Universiteit Leiden

ICT in Business and the Public Sector

Analyzing and improving the Dutch government ICT dashboard

Name: Jermaine Chin - A - Loi
Student-no: s1321072

Date: 23/09/2022

1st supervisor: Prof. dr. J.M.W. Visser
2nd supervisor: Prof. B.A.J. Klievink

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

ABSTRACT

Background

In 2011 the central Dutch government launched an ICT dashboard for big ICT projects. This was done in accordance with the "Law Open Government". This law states that there is a need for openness and transparency of the big ICT projects within the central Dutch government. The dashboard launched in 2011 has received substantial criticism and has gone through various iterations since its inception.

Objective

The main objective of this research is to design dashboard improvements that increase the usefulness of the dashboard. The purpose of these improvements is to increase the transparency and the accountability of large ICT projects within the central Dutch government. We propose that the dashboard should provide insights into costs and benefits after the initial development project has been concluded and the system has been taken into operation.

Method

We derived requirements for the dashboard improvements from the following three sources: letters from R.W. Knops to the house of representatives, an analysis of different governmental dashboards, and an analysis of the raw data of the Dutch government. For the analysis of the governmental dashboard, we used the method of Sarikaya et al. (2019) to learn how other ICT dashboards are designed. With the findings of the analysis, we were able to design more sensible key performance indicators (KPIs). These KPIs were designed with the help of the Goal Question Metrics (GQM) approach (van Solingen & Berghout, 1999). When creating the new KPIs we used the Create Sustain Benefit Model (CSB). This model was used to ensure that each section of the project "create", "sustain", "benefit" met the requirements that we established. To give an example of what a new dashboard could look like, we designed mock-ups of the dashboard with the help of the newly designed KPIs.

Results

After analysing the three governmental dashboards we found that these dashboards all share two goals: decision making for the politicians and creating awareness for the general public. With the help of Project R we analysed and processed the raw data from the Dutch governmental dashboard and we visualized this data through mock-ups.
Next to that we provided advice to the Dutch government on what steps they can take to improve their dashboard.

Conclusion

After completing this research, we would highly recommend the Dutch government to adhere to the newly composed requirements that can be found in Chapter 7.1. Within the current Dutch governmental dashboard not all the data is complete. Once the Dutch government adheres the mentioned requirements, they will be able to complete the missing data points. Next to the mentioned KPIs, there are also visual requirements that the Dutch governmental dashboard is partially missing. Within the Chapter 6.1 Modeling we created detailed mock-ups

to visualize how the data of the Dutch governmental dashboard could be presented in a more adequate way to various stakeholders. Our advice within the mock-ups can also be summarised the following points: *color coding adjustments, adding missing financial information, adding functionalities.* Our dashboard improvements can be used by the Dutch government to increase the usefulness and transparency of the Dutch IT dashboard.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor prof. dr. ir. Joost Visser for his expertise, time, guidance, and new insights given during my research process. It was a privilege having him as my supervisor as he was always understanding and patient throughout this project.

Second, I would like to thank Prof. B.A.J. Klievink, my 2nd supervisor, for the new insights he gave me through his feedback.

And finally I would like to thank my family, girlfriend and friends for the support, motivation and love they gave me during my research project.

TABLE OF CONTENTS

1. Introduction

In accordance with the "Government Information (Public Access) Act" also known as "Wet van Openbaarheid van Bestuur", the Dutch government wants to give more insight to the public into the ICT projects that are run. To do this they created the Rijks ICT-dashboard. Within this dashboard all the ICT projects that exceed an investment of 5 million are presented.

This chapter is arranged by first describing the background of the ICT Dashboard. Followed by the problem statement. Then the research question will be proposed and finally an overview of all the chapters within this thesis.

1.1 Background

In 2011 the central Dutch government launched the "Rijks ICT-dashboard". This dashboard is meant to show different stakeholders how big IT projects are currently being run (Arnoldus & Visser, 2011). According to Rijks ICT-dashboard (2021), the criteria for an ICT project to be entered within the dashboard is that the amount of money spent on the project is at least five million euro. The ICT Dashboard is also meant to give more openness of ICT projects within the government to the public eye. The reason for this transparency is to be in accordance of the "Law Open Government" (Open Overheid | Digitale Overheid | Rijksoverheid.Nl, 2021). Currently within the ICT Dashboard there are 342 projects that exceed the cost of 5 million. The total ICT expenses of the Dutch government only in the year 2021 were 2.886 million euros. The current active projects (101 projects) have an initial estimate value of 4.744 million and an actual estimate value of 5.870 million euros. This means that there is an increase of 24 % between the initial estimate value and the actual estimate value of the ICT projects. The biggest portion of these costs is being funded by the taxpayers in The Netherlands (Rijks ICT-dashboard, 2022).

Since the inception of the dashboard, there have been a few different iterations of it and the Dutch government is still working on a new iteration at this time. The reason for the different iterations was that the dashboard was not showing all the information that the stakeholders would like to see. On the 4th of March 2020, the Dutch Secretary of State of Home Affairs and Royal Relations R.W. Knops sent a letter to the House of Representatives about a new version of the ICT Dashboard. Within the letter they mentioned a newly added landing page. Within the landing page they implemented the following points (R.W. Knops, 2020b):

1. The total number of big IT projects
   The requirement is to give the stakeholders an overview of all the IT projects that are being implemented by the Dutch government
2. The current cost estimates
   The requirement gives the stakeholder the information about the cost of IT projects
3. The average run through time of projects
   The requirement gives the stakeholders an overview on what the average run through time is of a project.

These points will be a part of the new update of the ICT Dashboard. Currently there is no information about when this update will be delivered.

On the 1st of December 2020, there was another update about the ICT Dashboard, about the fact that the House of Representatives should be one of the stakeholders. The House of Representatives will be interviewed in order to identify the requirements. The research will be done by KPMG. The new dashboard updates are currently in the process of development (R.W. Knops, 2020a).

On the 4th of March of 2020 there was a letter from Ministry of the Interior and Kingdom Relations to the Speaker of the House of Representatives about the strategic Information-agenda of the Dutch government. The letter stated that they are further developing the ICT dashboard. That within the dashboard the cost for the maintenance and management of the IT systems will be added (R.W. Knops, 2020). This will be added to our requirements of the IT Dashboard.

In 2015 "Het Bureau ICT-toetsing (BIT)" was established in order to improve the control of ICT- projects within the Dutch government. In 2019 the decision was made by the Minister of Internal Affairs that BIT will proceed their work as an independent advice bureau (Jaarrapportage 2019Bureau ICT-Toetsing, 2020). Therefore the "Adviescollege ICT-toetsing" was established in 2020. Each year the "Adviescollege ICT-toetsing" publishes a year report directed to the Dutch government regarding the ICT-projects. In the most recent report regarding the year of 2021 the "Adviescollege ICT-toetsing" gave the following four general recommendations:

1. Make the intended benefits concrete and aim to achieve their realization
2. Think carefully about future processes
3. Increase focus on client-supplier relationship
4. When choosing a solution direction, also take the management and maintenance phase into account (Adviescollege ICT-Toetsing, 2022)

This will be considered for our advice.

1.2 Problem Statement

The research problem for this study is that the central Dutch government has been working on this dashboard since 2011. Within this period there has been a lot of criticism on the ICT dashboard. The most criticism regards the fact that the information given on the Dashboard is not sufficient (Ruud Leether, 2019).

At the moment, the central Dutch government is already working on a new version of the dashboard. However, this doesn't ensure that this new iteration will be the one to meet all the requested requirements. Also, the end date of this iteration has not been communicated yet (R.W. Knops, 2020a).

The goal of this research is to figure out how the ICT dashboard can be improved to get more information across the different stakeholders. We will contribute towards the goal of improving the dashboard, therefore we have taken the initiative to explore improvement directions within this thesis. The task at hand is to make the dashboard more transparent. A

way of doing this is selecting the right information. This information will also be visualized in a proper way so that a stakeholder can view the dashboard and immediately get the data they are looking for.

The academic value of this research is, firstly, the analysis of the governmental dashboards. By analysing multiple governmental dashboards, we will be able to deliver a conclusion on the key requirements that are needed for a well-functioning dashboard. We will also analyse how efficient the dashboards are in relaying information to different stakeholders. Another academic value of this research will be a model with governmental dashboards encoded through the method of (Sarikaya et al., 2019).

This research has a high practical value because there are currently 110 active IT projects with a cost estimate of 5,584 million euros (Rijks ICT-dashboard, 2021). When it comes to this amount of money put into different projects, it is very important that all stakeholders can have an insight of what this money will be used for. The goal is to make the dashboard more transparent. If the new iteration is finished before the end of this research, the study can be used to get another perspective of how the dashboard could be improved.

1.3 Research question

The main focus of this research is examining how the central Dutch government can represent ICT projects better on their already existing ICT dashboard. Within this research the main research question is suggested:

> ***What can the central Dutch government do to present their data on IT projects in a more adequate way to various stakeholders?***

The following sub-questions will help to answer the main research question.

> *What are other governments doing in the field?*

Over the past few years there has been a lot of criticism on the current Dutch ICT Dashboard. The most criticism regards the fact that the information given on the Dashboard is not sufficient and does not meet all the requirements requested by the Dutch government (Ruud Leether, 2019). This sub question was created to compare the current Dutch ICT Dashboard to the other existing governmental dashboards. The goal of this orientation is to understand how the other countries visualize the information of their IT projects within their governmental dashboards and if there is anything that we could learn from them and use within the newly designed Dutch ICT dashboard.

> *What kind of information is important to the involved stakeholders?*

Within the orientation phase we stated that the involved stakeholders are; the general public (citizens), government officials and IT business professionals. This sub question was created to

define the information need of the stakeholders for a governmental IT dashboard. We stated that the current dashboard does not fulfil all its requirements. As the governmental data dashboards display complex data to a user (such as general public (citizens), government officials and IT business professionals) the goal of the dashboard is to answer all the needs of all the involved stakeholders. To be able to advise upon a new governmental dashboard it is crucial to understand these needs of the stakeholders.

What information can be analysed from the raw data?

Within the current Dutch governmental dashboard, it is possible to download the data. For this research it is important to know what information can be derived from the data to define what data still needs to be added, to meet all the earlier mentioned requirements. This sub question was created to examinate what type of data can be found within the dashboard and determine what data is still missing.

What is the best way to visualize the analysed information?

Within the orientation phase we discovered that the current Dutch governmental dashboard went through different iterations. Within the different versions of the dashboard also the data was visualized in various ways. The most important purpose of the governmental dashboard is to visualize the data in an accessible and clear way for the stakeholders. This sub question was created to examine the best way of representing the data to all the involved stakeholders.

1.4 Overview of all chapters

Within in this section the overview of all the chapters will be described.

Chapter 2. Within this chapter the key concepts are highlighted and explained, followed by the literature review. The explained key concepts are: *Governmental Dashboard*, *KPIs*, *Stakeholders, Freedom of Information Act*. Also, the key framework *Create, Sustain, Benefit* is explained within this chapter.

Chapter 3. Within this chapter we delve deeper into the methodologies that have been used within the research. The methodology for the thesis was Design Science Research (Hevner et al., 2004). This gave us the structure for the research that has been done. For the creation of the dashboard, we used CRISP-DM (Chavez et al., 2020). This methodology was used to retrieve, prepare, transform, and model the data. The last used methodology was Goal Question Metrics (GQM) that helped to establish the KPIs.

Chapter 4. Within this chapter we analyse and compare existing dashboards. We looked at three existing governmental dashboards: the Australian, the American and the Dutch dashboard. These dashboards were compared in Chapter 4.1 and encoded by using the method of Sarikaya et al. (2019) . Our findings were used to establish the requirements that can be found in Chapter 4.2.

Chapter 5. Within this chapter we executed the *Data exploration, Preparation phase* and *Transformation phase*. These are the phases prior to the creation of mock-ups for the dashboard. For the *Data exploration*, *Preparation* and *Transformation phase* we used the CRIPS-DM method. The findings of the phases were also used to establish the requirements of the new governmental dashboard.

Chapter 6. Within this chapter we visualized the KPIs and designed mock-ups for the new Dutch governmental dashboard. These mock-ups brought together all the requirements that we found throughout the research.

Chapter 7. Within this chapter we answer the main research question by giving answers to the sub questions. Next to that we also discuss the reliability and validity of the research. Finally, we present the limitations of this research.

2. Conceptual Foundations

In this chapter the conceptual foundations are built. This chapter serves as foundation, explaining the issues driving this research. We will discuss the following key concepts in Section 2.1: *Governmental dashboard*, *KPIs*, *stakeholders* and *Freedom of Information Act* and review the literature in Section 2.2. In Section 2.3 we discuss a model, called *Create Sustain Benefit*, that we will employ to capture all aspects of IT projects.

2.1 Key Concepts

Within in this section all concepts that are of importance for this research are highlighted and explained. These principles are needed for the context of the research.

2.1.1 Governmental Dashboard

Governmental data dashboards display complex data to a user about the governmental departments and their decision making. These dashboards display different types of data across all the departments, from project planning, slippage to spending. Dashboards allow organizations to improve visibility in different areas of their business, as well as provide insights to citizens and support operational choices. KPIs are the metrics used in dashboards, measuring the performances of organisations. In general, KPIs help to drive organisations into the earlier determined direction (Laurent, 2007). Laurent (2007) emphasizes how the governmental data dashboards must serve the government itself the citizens. In addition to this, Use of Dashboards in Government (2011) states that there are two key elements for a governmental dashboard: the design of the dashboard and the performance metrics used in the dashboard (Use of Dashboards in Government, 2011).

Over the past few years there has been a growing interest in using digital tools to increase the transparency of the government, so that the citizens have a clear sight of the projects run by the government and the data around them. Transparency optimists argue that being more transparent with the public, by sharing this data with them, results in more trust in government (Grimmelikhuijsen & Meijer, 2014). Also following the core value of Freedom of Information Act (In The Netherlands known under the name of *Wet Openbaarheid van Bestuur*) the government should be transparent and promote democratic accountability, allowing the public to receive an accurate picture of what is happening inside the government. This includes access to the government records and documents, data, and insights of projects, all being visible on a governmental IT dashboard (Graham, 2012).

The main object of this research is the Dutch governmental IT dashboard. This dashboard has been redesigned multiple times, but still hasn't met all the needs of different stakeholders. Within this research we will investigate how the dashboard could be redesigned to better meet the needs of stakeholders (Laurent, 2007).

2.1.2 KPIs

The operational goals of an organization can be regarded as standards within the organization. These standards are derived from strategic goals, being decisive for the success of an organisation. Key Performance Indicators (KPIs) are factors that measure the extent to which the earlier mentioned standards are achieved in a specific period of time (Parmenter, 2015).

Drafting and application of these KPIs has become common practice within organizations, especially in the public and the non-profit sector (De Bruijn & Van Helden, 2006). The reason for increased use of KPIs in public and non-profit organizations is that these organizations are becoming more transparent about the allocation and usage of their resources, as well as trying to be more efficient about them. Measuring performance using KPIs, enables the organisation to have more insights into the use of their resources.

Within this research we use KPIs to model all the information needs of the stakeholders. Within the redesigned dashboard these KPIs will be visualized to provide the needed information to the stakeholders (Miles, 2017).

2.1.3 Freedom of Information Act

The USA's Freedom of Information Act states in general that all citizens should be able to request and receive access to government information, besides any records with protected information and confidential data related to people or companies. This right has in some form been implemented in over 100 countries around the world. In over 80 countries the principles of the Freedom of Information Act have been enshrined in their constitutions (Darch & Underwood, 2010). In The Netherlands the Freedom Information Act is known under the name of *Wet Openbaarheid van Bestuur* (WOB). For this research the Freedom of Information Act is very important, without it there would have not been any access to the data of the Dutch governmental ICT projects. To model the redesigned IT dashboard, we used the existing data of all ICT projects of the Dutch government above the budget of one million, as is available from the current national IT dashboard.

