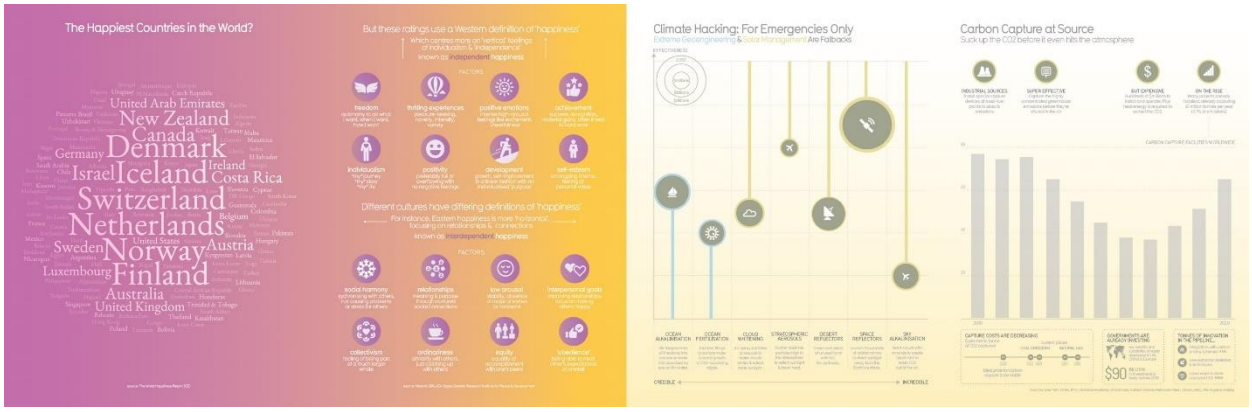


Figure 18. Examples of Advanced Data Visualisation. Author: David McCandless[29]

The system could display several visualisation as shown on Figure 17 of which most relevant visuals can be selected by the person working on the memorandum. Or the system is developing a more high-end and andvanced infographic as shown on examples of Figure 18, Figure 19 or Figure 20