2.1.4 Stakeholders

Stakeholders can be defined as people or parties with an interest in something, take part in it, have conflict with it, or could have influence on it. This could include individuals, organizations, or networks (Sarikaya et al., 2019). Stakeholders can take an internal or an external part in the project. Depending on the scope and complexity of a project there might either be very few or extremely large numbers of stakeholders. By researching who is directly and indirectly impacted by a project, it can be determined who these stakeholders are. The needs and requirements can vary per stakeholder, some might require little interaction, while others require constant communication. However, for a high stakeholder engagement it is crucial to satisfy the requirements of stakeholder (Rowley, 2011).

For the governmental dashboards there is no advanced domain expert needed to understand the data (Sarikaya et al., 2019). The data is also meant for the general public.

Within this research we define three types of stakeholders: the general public (citizens), government officials and IT business professionals. These are the key stakeholders and are directly impacted by the IT governmental dashboard. The dashboard developed in this research should meet requirements of these three stakeholders. As all these stakeholders have different types of backgrounds, it is important that the information of the governmental dashboards is accessible/ clear to all of them.

2.2 Literature review

Within the literature review, we will discuss related work to creating a governmental IT dashboard. We divided this in three subjects. These subjects are: *The Design of the dashboard*, *The content of the dashboard* and *the dashboard's KPIs*.

2.2.1 The design of the governmental Dashboard

When designing a governmental dashboard for the Dutch government, the most important aspect that a dashboard needs to adhere to is accessibility. Accessibility can have different understandings. In this context we refer to visual accessibility. As a governmental dashboard is being viewed by the general public, it also needs to be accessible and clear to the public with different types of backgrounds. One of them would for example be disabled people. Therefore, a governmental dashboard needs to adhere to the design restrictions and requirements spelled out in detail in the European Accessibility Act (Kloeckl, 2019). Within our research the European accessibility act will not be considered. The reason for this is that the research focuses more on the content of the dashboard. However, it is important to mention the accessibility act for the future dashboard. The future dashboard should adhere to the regulations that are set within the accessibility act and be accessible to individuals with disabilities (European Accessibility Act - Employment, Social Affairs & Inclusion - European Commission, 2022).

Another aspect is that the governmental dashboard should convey the needed information and be comprehensible to the users. It would be harder to educate a large group of people on how to use and understand the dashboard than to simplify it so that it is understandable to its audience. Within governmental data, there is a lot of advanced numeric information. The general public has a different knowledge base and cultural background. To ensure that the complex statistical information is equally accessible to all the general public, the information and the designs call for simplicity. Not only the quality of the data should be taken in consideration, but also pure visual aspects such as: the use of colors, shapes, screen size, and the positioning of the elements on the screen. All to guide the user's attention and display the complex information in an accessible way (Kloeckl, 2019).

Within the design mock-ups of the governmental dashboard, we will take into account how to make the dashboard more understandable to the general public. This can be found in Chapter 6.1 and Chapter 6.2.

2.2.2 The content of the governmental Dashboard

Within this research the main focus is the Dutch governmental dashboard that collects data from IT Projects within the government. We will also compare the Dutch governmental dashboard to other governmental dashboards that serve the same purpose. These dashboards all have in common that they are available to the public and the stakeholders to inform them about the state of the IT projects within the government. The data that the government collects from IT projects are for example: the project name, the description, the estimated cost of the project, the duration of the project, etc.

"In data science, the sharing, use and interpretation of data are key aspects in bridging the gap between the government and the public. Dashboards can be used to release information for governmental decision-makers, but also for the public to scrutinize government actions, to engage in the decision-making processes and to improve decision-making"(Maheshwari & Janssen, 2014).

"Dashboards should help to facilitate transparency, governance, trustworthiness and enable citizens' to participate in decision-making in smart cities" (Allio & Fahey, 2012).

As stated by Maheshwari & Janssen (2014) and Allio & Fahey (2012) the main purposes of a governmental Dashboard are: data transparency, improved decision-making, lowering expenses and monitoring service performance. However, a commonly occurring problem is that the data is not complete, not updated periodically and important data points are not being collected. Within this research we will give a recommendation on how the content of the Dutch governmental IT dashboard can be improved.

2.2.3 Dashboard project Key Performance Indicator

Within dashboards, Key Performance Indicators (KPIs) are commonly used. This is the case for the business sector as well as the public sector. The only difference within the KPIs for the business or the public sector is the type of performance that they want to track. The reason for this, is that the business sector often has different goals than the public sector. For example, the business sector KPIs can be driven by the revenue, cost, and profit. Usually this is not the case for the public sector.

This research focuses on the Dutch governmental dashboard which is the public sector. When defining the Key Performance Indicators for the Dutch government it is important to first define what will increase the performance of the Dutch government (Parmenter, 2015).

"While government inputs and internal processes matter in good policy-making, they are mainly assessed in terms of the final results and the impact of policies on the broader society." (Guillaume Lafortune, 2018).

KPIs are special performance tools and should not be manipulated to make the situation look better than it is. KPIs are essential to not be corrupted by the stakeholders to make them look better (Parmenter, 2015). KPIs are nonfinancial indicators that are measured daily or weekly. According to David Parmenter, it is a myth that performing monthly performance measurements will improve the performance. This means that for the dashboard to have an added value, the data needs to be updated regularly. At the minimum weekly, but optimal would be daily.

Within this research we will review the current KPIs of the Dutch governmental IT Dashboard. We will then remove, change, or create new KPIs, in order to make the dashboard

more transparent and able to support decision making of the stakeholders. The eventual goal of defining the KPIs for the Dutch government is to show the stakeholders whether the IT projects are meeting the initial expectations.

2. 3 Create Sustain Benefit (CSB)

To provide a holistic view on IT projects, a dashboard should not only provide information on the initial development cost of IT projects, but also on the operational costs incurred while the developed system is in operation as well as the benefits that stakeholders derive from the operational system. A concise rendition of how development cost, operational cost, and operational benefits are related to each other can be seen in the diagram in Figure 1. This diagram is taken from lecture notes by Visser (Visser, 2013).

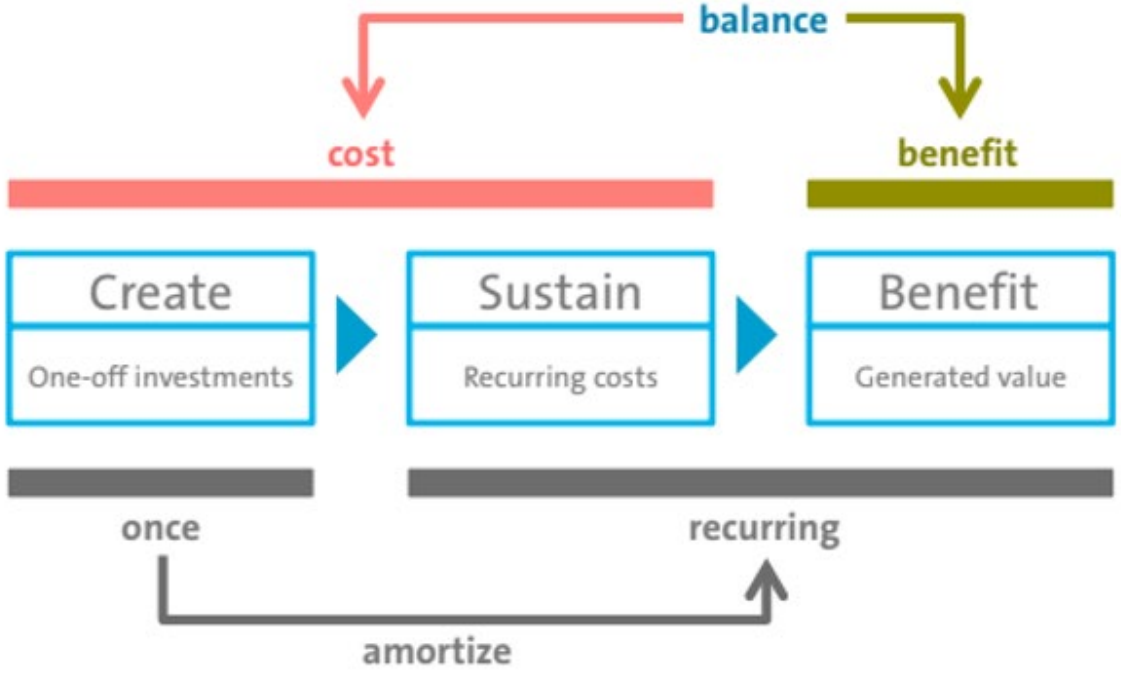


Figure 1 High-level representation of the relationships between development costs (Create), operational costs (Sustain), and operational value (Benefit) of IT systems (Visser 2013).

This model shows the balance between cost and benefit. According to Visser, there are three steps to the cost-benefit analysis; *create, sustain,* and *benefit*. *Create* and *sustain* are part of the cost of the project. The *benefit* is the generated value. When creating IT solutions, it is important that the benefit of the project outweighs the cost of it. To find this out it is important to know what the initial costs are of the entire project (one- off investment) and what the cost will be to sustain the system (recurring costs). This is then balanced with the benefits that the IT system will give the organization. This model will be used to analyze the cost and benefit (Fuguitt & Wilcox, 1999).

"Cost-benefit analysis has proven to be a useful decision-making tool with widespread application the analysis provides information to aid public managers who are considering any of an number policies with social goals or consequences, such as environmental policies, health and safety regulations, transportation and water resource project, recycling programs, youth programs etc. In depth knowledge of the practical steps in an objective cost benefit analysis can enable these decisionmakers to understand, interpret and critique a particular analysis and thereby make more informed decisions" (Fuguitt & Wilcox, 1999).

In reality, a lot of cost-benefit analysis for public sector IT projects have failed. The reasons behind it are that organizations fail to properly monitor and evaluate their projects (Liu & Lin, 2008). Since the benefit of public sector projects is not directly corelating to profit or to capital value, but to social and environmental goals, it is highly important to choose the correct method to perform the cost-benefit analysis (Liu, Y. C., & Lin, C. 2008) . Within this research we are going to define the benefit of IT projects within in the public sector. We are using the Create Sustain Benefit model to make a cost benefit analyses for the IT projects.

3. Methodology

In this chapter the research method for this specific research is described. This chapter also goes into the detail why the suggested method is the correct method for the research and elaborates how the research is organized. Furthermore, this chapter explains the following specifics: in what way the needed data will be collected, prepared, and analysed and how the literature review will be accomplished.

3.1 Design Science Methodology

This research makes use of the Design Science method by Hevner (2007). The motive for using Design Science is the fact that this research is solution oriented, and its main goal is developing knowledge that professionals in a company can use as a solution to their problems.

According to Hevner (2007) there are 3 cycles within this method:

1. The Relevance Cycle

The goal of the relevance cycle of the design science method is to research if the design artifact is innovative and if there is an added value to the environment. This is the first cycle of a design science method. In this cycle the following aspects are stated: the problem statement, the requirements, and what the final results should achieve. Design science research is an iterative method. Therefore, the relevance cycle can be executed multiple times until the final result is achieved. Within the relevance cycle the field testing for the design artifact is also performed. The relevance cycle is based on the environment with the goal to improve it, and to ensure that the designed artifact is innovative and does not exist yet within the environment (Hevner, 2007).

2. The Rigor Cycle

Within the rigor cycle of the design science method the goal is to ensure that the research is grounded on descriptive theory. However, Hevner (2007) argues that it is unrealistic to state that the design research must be grounded on descriptive theories. As the designed artifact should be innovative and it is not always possible to base it on already existing theories. According to Hevner (2007) these theories should serve as sources of creative ideas for the designed artifact, rather than the base of it. Within this cycle it is also of importance that the researcher does a thorough research to ensure that the designed artifact is a contribution to the knowledge base but does not exist within it yet (Hevner Alan, 2007).

3. The Design Cycle

Within the design cycle of the design science method, the artifact is constructed, refined, and evaluated. The requirements of the artifact are created within the relevance cycle. The designed artifact is being iterated until that artifact meets all the requirements. The design and evaluation theories are created within the rigor cycle.

The design cycle is dependent of the rigor cycle and the relevance cycle, even though the rigor cycle and the relevance cycle are executed independently. On top of that the design cycle is

iterative, in the same way as the rigor- and the relevance cycle. The created artifact needs to be tested in the environment (rigor cycle) as well as in the knowledge base (relevance cycle) (Hevner Alan, 2007).

4. Applying the three-design science cycle

Within the first phase also called the Relevance Cycle, we will do desk research on the current environment. The existing central Dutch government ICT dashboard will be analysed and evaluated. Also, the letter from drs. R.W. Knops to the House of Representatives about the current ICT dashboard, will be analysed. This letter gives feedback on the current ICT dashboard. We will use this letter to state what the criticisms and the requirements of different stakeholders are, regarding this current dashboard. These findings will be used in further phases of the research. In this phase we will also investigate different dashboards in the field of the public sector. The research will make a comparison of a few governmental dashboards. This will be done to see what requirements are important to the different dashboards.

Within the Rigor Cycle we will ensure the research project innovation. As described in the previous phase, we will investigate different dashboards in the field of the public sector and compare them. For this comparison we will use the dashboard evaluating method that has been used within the research of Sarikaya et al. (2019) to assess the different dashboards and code them within a table. Example of this table is shown in Table 1. This will be done to see what requirements are important to the different dashboards. All the important requirements found within the analysed dashboards will be merged to help create the new artifact within project R. Artifacts are the by-products that are produced of software development (Raymond Turner & Nicola Angius, 2013). This will be done to ensure that the new dashboard is not a redesign of just one of the existing dashboards, but to make sure that this artifact is truly an improvement. This can then be used to create a new artifact within the design cycle phase using project R.

				Purpose				Audience			Visual Features					Data Semantics		
Goal	Cluster	# Examples	Strategic	Tactical	Operational	Learning	Audience	Vis Literacy	Domain Expertise	Construction	Interactivity	Modify Data/World	Highlighting	Multipage	Alerting+Notification	Benchmarks	Updateable
Decision-Making	1 Strategic Decision-Making	16	Y	Y	-	N	O	-	-	-	Y	N	N	Y	-	-	Y
	5 Operational Decision-Making	14	N	Y	Y	N	O	-	-	-	Y	N	N	Y	-	Y	Y
Awareness	3 Static Operational	10	N	N	Y	N	O	L	-	-	-	N	N	N	-	Y	Y
	4 Static Organizational	8	-	-	N	N	O	M	-	N	N	N	-	N	N	-	Y
Motivation and Learning	2 Quantified Self	7	N	N	Y	N	I	H	N	N	Y	N	-	Y	-	-	Y
	6 Communication	13	-	-	-	Y	P	M	N	N	-	N	-	-	N	N	Y
	7 Dashboards Evolved	15	-	-	-	-	P	H	-	-	-	-	-	-	-	-	Y

Table 1 Example the coded dashboards (Sarikaya et al., 2019)

To create the artifact within project R we will use the *Logic Model* that will be redesigned where needed. This will be based on the same model and a similar application of it, found in the research Awan (2012). The described *Logic Model* can be found in Figure 2.

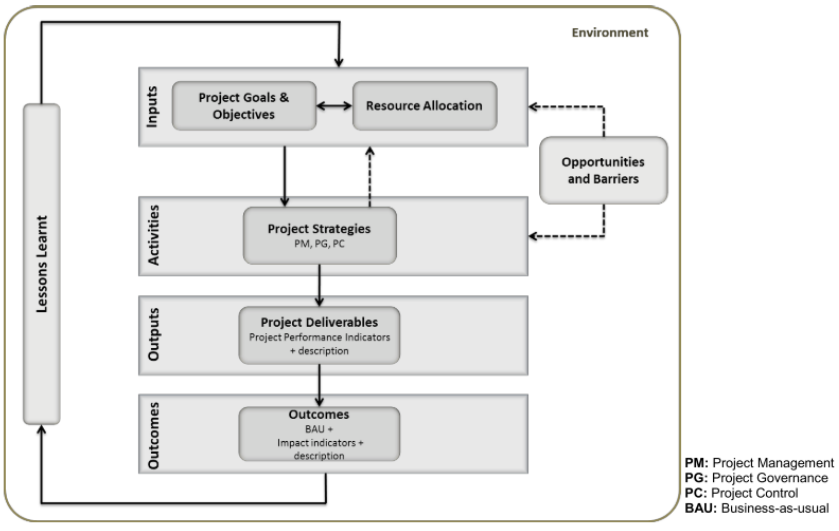


Figure 3: RepICT Model

Figure 2 RepICT Model (Awan, 2012)

Within the Design Cycle we will execute the main part of the research (Hevner Alan, 2007). We will analyze the open data from ICT Rijksoverheid. To improve the functional performance of the artifact that is The ICT dashboard of the central Dutch government, we will design a new artifact that will be a model within project R. This model will be a prototype for a new ICT Dashboard. The output of the model will be visualized with mockups. These can be found it Chapter 6.2. The research will start after the research proposal is finalized and approved.