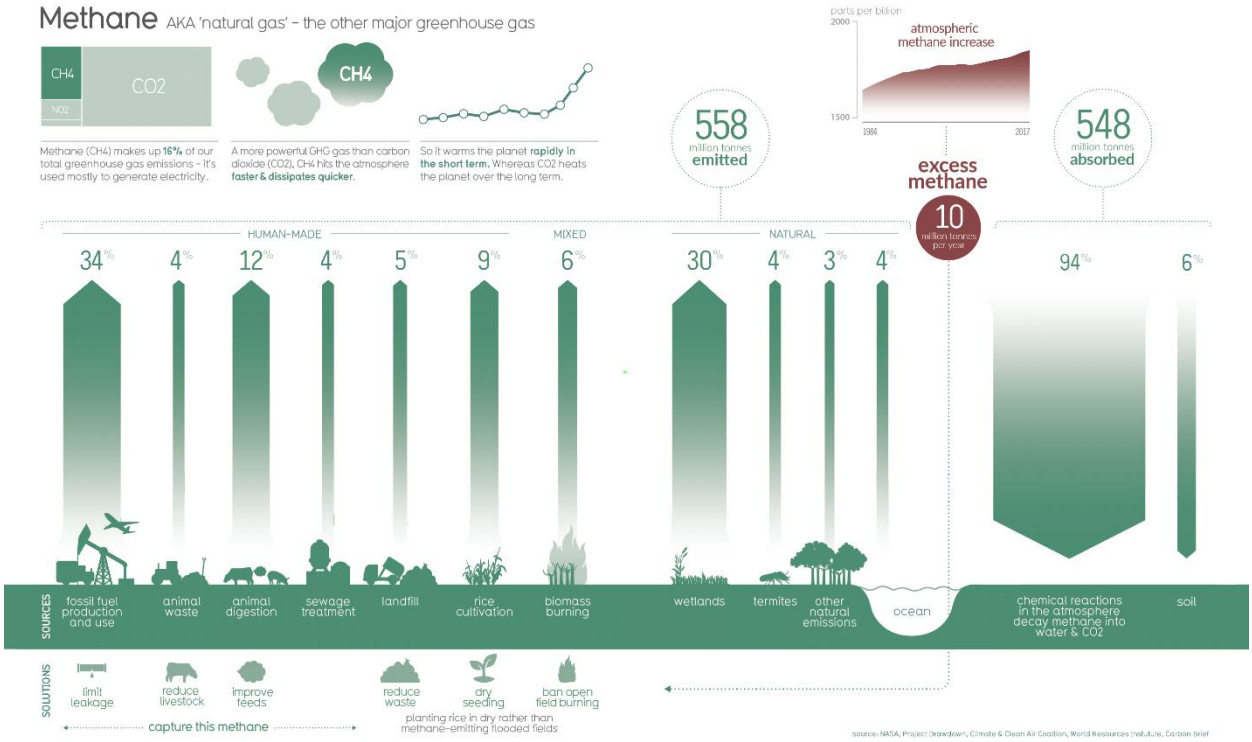


Figure 19. Example of Advanced Data Visualisation. Author: David McCandless

It is important to highlight that those advanced examples of infograpics are built manually by data scientists, data analystis, graphic designers focused on information design, front-end developers and animators. Often several of the mentioned specialists work together to produce an advanced custom-

[29] https://informationisbeautiful.net/visualizations/beautiful-news-the-book-by-david-mccandless/

made infograpic. Usually, a fixed data set (meaning that the data is static and will not be updating or changing) is used to produce a static outcome – an infographic – that cannot be zoomed in a dynamic way that is desired in the DDDM vision.