3.2 Data mining methodology

As explained in the previous chapters, designing a new IT Dashboard for the Dutch government is a part of this research. The CRISP-DM was used to create the data mining elements, that is the ingestion and processing of the data. According to Wirth & Hipp (2000) *"The CRISP-DM (CRoss Industry Standard Process for Data Mining) project addressed parts of these problems by defining a process model which provides a framework for carrying out data"*. CRISP DM was developed in 1996 and was a European union project in 1997 (Chavez et al., 2020).

Even though CRISP DM was developed in 1996 it is still seen as one of the leading methods in the field of data science. CRISP DM became "de facto" standard for data science projects and still is the most widely used analytic methodology according to many opinion polls (Martinez-Plumed et al., 2021).

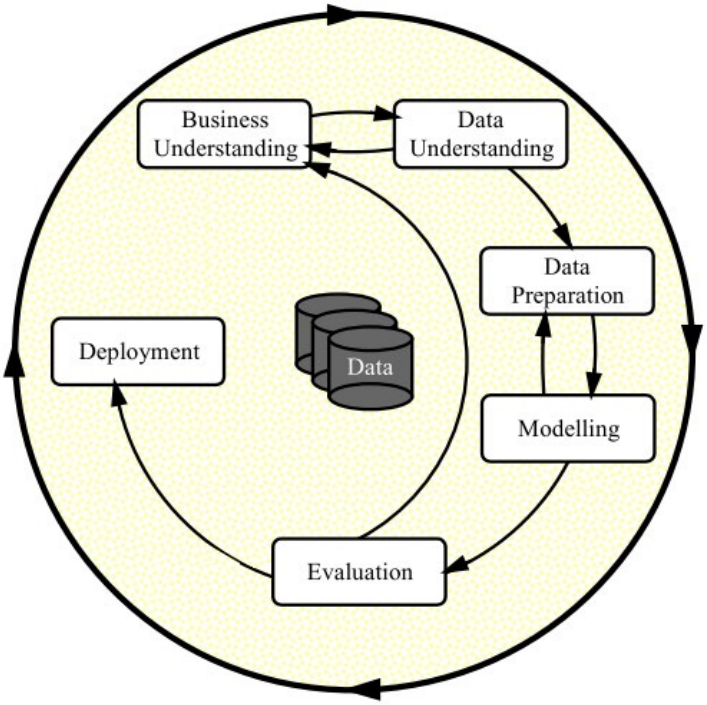


Figure 3 Phases of the Current CRISP-DM Process Model for Data Mining (Wirth & Hipp, 2000)

CRISP DM Process Model consist of six phases. These are shown in Figure 3. The *CRISP DM Process Model* is an overview of the life cycle of a data science project. The order of the phases is not strict. The arrows within in the model show only the important and frequent dependencies. The outer circle visualizes the continuously cycle of improvement of the data mining process (Wirth & Hipp, 2000).

Business understanding

The first phase of het process is *business understanding*. In this phase we establish the understanding of the problem at hand. The goal of this phase is to understand the objectives and requirements from the business perspective. This is then converted into a data mining problem definition and a project plan that is intended to facilitate the fulfilment of the objectives. This phase is well connected to the next phase, and it is iterative, so information sharing between these two phases will occur (Wirth & Hipp, 2000). As within this research we use the methodology of *Design Science*, this described part falls mainly under the *Relevance cycle*. However, a small part of the rigor cycle also takes place within this phase. This is mainly the establishing of the requirements through researching the other governmental dashboards within Chapter 4.

Data understanding

Data understanding phase is the second phase of the CRISP-DM. The first step within this phase is the data collection. After data collection the initial exploration of the data takes place to get familiar with the data. Within in this phase the first conceptions and hypotheses are made about the raw data (Wirth & Hipp, 2000). Within this research *data understanding* phase falls under the rigor cycle of *Design Science* methodology.

Data preparation

Within the *data preparation* phase all the steps are taken to construct the final dataset. This phase consists of multiple tasks that are not executed in any specific order. The tasks that are part of this phase are: creating a data frame or a table, attribute selection, data cleaning and transformation of the data for the modeling tools (Wirth & Hipp, 2000). These steps are concluded in Chapter 5. Within this research *data preparation* phase falls under the rigor cycle of *Design Science* methodology.

Modeling

Within in this phase modeling techniques are applied to the dataset. There are various ways and techniques of modeling the data. This phase is closely connected to the data preparation phase (Wirth & Hipp, 2000).

Within this research *modeling* phase falls under the design cycle of *Design Science* methodology. Data modeling finds place after the goal question metric and developing of the KPIs.

Evaluation

Within the *evaluation* phase the mock-ups are made. These will be then evaluated and reviewed by the researcher if they meet the requirements and if they visualize the KPIs in a sufficient way. Within in this phase it is important to evaluate if all the requirements are met

for the dashboard (Wirth & Hipp, 2000). Within this research *evaluation* phase falls under the design cycle of *Design Science* methodology.

Deployment

The *deployment* phase can be generating a report or be as complex as generating a repeatable data mining process (Wirth & Hipp, 2000). Within this research *deployment* phase falls under the design cycle of *Design Science* methodology.

Within this research the phase *business understanding* is mainly part of the relevance phase within *Design Science* methodology and can be found in Chapter 4 and Chapter 1. The *data understanding,* and *data preparation* are a part of the rigor cycle and can be found in Chapter 5. The *data modeling, evaluation and deployment phase* are part of the design cycle within the research methodology and can be found in Chapter 6. The *deployments* phase will also be the publication of the thesis.

We made the following illustration to show how CRISP-DM relates to design science. This illustration can be found in Figure 4.

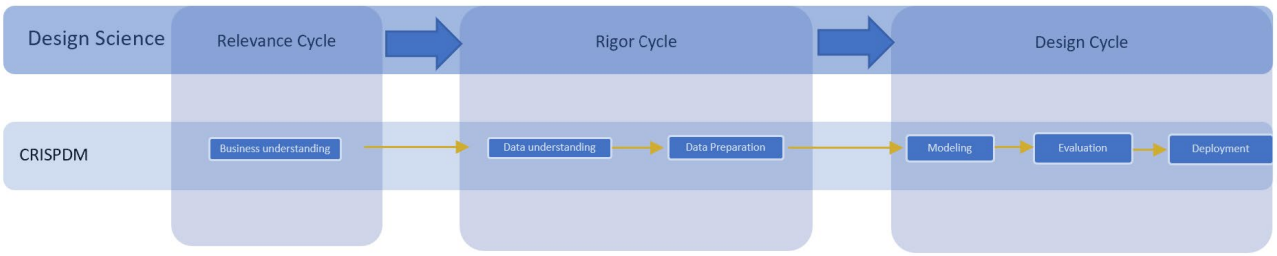


Figure 4 CRISP-DM relation to Design Science

3.3 Goal Question Metrics

To establish what metrics would be valuable to the Dutch government the *Goal Question Metrics* approach is used. According to Basili (1994) "The *Goal Question Metrics (GQM) method is based on the assumption that for an organization to measure in a purposeful way it must first specify the goals for itself and its projects, then it must trace those goals to the data that are intended to define those goals operationally, and finally provide a framework for interpreting the data with respect to the stated goals".*

This method is there to establish what informational needs there are within the organization. This information can be quantified when that is possible. This information can be analysed to whenever the goals are realized (Basili et al., 1994). The GQM is a methodical approach to integrate goals to models of the software processes. It is based on the needs of the project and the organization. The GQM method is a top-down approach, and it begins with defining a certain goal. This goal then gets refined into a question, then this question then gets defined

into a metric. According to Van Solingen & Berghout (1999) *"The questions within the Goal Question Metrics are there to break down the issues into its major components each question will then be refined into metrics. The results of the metrics should give the needed information to answer the Questions. The main principle of the GQM is that metrics should be goal oriented. In order to improve a process, the organization has to base their organizational goals on measurement goals. These goals need to be transformed into activities that can be measured during the project".*

Within this research the *Goal Question Metrics* is executed one step before creating the KPIs. As the GQM is very goal orientated, the metrics that get refined from this process should have the purpose to attain the main goals for the dashboard.

4. Dashboard Review

In this chapter IT Dashboards of different countries will be reviewed. The reason for doing this is to compare them and see where there are differences and where there are similarities, but also to discover features and design elements that could be transferred to the newly designed Dutch dashboard. Another reason for doing this review is to ensure that the new dashboard that is designed within this research will be innovative. We will attempt to improve the governmental dashboards that have been reviewed.

We found three countries that have a governmental IT project dashboard.
These are the following countries:

USA

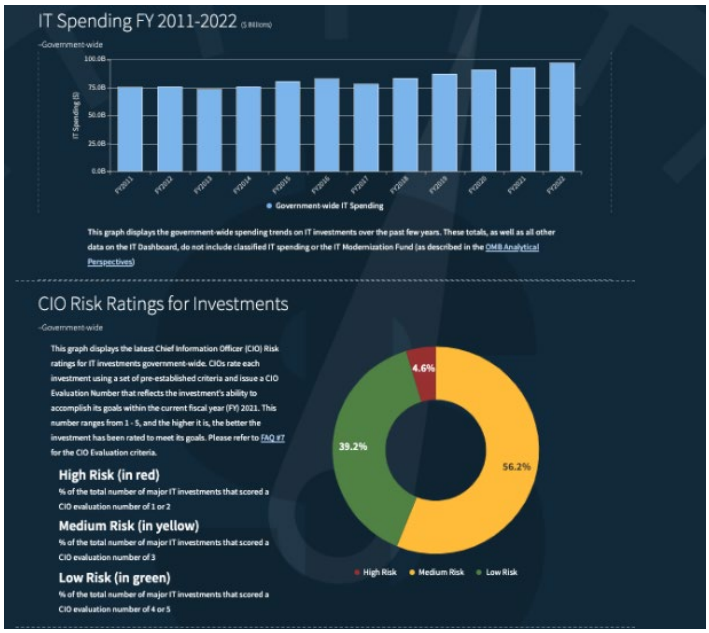


Figure 5 IT Dashboard of the government of the USA (https://itdashboard.gov/)

Australia

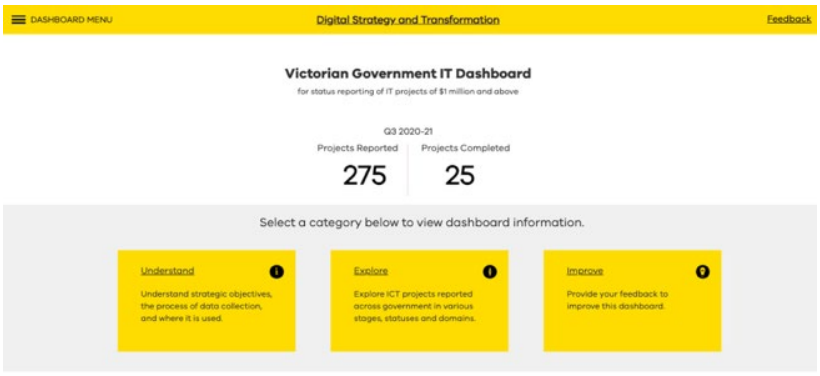


Figure 6 IT Dashboard of the government of the Australia (https://itdashboard.digital.vic.gov.au/#!/)

The Netherlands

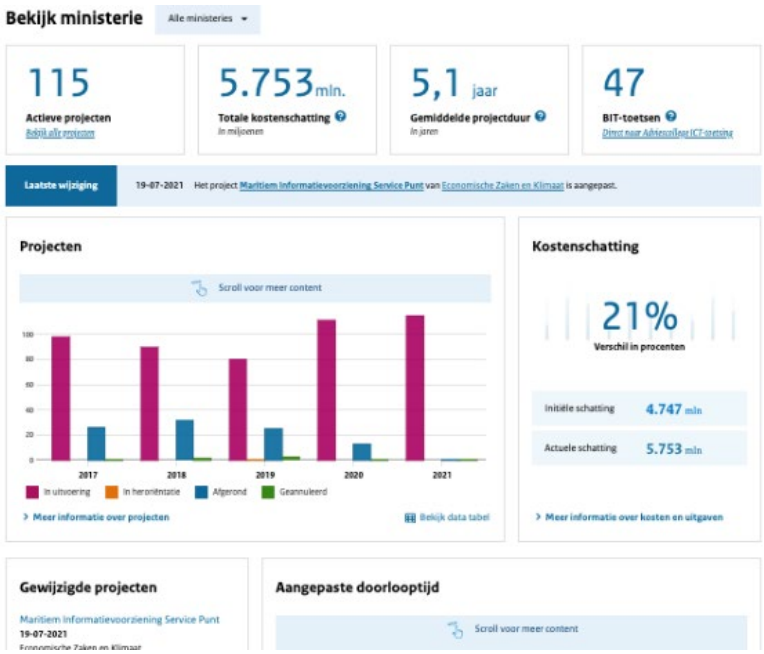


Figure 7 IT Dashboard of the government of the Netherlands (https://www.rijksictdashboard.nl/)

The dashboards have been reviewed by using the method of Sarikaya et al. (2019). The first step in this method is to find out what the purpose is of the dashboards. Governmental Dashboard have all the same purpose: communication and learning. According to Sarikaya et al. (2019). *"These dashboards exist to communicate or educate the reader, who may lack the context surrounding the presented data."*

The Governmental Dashboards are there to inform and communicate to the stakeholders about the IT Projects of the governments.

The following step within the method is establishing the audience for the dashboard. This is done within three sections: *Circulation, Required visualization literacy, Required advanced domain expertise*.

Circulation

The method speaks of 4 types of audiences: public, social, organizational, individual (Sarikaya et al., 2019) .

The governmental audience of the dashboards is public; however, it is also organizational. The reason why the dashboards have two types of audience is because it has multiple types of stakeholders. It is a publicly available dashboard for everyone to view, however it is also used by the government to control the use of governmental spending on IT Projects.

Required Visualization Literacy

The method speaks of three levels of complexity in visualization of the dashboard: Low, Medium, and High. Low literacy is basic visualization within the dashboard that are not complex for example: bar and line charts. Medium literacy adds combined dual axes, scatterplots for example. High literacy adds for example: radar, network visualizations (Sarikaya et al., 2019).

The visualization literacy of the governmental dashboards is low. Within the dashboards only basic visualization was used for example bar charts and pie charts. There is no medium or high literacy needed to understand the dashboard.

Required advanced domain expertise

This segment is about if the stakeholders need to have knowledge over the business data or that it is data that is general to understand for the public (Sarikaya et al., 2019).

For the governmental dashboards there is no advanced domain expertise needed.

The following section is about visual features. According to Sarikaya et al (2019) *"Tools may allow a user to design (or customize) the dashboard; they may allow faceting of the data through data filters and slicers; and they may allow modify- ing the state of the data and world based on the data presented within the dashboard".*

Construction and composition

This part is about if the dashboards allow the user to modify the composition and the construction of views (Sarikaya et al., 2019).

This where the dashboards start to differ. The USA governmental dashboard does not have this functionality. However, the Dutch and the Australian dashboards do allow this. Where the Australian dashboard has the most possibility to modify the composition and the construction of views.

Multipage

This segment is about if the data of the dashboard is all on one page or that the dashboards data is divided over multiple pages (Sarikaya et al., 2019). The governmental dashboards are all multipage dashboards.

Interactive interface

This subject is about whether the dashboard allows the user to observe different views. If the data can be filtered, sliced and cross-highlighting. Also, if it is possible to drill up and down within the data. It is about how the user can interact with the dashboard to get to the data that they seek (Sarikaya et al., 2019).