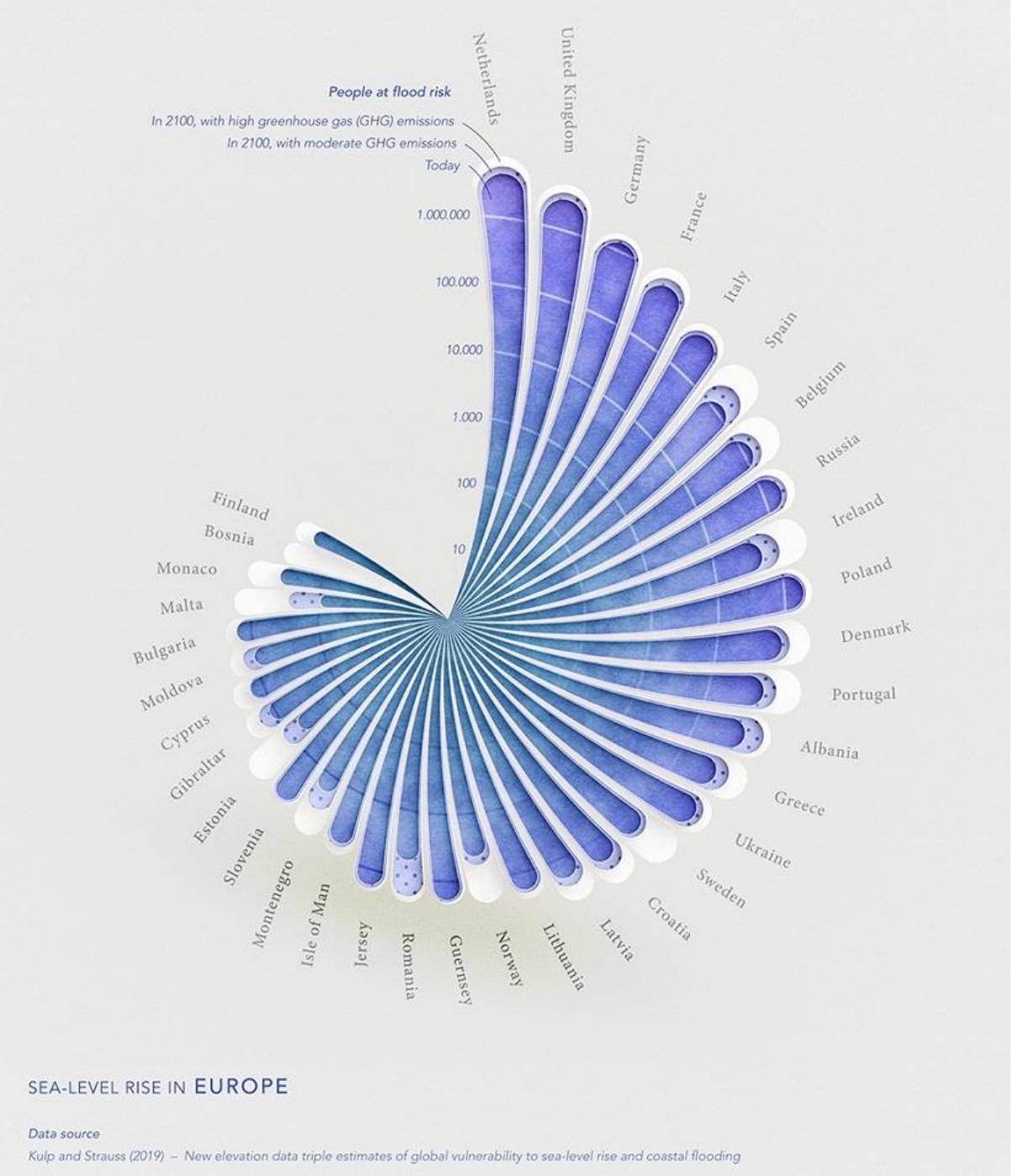


Figure 20. Example of Advanced Data Visualisation. Author: Visual Capitalist[30]

So it will be a challenge to achieve desired visualisation results aprovided by the automated machine, that would be similar to the advanced man-made infograpics and that would enable dynamic functions of zooming the visuals in and out to explore the data in greater detail.

[30] https://www.visualcapitalist.com/cp/sea-level-rises-2100-by-region/

In the next section (see Figure 21) analysis and discussion of the situation is created. In early phases of developing the DDDM system, this part of the analysis and interpretation is considered to be done manually by the official creating the memorandum. In later, more advanced phases of the TO-BE process, this part is visioned so that the DDDM system is able to intreprate data by itself. By producing intelligent analytical description of the situation and suggesting the decision proposals automatically.

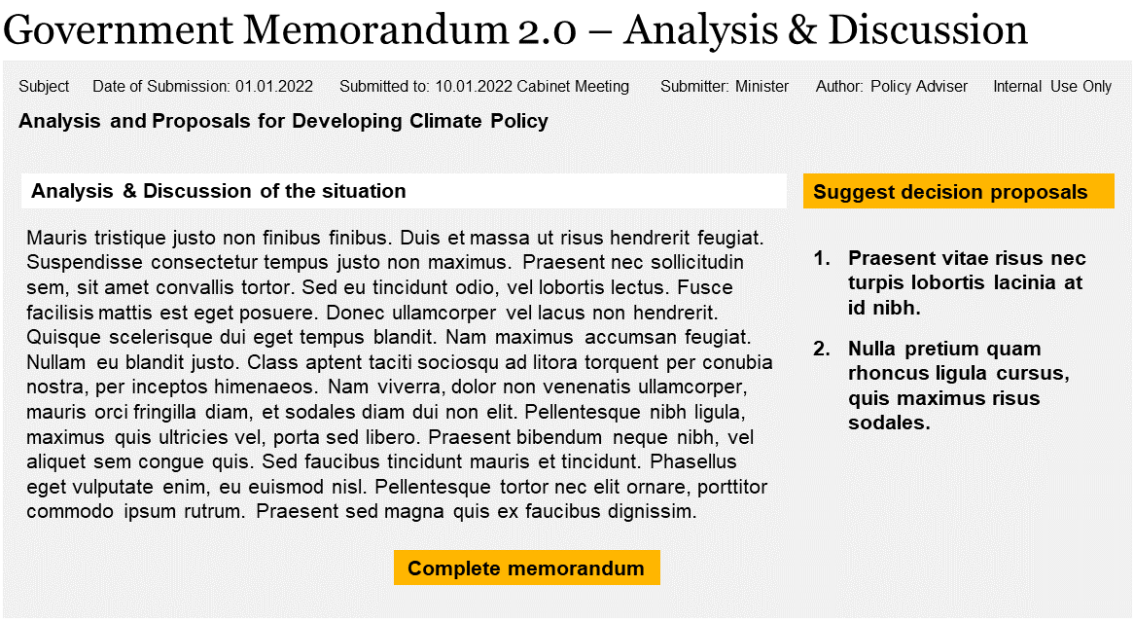


Figure 21. Government Memorandum 2.0 – Analysis & Discussion

By clicking the button “Complete memorandum,” the system will produce the Government Memorandum 2.0 output (see Figure 22) that includes a compact overview of memorandum components, which could be clicked to discover more details and visuals could be zoomed in or out to analyse the data content in deeper detail.

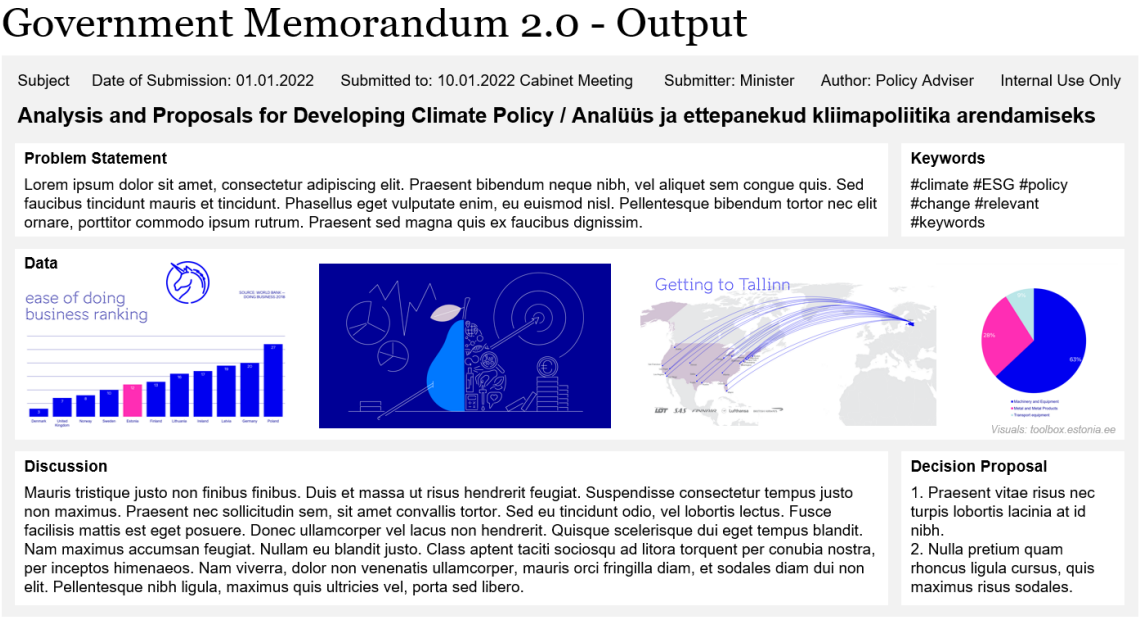


Figure 22. Government Memorandum 2.0 - Output

But to ensure that the potential decisions made by the government are aligned with stakeholders of the specific matter that the memorandum analyses, additional final steps of creating the government memorandum are suggested to take place in the TO-BE process:

1. Collaboration with stakeholders and stakeholders commenting the memorandum created,

2. Needs-based iteration of the memorandum content,
3. Confirming and approving the memorandum,
4. Memorandum discussion at the Government's cabinet meeting.

Step 1 could potentially take place within the DDDM system when system has the Memorandum Viewer role developed and applied. Step 2 can be done in the system by reviewing or re-selecting the data and visualisations. Step 3 of confirming and approving the memorandum is not visioned yet and within step 4 the dynamic output of DDDM system can be used instead of current PowerPoint slideshows to communicate key messages and present decision proposals together with data visualisations supporting the data-driven decision-making process.

Visit our website:

Find out more about the Technical Support Instrument:

pwc

Funded by the European Union