All the governmental dashboards have an interactive interface. However, all dashboards have a different approach how they can be interacted with. The IT dashboard of the USA enables the user to drill down from department to a specific investment. This is a bit limited and makes it difficult to get a good overview of all the projects of a department. The IT dashboard of the Netherlands gives the user freedom to get an overview of all the projects, filter between projects and drill down to individual projects. In some cases, it also gives the user the possibility to get more information about a subject. When drilling down to an individual project we observed that there were missing data points. The IT dashboard of Australia lets the user freely interact with almost every aspect of the dashboard. It is possible to drill down on all the projects or to filter between departments. It is possible to drill down from all projects to one individual project. We found that the Australian dashboard was the most interactive of all dashboards.

Highlighting & annotating

This subject is about whether the dashboards allow the user to highlight parts of the data or allow them to add notes to the data. These won't change the data of the dashboards but allows the user to view these highlights or annotating in the future (Sarikaya et al., 2019). None of the governmental dashboards lets the user highlight or annotate any subsets of the data.

Modify Data or the world

The following section of the method goes in to if the dashboard allows the audience to change the data or control the external state of the world (Sarikaya et al., 2019). None of the governmental dashboards allow the users to alter the data or the external date of the world. Within the next section the method goes into extra data semantics. Dashboards can allow the user to set actions for example: alerting, benchmarks, or update.

Alerting notifications

The following part is about if the dashboard gives the user warnings when outliers occur or a certain threshold is reached (Sarikaya et al., 2019). As far as we know none of the governmental dashboards sends out notifications or an alert.

Benchmarks

The following subject is about if the dashboards display benchmarks in the form of indications of defined Thresholds (Sarikaya et al., 2019). The three governmental dashboards do give some form of benchmarks. The Australian and the USA dashboard do that with percentage thresholds of the increase of the costs or duration of IT Projects. These are color coded in the colors green, amber and red. The Dutch Dashboard does it with color-coding when the the run through time of the project is changed. Their thresholds are; earlier, unchanged or shorter than a year, or longer then a year.

Updateable

This following part is about if the dashboard is updated regularly and the data gets automatically refreshed (Sarikaya et al., 2019). All three governmental dashboards are regularly updated with new data, however these updates have a low frequency.

4.1 Encoded Dashboard

We encoded the three governmental dashboards in the same way as Sarikaya et al (2019) this can be viewed in Table 2.

			Purpose				Audience			Visual Features					Data Semantics		
goal	Cluster	Country	Strategic	tactical	operational	Learning	audience	Vis literacy	Domain expertise	construction	Interactivity	Modify data/World	Highlighting	Multipage	Alerting+notifications	Benchmarks	Updateable
Decision making/ Awareness	Strategic Decision making/ Static organizational	USA	y	y	N	N	P/O	L	N	N	N	N	N	Y	N	N	Y
Decision making/ Awareness	Strategic Decision making/ Static organizational	Australia	y	y	N	N	P/O	L	N	Y	N	N	N	Y	N	N	Y
Decision making/ Awareness	Strategic Decision making/ Static organizational	Netherlands	y	y	N	N	P/O	L	N	y	N	N	N	Y	N	N	Y

Table 1 Encoded governmental dashboard

We found that within the coding of the different dashboard there are not many differences. However, by using the method we reviewed all the dashboards on different aspects. This showed us that even though the dashboards are very similar in their goal and functionalities, they do visualize their data in different ways. This information that we gathered will be useful in the future phases and with the result to ensure that the dashboards meet the same requirements the reviewed dashboards had to meet.

4.2 Findings

Within in this section of the analysis we will go into our findings of the three governmental dashboards. We found that the dashboards shared a lot of similarity, and they all share a common goal. This is very visible within Chapter 2.1 . We mentioned in this chapter that where they differ the most, is in how they visualize the data.

Because the similarity of the three governmental dashboards, we can determine some goals and requirements of what is needed within the new dashboard of this research. One of these requirements is the goal. The governmental dashboards all share two goals. On the one hand it is decision making for the politicians, and on the other hand it is creating awareness for the general public. One of the main requirements of the new dashboard is that it can deliver information in a clear and visual manner to the two different stakeholders. One of the requirements of the dashboard mentioned in Chapter 1 is that the dashboards need to service the different stakeholders. With these requirements comes two aspects: the visual literacy

needed to comprehend the information within the dashboard, and the domain expertise that is needed to understand the information that is displayed within the dashboard. The required visual literacy needs to be low so that the dashboard is understandable for the general public. The domain expertise should not be needed so that all stakeholders can get and understand the information within the dashboards. This will also be one of the requirements of the dashboard that will be designed with in this research. In one of the aspects that the dashboards differentiated from each other was the construction and composition. The dashboards of the government of the USA did not have this functionality, where the Dutch and Australian had. However, the Australian dashboard had far more possibilities to view the information. This makes it easier for the end user to make comparisons.

5. Data exploration and preparation - Analysis of data from existing dashboard

Within this chapter the raw data from the existing dashboard is analysed. The reason for this analysis is to find what information is already available within the raw data of the dashboard and what information can be derived from this data. This is done through the first two phases: the data exploration phase and the data preparation phase. These are phases from the method CRISP-DM that is introduced in Section 3.2 Research Methodology. Within this methodology these phases are iterative and closely connected to each other. There is a lot of information sharing and improvement cycles within these two phases. Within these phases the raw data analysis will be done. With the findings of this analysis requirements will be established. The findings will be presented in Section 5.3. with all the requirements a GQM is created in Chapter 5.5 .

5.1 Data understanding

Within the data understanding phase, we will collect the data and get familiar with the data from the Dutch government. The data understanding phase consists of a few subjects: retrieval of the data, exploration of the data, assessing the data quality. Within the following sections this will be executed.

5.1.1. Data retrieval

In this step the data retrieval step which is part of the data understanding phase the objective is to collect the raw data that will be used within the dashboard. The data of the Dutch governmental IT dashboard is in accordance of the WOB. The explanation of the WOB can be found in the conceptual foundations. This raw data Is available from the website: https://www.rijksictdashboard.nl. The data was downloaded on the 23rd of May 2021.

The data is available in two different file formats these are XML and JSON. Within this research the JSON file format is used. This data was loaded into R Studio. R Studio is a free data science framework that works in the R language. To be able to load the JSON within R there was a library used and syntax needed. The code used can be found in the appendix of this paper, see Appendix II. This concluded the data retrieval step.

5.1.2. Exploration of the data

In this section we discuss exploration of the data. Within the data exploration phase. The goal is to get to understand the data set. During this phase we found how the information was stored within the data base. This is important for the next phase within the CRISPDM method. Within the data exploration phase a part of the analysis will be done. The reason that a part of the analysis is done within these phases is because all the data is first handled within this phase. To be able to complete the data cleaning correctly the data must be understood first.

We had the following observations while executing the exploration of the data. The initial observation of the data was that the raw form is saved in different lists. This made it very unreadable because it made it harder to see all the information of one project. As was explained in chapter Research Methodology, the phases within CRISPDM are iterative. So, there is no issue in switching between the different phases. To make the data easier to understand, we did a part of the preparation phase and flattened the data. How this is done is explained within Section 5.2 Data Preparation phase. The following observation we made is that data had a lot of empty fields. This could be an issue if this data is needed to do some of the analysis. The last observation we had is that the data fields were not al formatted in the same way. For example, the yearly maintenance costs were sometimes saved as text and the other times as a value or a combination of both.

5.2 Data Preparation phase

The data preparation phase is the phase before the modeling phase. Within this phase the following steps will be taken; data cleaning, creating tables or data frames, construction of new attributes, transformation of the data. So that the data is prepared for the next phase; the data modeling phase.

5.2.1 Data cleaning

According to Ilyas & Chu (2019) the quality of the data is one of the most important obstacles in data management. Dirty data leads to inaccurate data analytics results which then leads to incorrect business decisions. Therefore, data cleaning is a very important phase in the building of a database.

For the cleaning of the data collected from the current IT Dashboard of the Dutch government, a three-step data cleaning framework was used similar to Jan Van den Broeck & Jonathan R Brestoff (2013), see Figure 8.

In Section 5.3 the recommendations of the report made based on these steps, will be elaborated.

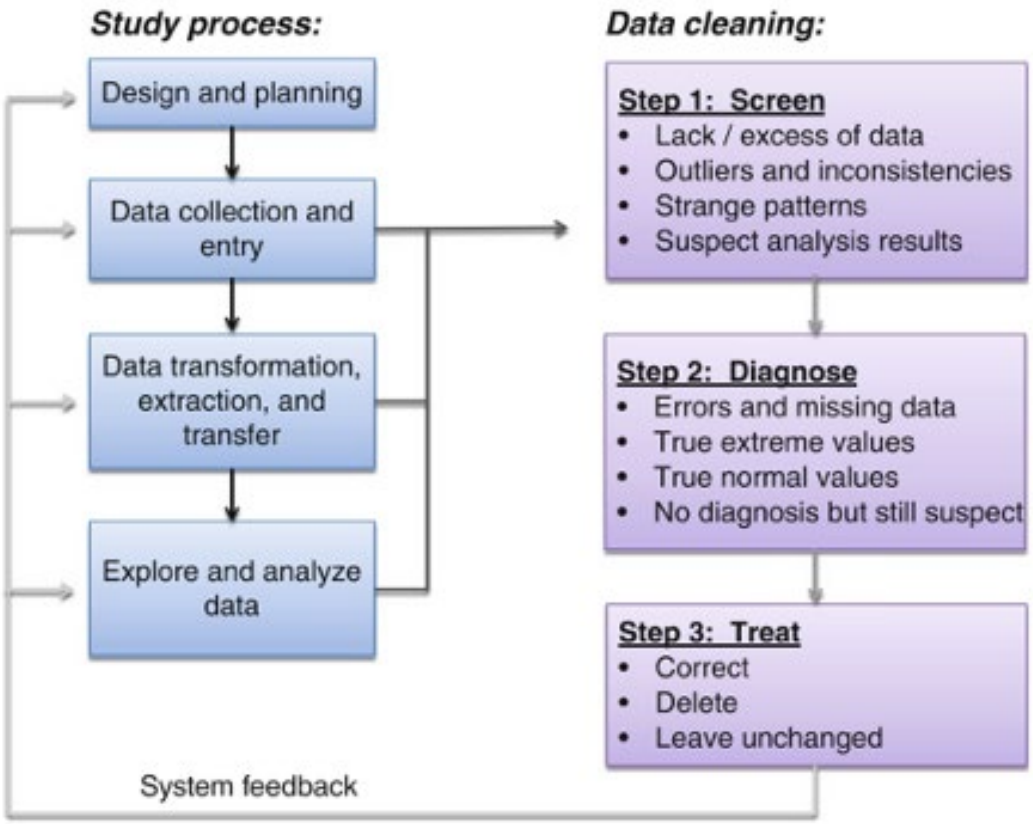


Figure 8: Data cleaning Process (Jan Van den Broeck & Jonathan R Brestoff, 2013)

The first step of this framework is screen. In this step the data gets analysed on the surface level (Jan Van den Broeck & Jonathan R Brestoff, 2013). This step was for the biggest part completed in the data understanding phase. The findings of this phase were as follows: the data was not stored in a practical form. The data was stored in multiple lists, which did not make it accessible. Furthermore, what was established in this phase was that there were a lot of missing data points. With all the missing data points it is difficult to get accurate data analytics, which has said before, can lead to the wrong business decisions or in this case a wrong governmental decision.

The next step of the data cleaning framework is the *diagnose*. During the analysis of the data there were a lot of inconsistencies found. This should be improved for the future so that the stakeholders are able to get access to the correct data that is needed to inspect the IT projects. What was found within the data was that there were a lot of missing data points or that data points were formulated not in a uniform metric. This makes the data very inconsistent. For example, within the project the data of the maintenance cost was not complete or ambiguous for most of the projects.

The last step of the data cleaning method is the *treat*. In this step the decisions are made if data needs to be deleted, corrected, or left unchanged. The decision making on whether the

data was important or not, was based on the previous step of data cleaning: the diagnosis phase. Within the data there was not much data that needed to be deleted. It was more data that was not needed for the analysis. This data was excluded out the data frames in the following phase: the data transformation phase. This can be found in Chapter 5.2.2 Data transformation. The more important problem of the data was data points that were missing. However, this was not data that could be filled in, so this had to be left unchanged. The data set had a lot of data missing. Because of the missing data, some analysis cannot be done for all the projects. To audit the projects, it is important that the stakeholders have all the information about the IT Projects.

5.2.2 Data transformation

Within this step we started to transform the data. The data transformation phase is an important phase before the data modeling phase. Within in this phase the data will be transformed so that it can be modeled and used within this dashboard. This phase is an iterative phase and closely connected to the modeling phase. So, when needed we would come back to this phase to enable the needed transformation of the data for the modeling. Within this chapter we will go through the actions that have been taken during this phase. This is not the chronological order of these action because some actions had more iterations. What was done in this phase was: flattening, creating data frames, mutating data types, calculations, combining tables and data frames.

Flattening

As mentioned in the earlier chapters the raw data loaded within R studio was saved in a list containing multiple nested lists. Every project had the same data fields and lists. Lists are hierarchically saved data points what makes it difficult to call on the right fields.

We made a visualisation of how the data was formatted. The first level of the hierarchy was the projects. There were 313 entries. The next level was the project numbers. Every project number had 31 entries. Within the 31 entries there were six entries with their own list, one of them is the “herijkingen”. Within each herijking the entry type was written out in lines separated by quotation marks. To illustrate how this data was presented, we made an example that can be found below:

[Projects] (313)
- [Project 1] (31)
  - Name
  - Initial Cost
  - [Herijkingen] (9)
    - ‘Date1”Date2”Date3’
    - ‘Cost1”Cost2”Cost3’
    - ‘end Date1” end Date2’ ‘end Date3’
- [Project 2] (31)
- [Project 3] (31)

To excess these lists, we needed to flatten them. Flattening means we needed to remove the hierarchy within in the list. We did the flattening by executing a lapply function. A lapply function has the following effect on the data; it loops over the list and applies the functions over all the elements within this list. By doing this, all the hierarchy saved data or list was transformed into one table. We flattened the "herijkingen" after flattening the data we started creating the data frame with the metrics that we needed to calculate the KPIs. This will be further touched upon in the segment creating data frame.

Creating data frames

After flattening the data, it made it possible to start creating data frames. When creating the data frames, it is important that the phases before were done correctly so that the required data for the modeling phase gets implemented within the data frame. However, if some of the required data was missed, we went back to add this to the data frame. When we completed the main data frame, we filtered per ministry and gave them separate data frames. This way we could get a better view on the IT Projects for every governmental department.

Mutating data types

To complete some of the calculations that are planned during the modeling phase it was important that the data would get the right attribute to achieve that we needed to code that in. For example, dates within the data frame should be recognized as a date within R studio. We also needed to give values within the data frame the numeric attribute so that calculations could be made within R studio.

Creating new fields

For some of the KPIs we needed additional metrics to be calculated so that we could display them within graphs and models. One of the metrics is *number of workdays*: workdays is built from start date to end date. What the function does is calculating the number of workdays between the two dates. This function was needed to calculate the burn-rate in a later phase.

Combining tables and data frames

Within the raw data of the Dutch government there was a lot of data that we did not need for the modeling. We created a data frame with the information that we were going to use for the modeling of the dashboard. There were a few problems we had to work around. One of the problems was that the data from the list "herijkingen" needed to be combined with the project data. Every time a herijking is made within the data of the Dutch government, this is saved in a separate list. So, if the project would have four "herijkingen", there would be four entries within that list saved under "herijkingen". For the analysis we needed the data from the "herijking" within our data frame. In the way that the data of "herijkinen" was originally

formatted in, made it impossible to recall it. The reason for it is that the “herijkingen” was not a unique data entry.
We presented an example in a subsection of Chapter 5.2.2 Flattening, to show how the data was formatted. For every project herijkingen was saved as followed:

- [Herijkingen] (9)
  - ‘Date’ ’Date2’ ’Date3’
  - ‘Cost1’ ’Cost2’ ’Cost3’
  - ‘end Date1’ ‘end Date2’ ‘end Date3’

To combine this to the project data, it was of importance that we could find the corresponding information. To connect this data, we needed to give all the data of the “herijkingen” the project numbers, sothat we could merge these two unique keys into one data frame. We also wanted to recall the last entry for every herijking. After combining these data entries we had all the necessary data within one data frame and we could start the calculations. An example of how the created data frame was visualised can be found below:

Project number	Name	Initial Cost	“Herijking” Cost	“herijking” End Date
1	Project name 1	Cost of project 1	Cost 3	End Date 3
2	Project name 2	Cost of project 2	Cost6	End Date 6
3	Project name 3	Cost of project 3	Cost 2	End Date 2

5.3 Findings from data understanding phase and data preparation phase

During the data understanding phase and the data preparation phase the raw data of the Dutch government was analysed. Within this analysis we had a couple of findings for improvement.

There were a lot of missing datapoints on how much it would cost to maintain the IT application. For the datapoints that were filled in there seems to be no guidelines on how this should have been filled out. Some datapoints were text only and some data points were text combined with some numerical values. To get the information that is needed the cell should be filled out with a numerical value and a time indication for example per year or per month.

Overall, within the data there were also a lot of data points missing. As mentioned earlier the importance of accurate data for precise analyses is imperative. To get this data, it would be valuable if there were some required fields that are important to get the metrics that is needed to calculate the KPIs.

We found that the run through time of the project and the budget is often changed without clear reasoning. For transparency this would be better if the explanation would be more elaborated than for example what happened in project ”300674” and the reasoning was “setbacks in the project”. This information does not seem sufficient for a project that is extended for 5 years, and the budget went up from an estimated cost of 1.5 million to an estimated cost of 27.14 million. Within the data we found more similar cases within the IT projects of the Dutch government.

There are also some additional data points that we would like to get from the projects. For example: The reason why a project will cost more than originally estimated. Within the data there is often no good stated reason for the project going over the first estimated budget.

It is important to follow the project after it is finished so that the benefits can be more clearly identified.

There are a few subjects that can be followed by the finalization of the project. The fact if the expected lifetime of the project is met, keeping track of the yearly maintenance cost and where possible quantifying the benefits the system has had. It is recommended to do this so that the balance can be made and that it is possible to learn from past projects.

A question that we have after going through the data is whether there are go and no-go moments. When a project is extended for so long and the budget is far exceeding the initial estimation is there a point in time where the project is stopped. These findings will be looked at within the GQM and will then result in KPIs.

5.4 Requirements

Within in this chapter the total list of requirements that we derived from the findings will be established. These requirements will be combined with the requirements given by the Dutch government by R.W. Knops. These requirements will form the basis of the Goal Question Metrics that will then be the basis for the KPIs. Within the letter there were three changes the implementing within the landing page of the dashboard these were the following to R.W. Knops:

1 The total number of big IT projects
The requirement is to give the stakeholders an overview of all the IT projects that are being implemented by the Dutch government
2 The current cost estimates
The requirement gives the stakeholder the information about the cost of IT projects
3 The average run through time of projects
The requirement gives the stakeholders an overview on what the average run through time is of a project. With this information the stakeholder can estimate when other projects will be done
(R.W. Knops, 2020a)

Also, the "Adviescollege ICT-Toetsing" gave the following points of improvement which form the additional requirements:

4 Make the intended benefits concrete and aim to achieve their realization
5 Think carefully about future processes
6 Increase focus on client-supplier relationship
7 When choosing a solution direction, also take the management and maintenance phase into account
(Adviescollege ICT-Toetsing, 2022)

These are requirements that are already fulfilled within the current ICT dashboard and should also be a part of the newly designed dashboard.

The requirements that we established from findings of the different governmental dashboards are as follows:

1. The governmental dashboard has two goals: decision making and creating awareness for the general public
    a. Visual Literacy requirements need to be low so that is understandable for all the stakeholders
    b. Domain expertise should not be required
2. The stakeholder should be able to edit the construction and the composition.
3. The stakeholders should have the ability to drill down from overview of the total projects to one specific project.

The requirements that we established from findings by analysing the governmental data are as follows:

1. The number of workdays between start and end date should be calculated and presented within the dashboard

2. Guidelines that are of importance should be presented within the fields of the dashboard
3. To calculate and present the percentage of slippage between “herijkingen”

These are the requirements that we established for the research. Within in the Goal Question Metrics these requirements will be used as foundation for the KPIs.

5.5 Goal Question Metric

Within this section we will use the findings of the analyses that resulted in the requirements of the new dashboard. With this information, we will create a Goal-Question-Metric hierarchy. This GQM hierarchy will then be refined into KPIs. These KPIs can be found in Chapter 5.6 KPIs.

Within CRISP-DM the GQM analysis falls in the modeling phase and partly in the phase of evaluation. The method has a top-down approach, and it works from the goal that you want to achieve within the dashboard and not from the metrics that is available within the data. This is also stated in Chapter 3.3 Goal Question Metrics. What was concluded after using the GQM approach is the goal that the stakeholders have, the questions that need to be asked to get the goal, and the metrics that are needed to fulfil the information needs for them.

In our research, we take as starting point a GQM analysis by Visser (2013), which was informally communicated in a presentation. Figure 9 shows the relevant slide from that presentation, made by Visser (2013). Within this slide a model of the Goal Question Metrics can be found. This model will be used and redesigned within our research. The model can be found in Figure 10.

Within the current dashboard of the Dutch government, the focus is only on the create phase. By using the CSB we are looking at the full life cycle of the applications that are created, so that the cost of sustaining the created systems can be taken in account. This can then be balanced against the benefits the application will bring to the Dutch government.

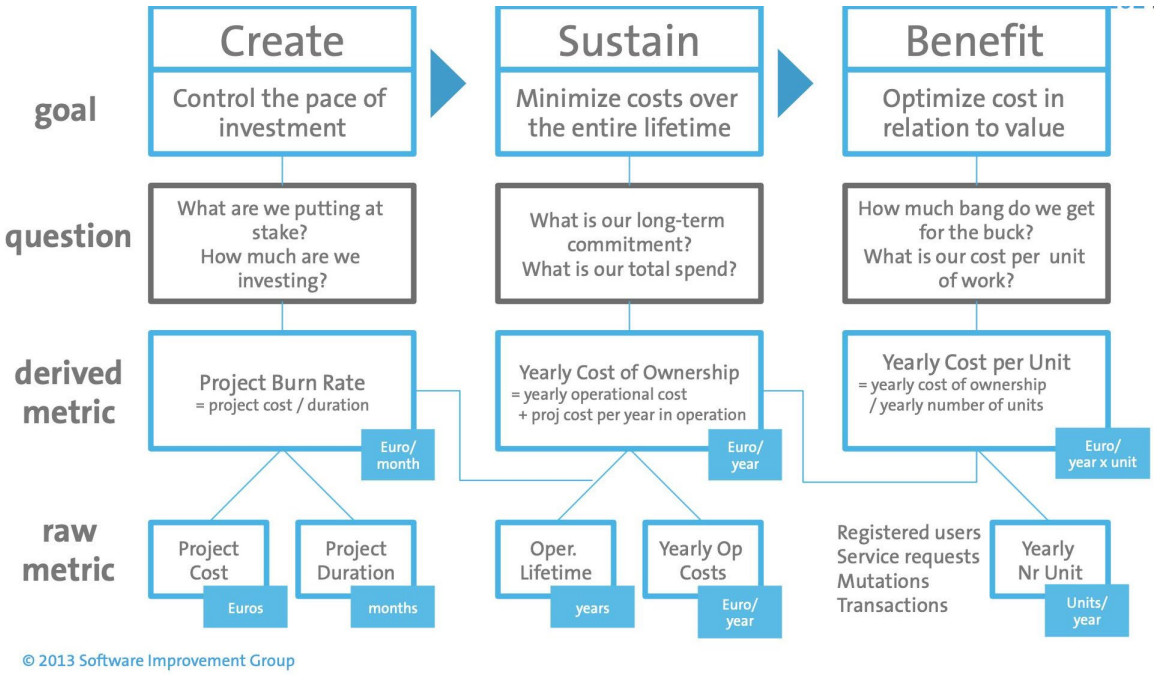


Figure 9 CSB Model by Visser (2013)

The model takes as basis the view that software projects have three aspects: (1) during the project itself, a software system is created, (2) after the project has delivered a software system, this system needs to be sustained by operating and maintaining it, and (3) value is created with the software systems by giving certain benefits to various stakeholders. CSB stands for *Create, Sustain, Benefit*. The *create*, the *sustain* and the *benefit* part. However, there are some limitations to the model. One of these limitations is that without the data that is needed to get the derived metrics the goal can't be achieved.

These phases all have different goals. Within this research this model has been redesigned to fit all the earlier established requirements and can be found in Figure 9. In the sections below we will go through the three segments of the CSB model.

5.5.1 Create

Within this section we complete the Create Segment of the CSB that applies to this research. We go through the following subjects: Goal, Question, Derived Metric and Raw metric needed. This section is an addition to the CSB model made by Visser (2013).

Goal

The original model of Visser (2013) established the following goal: to control the pace of investment of governmental IT projects. Within our research we found that this goal is

important for the governmental dashboard. That goals aids to achieve some of the requirements: the current cost estimate and decreasing the percentage of slippage.

However, we found that an additional goal was needed to meet the requirements: to detect rising of cost and duration of the IT Projects. During the analysis of the raw data, we found that there seems to be a lot of increases within the cost and the duration of the projects.

Questions

The original model established the following questions: What are we putting at stake? How much are we investing?
The question that we added is: Does excessive slippage occur?
During the analysis of the data of the governmental IT dashboard we found that in a lot of projects slippage occurs frequently. With the answer to this question, we can grasp how much slippage occurs during the execution of the IT governmental projects.

Derived metrics

Derived metrics can be formulated utilizing the metrics existing within the data base. To answer the questions stated in the previous segment, the following derived metrics is needed: project burn rate and project slippage (burn rate, project cost, project duration).

Raw metrics

Raw metrics can be found within the data base. These metrics are needed to create the derived metrics. We found the following raw metrics within the IT governmental dashboard data base: project “herijking” cost and project “herijking” end date.

5.5.2 Sustain

Within this section we complete the Sustain Segment of the CSB that applies to this research. We go through the following subjects: Goal, Question, Derived Metric and Raw metric needed. This section is an addition to the original CSB model.

Goal

The original model established the following goal: minimize cost over the entire lifetime. We found that this goal is sufficient to meet the following requirement of the dashboard: the current cost estimate.

Questions

The original model established the following questions: What is the expected periodic maintenance needed? What is the total estimated downtime within its lifetime? How many errors are allowed for the application to stay operational. What is the cost of the downtime? By answering these questions, we can find out what the operational costs are for the IT governmental projects and the costs of ownership.

Derived metric

Derived metrics can be formulated utilizing the metrics existing within the data base. To answer the questions stated in the previous segment, the following derived metrics is needed: yearly cost of ownership.

Raw metric needed.

Raw metrics can be found within the data base. These metrics are needed to create the derived metrics. We found the following raw metrics within the IT governmental dashboard data base: operational lifetime in years and yearly operational costs.

5.5.3 Benefit

Within this section we complete the Benefit Segment of the CSB that applies to this research. We go through the following subjects: Goal, Question, Derived Metric and Raw metric needed. This section is an addition to the original CSB model.

Goal

The original model established the following goal: personal cost and time saved. However, we found that an additional goal was needed to meet the requirements: actual significance/value of the application. While analysing the data we found that the potential benefits of the IT projects were not always identified. Identifying the potential benefits is important to see if they final goals are achieved. These projects can have financial, non-financial or both benefits. We found that these goals are sufficient to meet the following requirement of the dashboard: make the intended benefits concrete and aim to achieve their realization.

Questions

The original model established the following questions: How much bang do we get for the buck? What is our cost per unit of work?

The question that we added is: How many mutations will this application do in its lifetime? How much time is saved by this application? What resources are saved by this application

being in operational? How much money is saved on employee cost? What is the benefit of the application financial or non-financial?
During the analysis of the IT governmental dashboard, we found that the financial and non-financial benefits are not given. By answering these questions, we can find out what the financial and non-financial benefits are of the governmental IT projects.

Derived metric

Derived metrics can be formulated utilizing the metrics existing within the data base. To answer the questions stated in the previous segment, the following derived metrics is needed: yearly cost per unit, the financial benefit of the application, the non-financial benefit of the application.

Raw metric needed.

Raw metrics can be found within the data base. These metrics are needed to create the derived metrics. We found the following raw metrics within the IT governmental dashboard data base: the financial or non-financial benefit.

After completing the three sections of the CSB model, we redesigned this model in Figure 10.

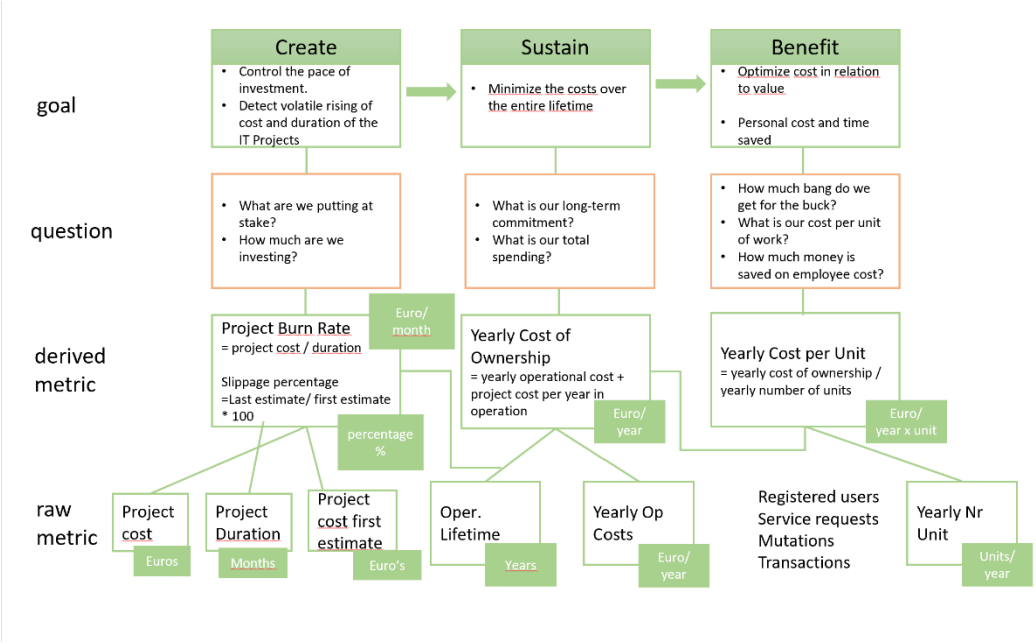


Figure 10 Redesign CSB Model by J.L. Chin A Loi

5.6 KPIs

Within the GQM metrics we established the goals the questions and the needed metrics. With this information we created new KPIs that would accomplish these goals.
Within in this research we differentiate between metrics and KPIs. We see metrics as datapoints and KPIs are tied to one or multiple business goals or requirements.
Within this chapter we will divide the KPIs in the three different segments of the CSB.
We do this to show how every goal is met with the set of KPIs. The next step is to check if the KPIs meet all the requirements.

5.6.1 Create

Before entering the create phase of the KPIs, we need to emphasize that the "Bureau Adviescollege ICT toetsing" advised in their report from 2021 upon the business cases, prior to the IT projects initiation. According to the "Bureau Adviescollege ICT toetsing" the costs and the benefits were not elaborated enough within the previous business cases. This means for the IT projects that the intended result could differ from the result achieved.
These requirements are not within the scope of our research but have consequences to the IT projects. The requirements that "Bureau Adviescollege ICT toetsing" gave are as follows:

1. Think carefully about future processes
2. Increase focus on client-supplier relationship

For the create section of the GQM we had the following goals: to control the pace of investment and detect rising of cost and duration of the IT Projects.

The KPIs that we created to achieve these goals are:

Project burn rate for each workday
With this KPI the stakeholders can see how much is being spend every day while the project is in progress

Total number of projects and their status
The number of projects in total and partitioned by status where status can be the following:

- "Afgerond" (Finished)
- "Geanuleerd" (Canceled)
- "In uitvoering" (In progress)
- "Niet gestart" (Not started yet)

This KPI is important to give the stakeholders a quick overview of all the projects that are active or inactive within the Dutch government. This was also one of the requirements that was already implemented by the Dutch government.

Percentile difference of the total cost, the duration and the burnrate at the beginning of the project and the last "herijking"
This KPI can give the stakeholders a good idea of what is happening with the cost and the duration the of the project. With IT projects it is important to keep control over the slippage of the project. This information could be expanded on with the reasoning for the project slippage to happen. The hypothesis is that when the cost increases the duration of the project also increase with almost the same percentage. This is because if the burn rate stays the same this should be te conclusion.

Color indication if the percentile absolute difference is above the thresholds that are set
This color indication will be used on the percentile difference of the slippage
These thresholds are set to the same percentage as the dashboard of the Australian government's dashboard.

- Green 10%
- Amber 10- 30%
- Red 30% and above

The explanation for the slippage
Within the current IT Dashboard there is an explanation given why slippage occurs, however this could be elaborated in a more concise explanation. Now these explanations are often vague, for example: *the scope of the project changed*. With additional information about the reasoning for the slippage, it will be easier for the stakeholders to understand what is happening within the IT projects of the Dutch government.

Color indicators in which period the challenges occur
This can give the stakeholders a quick view of when the project changed from green to amber

or red. This with a brief explanation of what happened can give the stakeholders the control they need to intervene if this is needed.

5.6.2 Sustain

For the sustain section of the GQM we had the following goal: minimize cost over the entire lifetime.

This connects to the advice that the "Bureau Adviescollege ICT toetsing" had given in their report of 2021; "When choosing a solution direction, also take the management and maintenance phase into account". This advice can be translated into the fact that it is important to think of the "Sustain phase" while decision making around the IT projects within the "Create phase" of the Create, Sustain and Benefit model (Adviescollege ICT-Toetsing, 2022).

The KPIs that we created to achieve the goal "minimize cost over the entire lifetime" are:

The yearly operational cost
We found that within the data of the Dutch government this was not always a clear value. However, this is a very important KPI for the cost sustain and benefit analyses. This KPI can be important for the stakeholders when the total cost outweighs the benefit of the IT project. What than can be used to decide to go through with the project or cancel it. This KPI should be monitored even when the project is active to present what the actual cost is to sustain the IT applications. If the difference of the actual operational cost and the estimate operational cost is large the data can be used to make better estimates of future projects.

The yearly cost per unit
The yearly cost of unit can also be seen as the yearly cost for every mutation. For this KPI to be available within the dashboard Data points need to be added.
This KPI can give a warning when the cost per unit goes up or down.

5.6.3 Benefit

For the benefit section of the GQM we had the following goal: personal cost and time saved and actual significance/value of the application.

This connects to the advice that the "Bureau Adviescollege ICT toetsing" had given in their report of 2021. "Make the intended benefits concrete and aim to achieve their realization". This can be translated into the fact that it is important to establish in advance what the added value is going to be of the ongoing IT projects.
The KPIs that we created to achieve these goals are:

The cost saving for every mutation done within the application
What the goal is of the project and what problem is it solving?
Is the project achieving the established goals and what are the chances of it achieving the goals? This quantified by percentage of the probability of it achieving the set goals.

6. Dashboard Development

Within this chapter we are going to put all the KPIs into a model, to visualize them within the dashboard. This is done after looking at all the requirements of the previous dashboards that were then used to define KPIs using the GQM method. Also, in this chapter we give priority to how the dashboard is perceived and operated, what the possibilities are for the stakeholders and how they can interact with the dashboard and what are the design goals for the dashboard and how do we achieve them. Within this chapter all the research that was done comes together to create the views that we describe. These views are mock-ups that we created and make use of a combination of data from the Dutch governmental IT Dashboard and data we would advise to be created through our guidelines.

6.1 Modeling

Within this section we go through all the KPIs we created in Section 5.6. These KPIs will be visualized and shown within this section. Some of these visualizations are there as example because the data was not there to create one based on real data. The visual representation of the KPIs will be categorized the same as within the GQM in the section create, sustain, and benefit. This visualization will then be collected and grouped to create the different views of the dashboard that can be found within Chapter 6.2.

6.1.1 Create

Within this section we will visualize all the KPIs within the *create* section of the goal question metric. The *create* section of the goal question metric entails the project execution phase where essentially an IT system is created.

Total number of projects and their status

In Figure 11 we visualize the following KPI: all projects summarized and their status. This KPI is important for the stakeholders and gives them an overview of all running projects and their current status. We created Figure 10 using the data we retrieved from the Dutch governmental dashboard. Within Figure 10 all the projects within the dashboard are grouped according to their status. These statuses are as followed: “Afgerond” (Finished), “Geanuleerd” (Canceled), “In uitvoering” (In progress), “Niet gestart” (Not started yet).

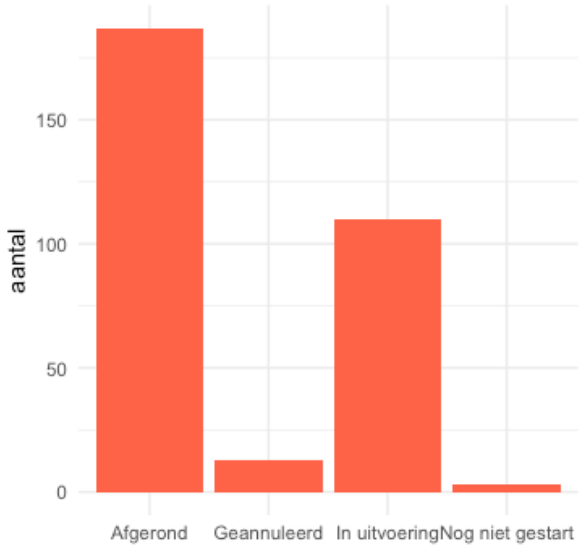


Figure 11 Visualization of all the projects and their status (from the Data)

Project burn rate for each workday

The KPI *Project burn rate for each workday* is added as the last column of Table 3. This is total burn rate for working days. Within the retrieved data the total number of working days were not given. We calculated the working days by using the start date and the end date of the project. We then divided the total cost by the working days. This outcome gives the burn rate for the IT project.

With more information from the government this could be done more accurately. Within the data there is no information about initial investments. If this information would be added to the data, it would enable to create a more accurate burn rate. This will enable projects that have similarity to be better compared to each other.

With all the total burn rates we made a graph of the burn rate for each working day. This can be found within Figure 12. This is now only done for all the projects; however, this can be configured to do this to filter it on a specific group or subject.

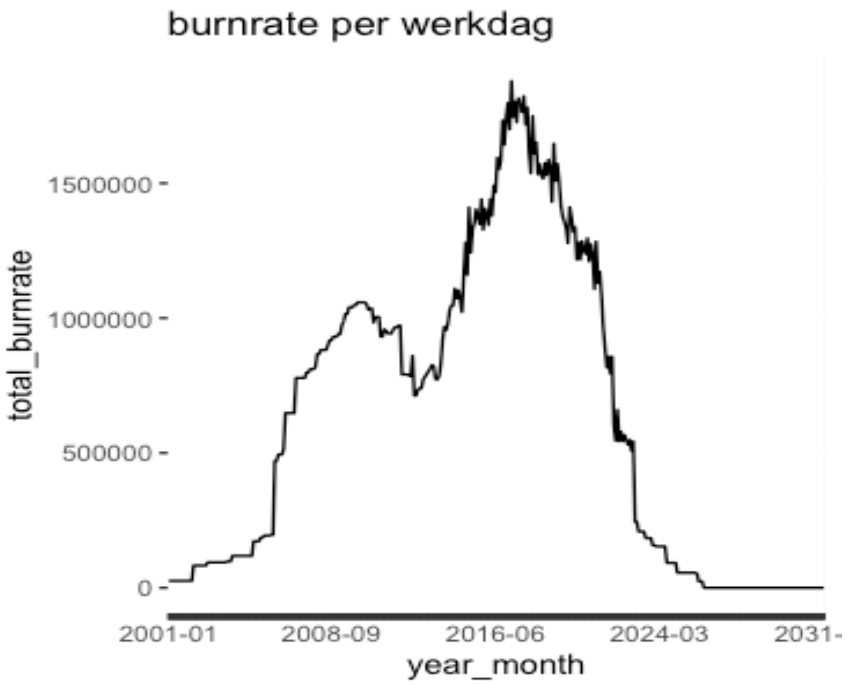


Figure 12 Total burn rate of Projects

Percentile difference of the total cost, the duration, and the burn rate at the beginning of the project and the last "herijking"

This KPI is added as a column of the table view for each project. This is color coded as discussed within Segment 5.6.1. this is part of the table view for each project. How this is visualized is shown in Chapter 6.2. This KPI can be achieved with the data of the IT dashboard. Within the table the columns are indicated and how it is presented. In the detailed view will be shown the total table is shown with all the KPIs corresponding metrics.

Project	Slippage cost difference%	Slippage Runtime difference %	Burn rate
92912	130	105	145

Table 2 Slippage burn rate

Within the earlier mentioned *create* section of the Goal Question Metrics in Chapter 5.6.1 Create, we found that we should calculate the percentage of slippage.

Percentage of slippage represented with indicators

Within this view the stakeholders can get an overview of all the projects and the percentage of slippage between initial cost and the last "herijking" (calibration) of the cost. Within this view the increase or decrease of the duration of the project will also be presented. This can show the correlation between increased cost an increased duration.

The explanation for the slippage

The KPI *The explanation of the slippage* is added to the detailed view for every project. This data is existent within the data; however, this could be elaborated better. For this data point it is important that guidelines are added. These guidelines will make sure that the reasoning for slippage will be explained better to the stakeholder. The stakeholders can then better understand why there is slippage within in the project. What we would suggest as a guideline is that for every time a project goes over the threshold and changes from green to amber or red, there is a clear and concise explanation why this is happening. This explanation should be provided by the project owner or the project manager. This cannot be left open or explained within a few words.

Color indicators in which period the challenges occur

The KPI *Color indicators in which period the challenges occur* this can be found in Figure 13 and is added to the detailed view of every project. This will give the stakeholder a good grasp of what is happening within the life cycle of the project. This can also be enriched by the explanation of the slippage.

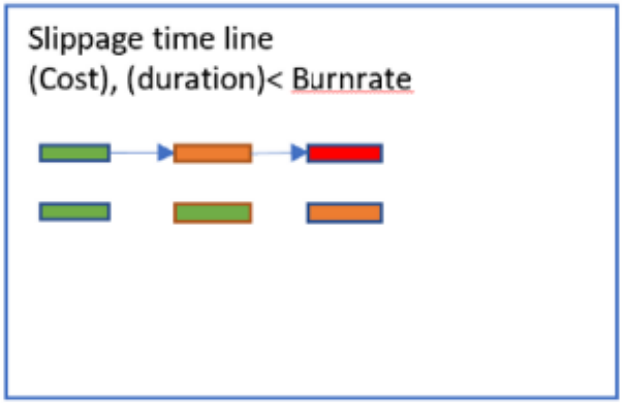


Figure 13 Color indicators in which period the challenge occurs

This KPI is possible to create within the data of the Dutch government. Within the table, the columns and how they will be color-indicated, is shown. In the detailed view will be shown the total table is shown with all the KPIs corresponding metrics.

Within the overview of all slippages and burn rate of the projects that can be found within Table 4. All the percentages are color coded so that the viewer can see in one glance what projects are over the set thresholds.

Project	Slippage cost difference%	Slippage Runtime difference %	Burn rate
92912	130	105	145

Table 3 Slippage and burn rate (colour coded)

6.1.2 Sustain

Within this section we will present the KPI within the sustain section of the goal question metric.

The yearly operational cost

The KPI *The yearly operational cost* is added to the detailed view for every project. This will give the stakeholder information on what the cost will be to keep the IT solution operational. For some projects, the raw data is available to derive this KPI, however there should be better guidelines on how this field is filled in. Within our mock-up we give examples how it could be presented to demonstrate what the added value is of this metric is. This can be found in Chapter 6.2.

6.1.3 Benefit

Within in this section we will present all the KPIs within the benefit section of the goal question metric.

The cost saving for every mutation done within the application

The KPI *The cost saving for every mutation done within the application* is added to the detailed view for every project. This will give the stakeholder information on what the cost will be for every action done by the program.

Project	Average time for the replaced task	Average employment cost for each hour	Mutations done within the first year	Time Saved (hours)	Cost saved (euro's)
928383	5 min	50 euro	7688	640,67	32033,33

Table 4 cost-saving for every mutation

The yearly cost per unit

The KPI *The yearly cost per unit* is added to the detailed view for every project. This will give the stakeholder information on what the cost will be for every action done by the program.

Quantification of the project goals

With the goals added for every project it is important to quantify these goals so that stakeholders can understand the importance of the application. There are two different main goals: financial and non-financial goals. Financial goals are simple to quantify this can be done for example: in time reduction, cost reduction or income. The non-financial goals are more difficult to quantify. A few of these non-financial goals are for example: customer satisfaction, safety, and security, providing information. The reason for this is that the quantification can only be subjective. This should also be done by the minster responsible for the project. Even though the quantification is subjective we added this as a metric to the data. They will be quantified with a rating of one1 till five with one being least important and five being most important.

We did this because even though this metric is subjective together with the objective data that is within the dashboard the stakeholders can make better informed decisions. It will give the stakeholders a good understanding of de importance of the specific project.

Added information to the dashboard for context

We found that in order to make the dashboard more legible it is important to give context to some of the data of the projects. One of the context points is "What is the goal of the project?" and another one is "What problem is the project solving?". This information will be added to the detailed view in Chapter 6.2. This will give the stakeholder information of what the goal is of the IT application. This can be split in two different main goals: financial and non-financial. Within these two main goals there are different sub goals the IT project can fulfil.

6.2 Interaction with the Dashboard

Within this section we will give a preview on how the stakeholders will interact with the new IT dashboard. We will go through the different levels of the dashboard. The goal of these levels is to give the stakeholder all the information they need to make decisions or audit IT projects. The views will go from landing page to the most detailed page on a specific IT project.

The views have been created by taking in account what the requirements, goals and KPIs are of the Dutch governmental dashboard. The views are made up out of data that is retrieved from the Dutch government combined with some fictional data points where the data was not sufficient to support the information need. Within the design we took the *Create, Sustain and Benefit* as guideline. Within the different design the stakeholder can see these subjects being highlighted within in this section we will present the landing page.

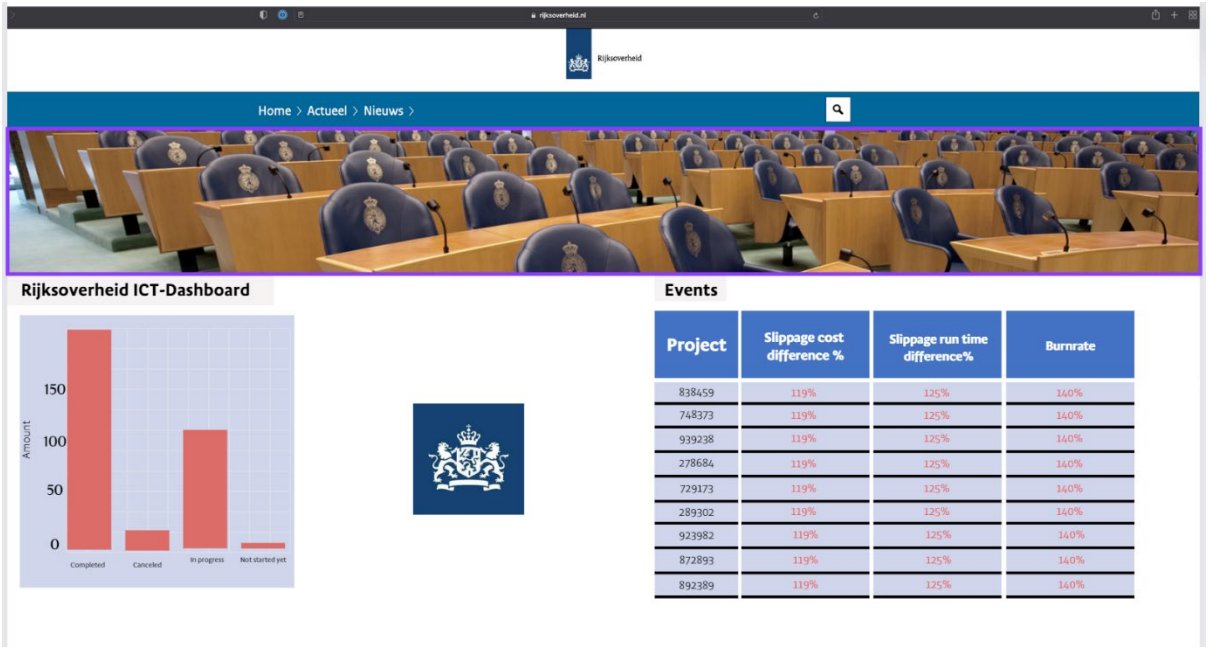


Figure 14 Mock-up Landing Page

Within designing the landing page, we found it important that the stakeholder can get overview of the *Create, Sustain, and Benefit*. We did this through three section the first section is the *create* section.

Within the *create* section of the landing page as shown in the figure landing page. The overview of all IT projects that are currently being run within in the Dutch government will be presented. These projects are grouped with their status. Next to the total projects there is a table with ten projects that have the most negative fluctuation of the slippage cost, duration, or runtime increase. Within the first page it is important to show what the current events are. This will give the stakeholder the opportunity to intervene if needed. All projects in the table are clickable so that the stakeholder can dive deeper to the detailed view of the project. The detailed view can be found in Figure 17. The goal of the landing page is that the stakeholder gets a bird's eye view of all the projects that are run within the Dutch government. When the stakeholder is interested, it needs to be possible for him to drill down within the dashboard.

This is what we found that is important through the dashboard reviews in Chapter 4 Dashboard Review. Within the sustain section of the landing page as shown in 5.

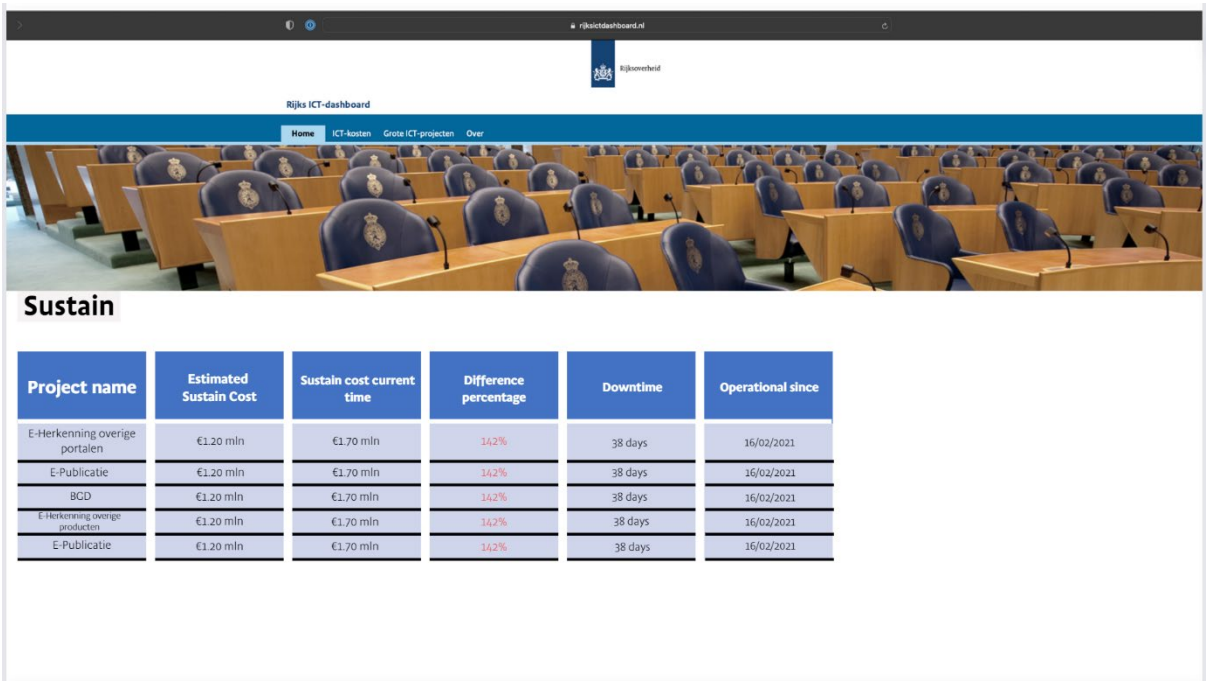


Figure 5 Mock-up Sustain view

The stakeholders can see the projects that are operational and what their yearly cost is. This is compared to the estimated maintenance cost. The difference between the estimated maintenance cost and the real maintenance cost is presented and in the differential percentage. This information is not available in the current dashboard. It is important that this information is added so that the stakeholders can get a better grasp of the total cost of projects. This can then also be used to calibrate future estimations of the maintenance cost. Better estimations will lead to the ability to assess financial benefit to the total cost of the project. Within the Figure 15 for each project the downtime is given. You can also see since what date the application is operational. This mock-up is made out of data that can be found within the dashboard there is also information that is not found within the dashboard. This data is created to give an indication how they can be presented. We would suggest that after the project is done information is still gathered. The information that we want to be gathered is the yearly sustain cost, the amount of downtime since its inception, the downtime percentage for the last year and the date since the program is in operation.

Filtered project view

We found through our research for other dashboards that can be found within Chapter 4, it is important to give the users the freedom to filter. This gives the stakeholder the possibility to find connection between the IT Projects. Some of the examples of the project is: the department, the project health, the described benefit. The filtered view can be seen in the Figure 16.

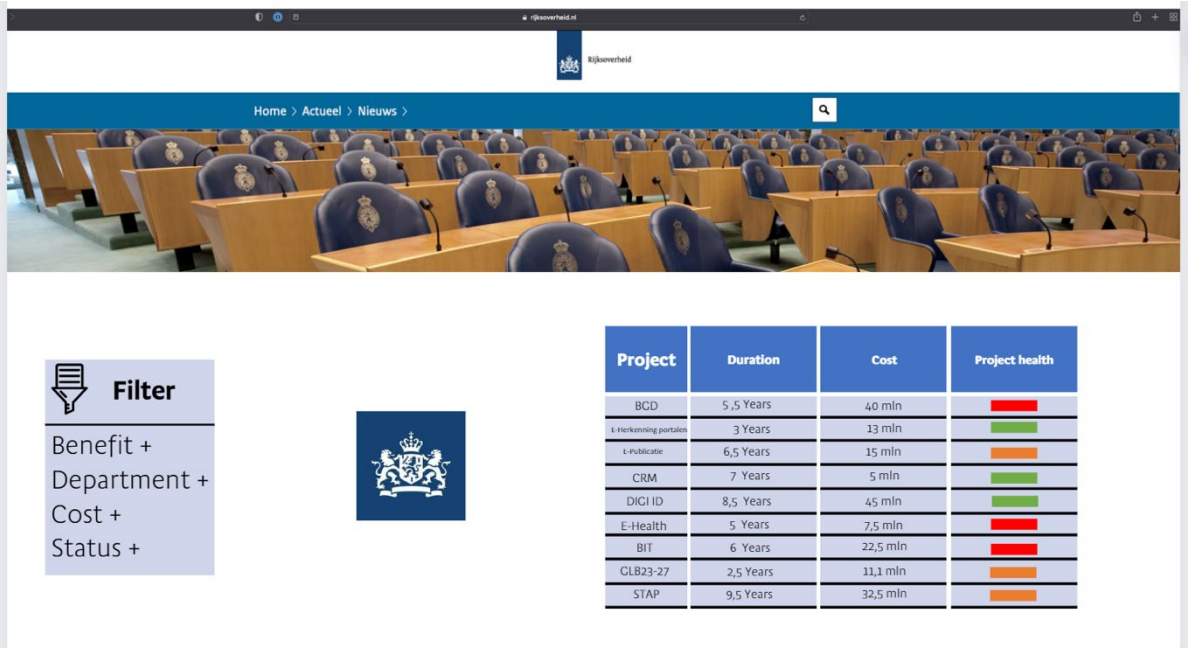


Figure 16 Mock-up Filtered view

Within this screen the stakeholder can filter within the different meta data. For every project within the Dutch government there is a lot of meta data saved. Some examples of the information saved with every project is the: department, the status, the minister that was in charge at that given moment. Enabling the user to filter and select what projects they want to see can give them the opportunity to analyses the data.
Within Chapter 4 we found that this is an important functionality.

For every project that meets the requirements of the filter the following key information is showed. The project, the duration the cost and the project health. The project health is color coded. This can have the same colors as the thresholds that are being used through the whole dashboard, in this case in colors: green, amber and red. The project health will change from color when the slippage hits the threshold the duration gets exceed or the cost of the project oversees the financial benefits. The last column will give the reasoning what caused the project health to chance color. Every project will be clickable so that the stakeholder can see the detailed view of the project.

Detailed view

Within the dashboard it is important that the stakeholders can get an overview of all the projects, however they should also be able to drill down to the most detailed view of a specific project. In this section we will present the detailed view. The following mock-up is here to give an indication of how the new dashboard page can show all the requirements within the detailed view.
Within in this view it is important that the stakeholder is presented with all the important information about the project.

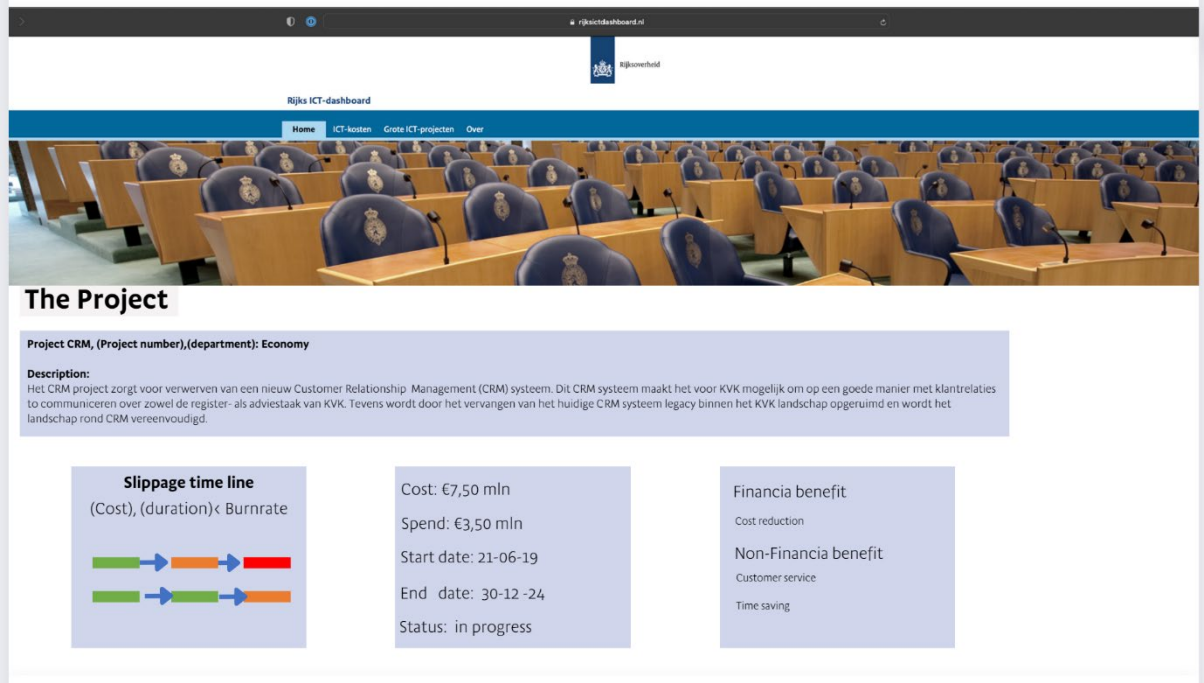


Figure 17 Mock-up Detailed view

The detailed view is split up into different parts. The first section is the relevant information about the project. This information is given so that the stakeholder can get the overview of what the project is entails. Within in this part information is given as the: project name, project number, department, minister, and the description. This information can already be found within the data of the Dutch government.

Slippage timeline

In Figure 17 of the detailed view the slippage timeline is shown this timeline presents in what period the slippage percentage changes more than the given thresholds. This way the stakeholder can observe when issues happened within in the project. When clicking on the timeline for more information. You can see the reasoning why the slippage occurred. This information will help the stakeholder to understand if there is need for further elaboration or that the information is sufficient.

Financial

In Figure 17 of the detailed view the financial information is given. This information gives the stakeholder the information about the financial investment of the project. The information that is given is: The estimated cost of the project, the amount spends at the current date, the burn rate of the project. The maintenance an operational cost for when the project is finished. In the case that the project has financial benefits the following information is also given the breakeven point in the case that the benefit of the project is a financial benefit

Schedule information

Within Figure 17 of the detailed view the financial and schedule information is given. This information gives the stakeholder the information about the financial and time investment the project will need. It also will give the stakeholder on how much of the estimate cost is spend.

Benefit

Within Figure 18 of the detailed view the benefit of the projects is shown. This will give the user the information of why the project is needed and what it is going to improve. This information can help the stakeholder understand why the project is initiated. This is information that can be divided within into main subjects. Financial and non-financial benefits. The financial benefits can be measured by numerical values such as money saved or time saved, what also can be estimated into financial amount. The non-financial benefits this is not that simple. We suggest that the nonfinancial benefits will be graded by the project owner on a scale. This is very subjective however with the rest of the information about the project stakeholder will still be able to get a sense of why the IT application is needed. We also suggest that when the project is finished this information will be updated. This way the stakeholder can see if the application meets the estimated benefits. This can give the stakeholder the information if the project is running as expected or that the project is exceeding the estimation. The benefit view is shown in Figure 18.

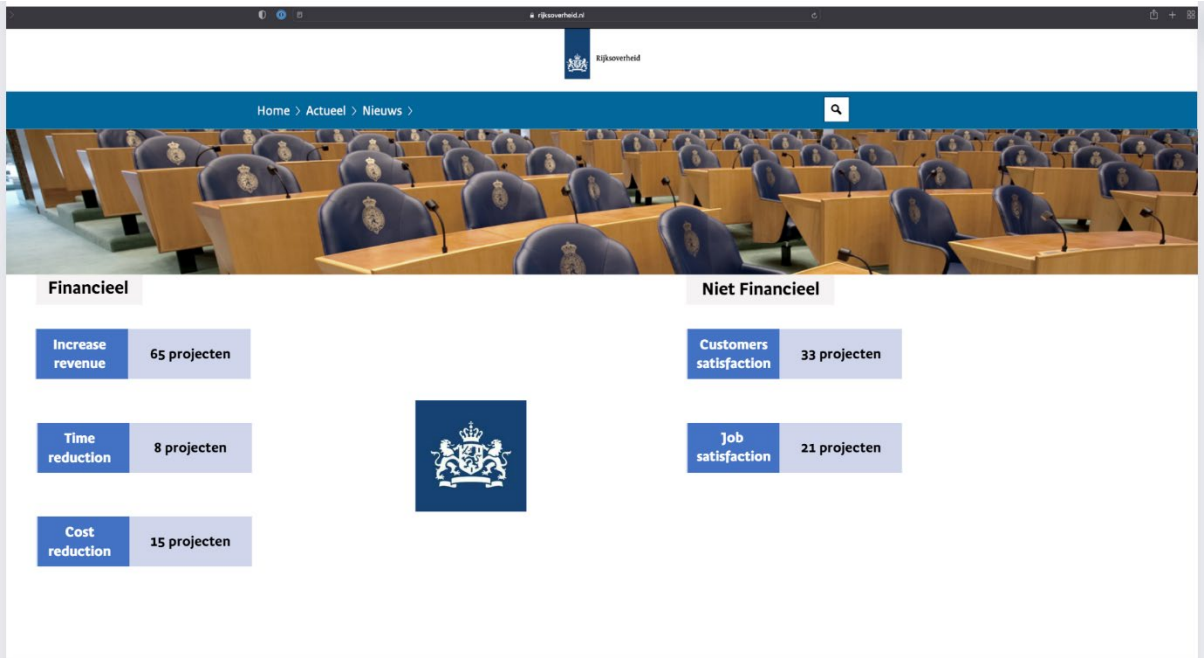


Figure 18 Mock-up overview Benefits

Within in the view of benefits all the projects are presented by their goal and what they will achieve. The number of projects that have the same benefit and what their status is.
This will give the stakeholder a good overview of all the projects catering to the different benefits. What we include within the dashboard are the benefits. These benefits can be split in financial or non-financial benefits. This split is not enough cause for the non-financial benefits you can also split this up in functional benefits. We are going to make this apparent with separate views. We divided the benefits in different subjects that will be presented in different

ways. These are the following: the cost saving, time saving and functional benefits.

Mutations

Within in this Figure 18 the benefits are shown. These benefits are the base of the number of mutations the program has done in its time of being operational. This can be seen in Figure 19.

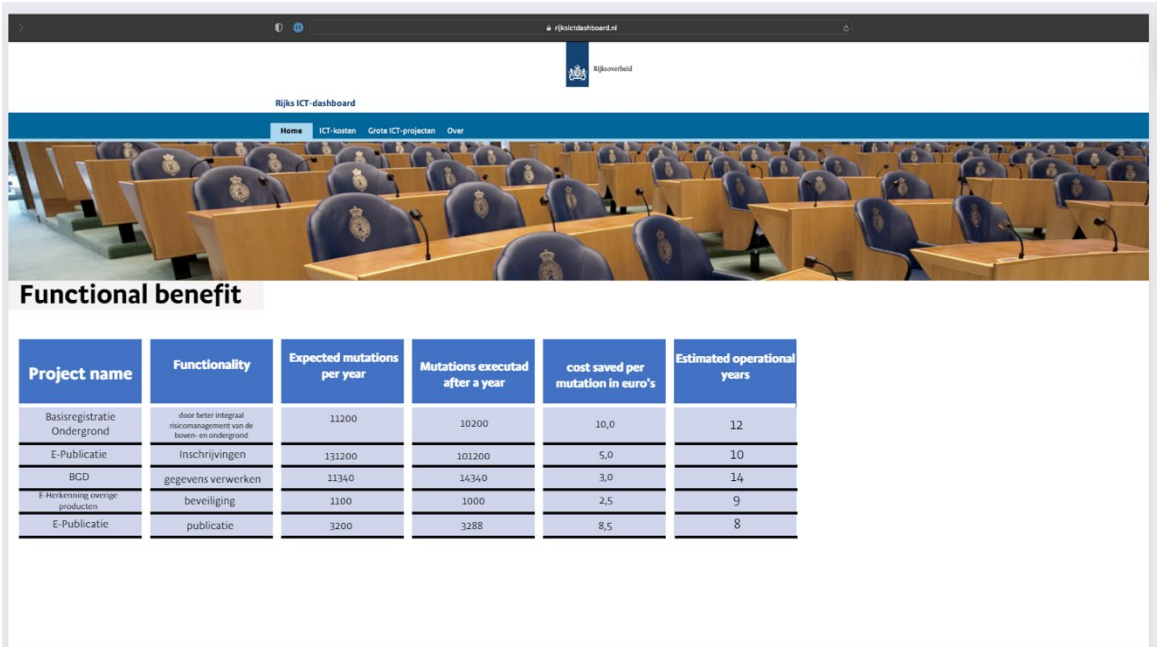
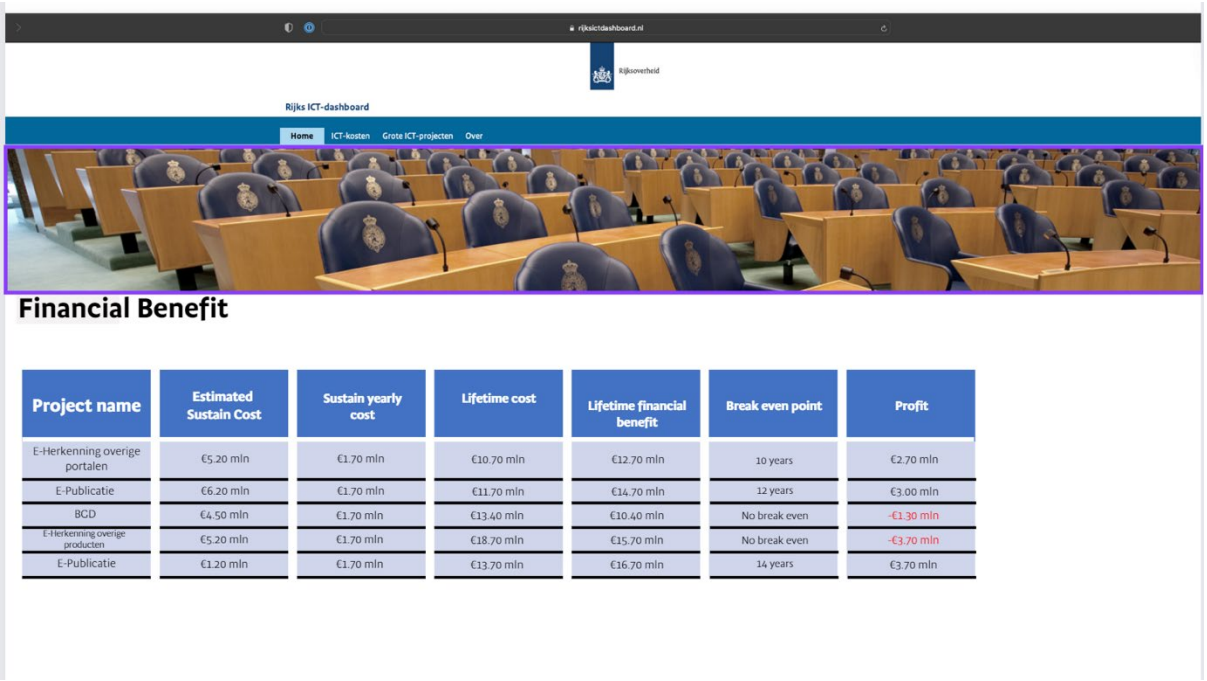


Figure 19 Mock-up Functional Benefits

Within in this figure the first column tells the project name of the project. The next one describes what function the program executes. The next column shows the number of functions the program is expected to execute for every year that it is operational. This is not included in the data of the current dashboard. However, with this information stakeholders can understand what the impact is of the IT- project. This is followed by the number of functions after a year of being operational. With this data it will enable to calculate an estimate of mutation the application will do through its lifetime. In the next column the total lifetime calculation can be used to calculate the cost for every mutation. This can give a better indication for the added value that the application gives and if the mutation is worth the cost.

Financial Benefit

Within Figure 20, the benefit section presents the financial benefits. The financial benefits are shown side by side, so the user can compare with the total cost of the application to create and to sustain. When the cost to create and sustain becomes higher than the potential benefits of the project, an alert will go off and the difference will be color coded.

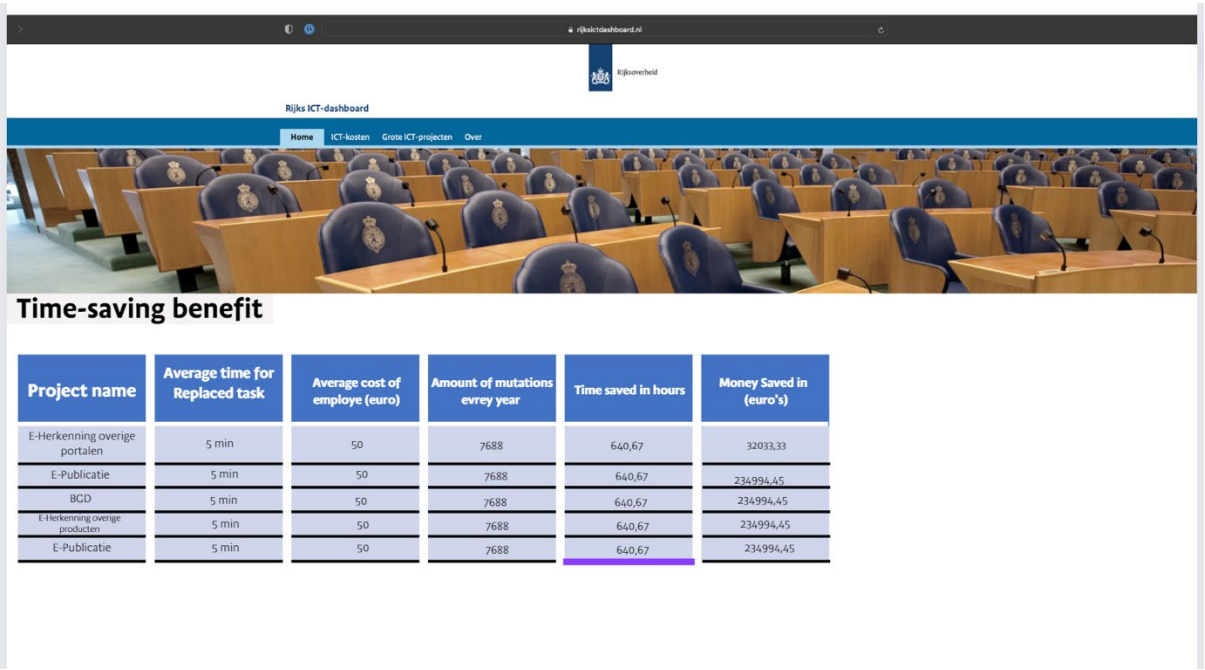


Project name	Estimated Sustain Cost	Sustain yearly cost	Lifetime cost	Lifetime financial benefit	Break even point	Profit
E-Herkenning overige portalen	€5.20 mln	€1.70 mln	€10.70 mln	€12.70 mln	10 years	€2.70 mln
E-Publicatie	€6.20 mln	€1.70 mln	€11.70 mln	€14.70 mln	12 years	€3.00 mln
BCD	€4.50 mln	€1.70 mln	€13.40 mln	€10.40 mln	No break even	-€1.30 mln
E-Herkenning overige producten	€5.20 mln	€1.70 mln	€18.70 mln	€15.70 mln	No break even	-€3.70 mln
E-Publicatie	€1.20 mln	€1.70 mln	€13.70 mln	€16.70 mln	14 years	€3.70 mln

Figure 20 Mock-up Financial Benefits

Time saving benefit

Within Figure 21 the projects are shown that will save the government time. The first column shows the project name. The next column shows the average time saved by the application by replacing the task. The average cost of a government employee is needed to calculate how much money is saved by completing the task within the application. The next column shows the number of mutations the application processes every year. With this data point it will enable the ability to calculate the saved amount of money in euro's.



Project name	Average time for Replaced task	Average cost of employe (euro)	Amount of mutations evrey year	Time saved in hours	Money Saved in (euro's)
E-Herkenning overige portalen	5 min	50	7688	640,67	32033,33
E-Publicatie	5 min	50	7688	640,67	234994,45
BCD	5 min	50	7688	640,67	234994,45
E-Herkenning overige producten	5 min	50	7688	640,67	234994,45
E-Publicatie	5 min	50	7688	640,67	234994,45

Figure 21 Mock-up Time saving benefit

7. Conclusion

Within in this thesis we explored what the Central Dutch government can do to present their IT projects in a more adequate way. Through this research we presented new KPIs and how these can be presented to all the stakeholders of the government ICT dashboard. The following section will recapitulate the research objectives and will elaborate how these were attained.

7.1 Answer to research questions

Within this chapter all the research questions will be answered.

7.1.1 Sub questions

Through answering the sub questions, we will answer the main research question that can be found in chapter 7.1.2

What are other governments doing in the field?

We studied several governmental ICT Dashboards that display big IT projects. At the time of this research, we identified three countries with ICT dashboards. These countries were the United States of America, Australia, and The Netherlands. To get an overview of what these governments were doing in the field, we reviewed the different dashboards using the method of (Sarikaya et al., 2019) . While reviewing the dashboards we compared them and discovered that the American and Australian dashboards have more functionalities than the Dutch governmental dashboard. The functionalities such as color coding, filtering of the projects and comparing possibility of the projects, but also fully completed data make the data more transparent to the different stakeholders. The data is easier to be found and understood as well. These findings were noted down and were used to establish the requirements for the dashboard improvements of the Dutch governmental dashboard that were developed within our research.

What information can be inferred from the raw data?

Within this research we analysed all the available data from the Dutch governmental ICT dashboard. The following information can be inferred from the analysed raw data: the total number of projects, the estimated cost per project, the duration of each project, project owner per project. With the knowledge of the analysed raw data, we designed KPIs using the GQM approach.

What kind of information is important to the involved stakeholders?

To establish what information is important to the involved stakeholders we had to establish what the exact purpose is of the ICT Dashboard. The main purpose of the ICT Dashboard is to display all the current ICT projects within the government, how they are proceeding and what their exact status is. The most important information to different stakeholders is the content and the KPIs. With this information the stakeholders are able to understand how the ICT projects are being run within the government. It is also important that the dashboard is giving transparency to all the stakeholders and being understandable to stakeholders with a wide variety of different backgrounds.

What is the best way to display the analysed information?

The best way to display the analysed information of the Dutch governmental dashboard is to create a visualization of it. We created mock-ups to present how the new dashboard could be navigated and what functionality the different views have.

7.1.2 Main research question

What can the central Dutch government do to present their data on IT Project in a more adequate way to various stakeholders?

Within this research we discovered that it is highly important that a governmental dashboard is comprehensible for the general public. Therefore, it is important to visualize the data on the dashboard in a way that it is understandable for a large group of individuals.

After completing this research we would highly recommend the Dutch government to adhere to the following requirements: “total number of projects and their status”, “project burn rate for each workday”, “percentile difference of the total cost”, “the duration and the burn rate at the beginning of the project and the last “herijking”, “the explanation for the slippage”, “color indicators in which period the challenges occur”, “the yearly operational cost”, “the cost saving for every mutation done within the application”, “the yearly cost per unit”, “quantification of the project goals”.

Within the current Dutch governmental dashboard some of these requirements are not met. Therefore, their data is not complete. The Dutch government should adhere the requirements mentioned above to complete the missing data points.

Next to the mentioned KPIs, there are also visual requirements that the Dutch governmental dashboard is partially missing. Within the Chapter 6.1 Modeling we created detailed mock-ups to visualize how the data of the Dutch governmental dashboard could be presented in a more adequate way to various stakeholders.

Our advice within the mock-ups can also be brought up to the following points:

Color coding adjustments
By adding color coding to the data, it will be easier to discover whenever there are financial outliers. For example, whenever the costs of a certain project pass the estimated costs, there will be a visual que presented by a certain color.

Adding missing financial information
All the financial data should be filled in, not only the total costs per project should be visible, but detailed split.

F(
In order to present the Dutch governmental dashboard in a more adequate way the Dutch government could also add some new functionalities. For example, enabling to compare projects, to filter projects, to delph deeper into a certain project, to give an overview of an IT project.

7.2 Discussion

Within this research no particular hypothesis was established. We did however have our own speculation about the topic of the Dutch governmental dashboard. The speculation was that the Dutch governmental dashboard was not transparent enough and could be improved. The speculation was caused by a lot of criticism from media outlets and stakeholders around the dashboards and the fact that the dashboard has gone through different iterations since its inception.

After conducting our research and looking back at our first speculation we can state that that the speculation was correct. By analyzing the requirements given by the Dutch government within the letter by R.W. Knops (2020), analyzing and comparing the existing governmental ICT dashboards from the USA, Australia and The Netherlands and also retrieving and analyzing the data of the current Dutch governmental dashboard we concluded that the current Dutch governmental dashboard was not transparent enough and comprehensible for the general public.

Reliability

Within this research we used three different methods to establish the requirements of governmental ICT dashboard: *analysed the requirements given by the Dutch government within the letter* by R.W. Knops (2020)*, analysed and compared the existing governmental ICT dashboards from the USA, Australia and The Netherlands, retrieved and analysed the data of the current Dutch governmental dashboard.*

Within this research we also used the following models to establish the KPIs for the new Dutch governmental dashboard: *Goal Question Metrics model (GQM)* and *Create, Sustain, Benefit model (CSB)*.

We could state that this research has its limitations when it comes to being reproducible for future researchers. The steps of this research are fully traceable; however it is not a given that the future researcher will come to the same results by following the same steps. This is caused by a fact that we executed this research without having any contact with the stakeholders. Therefore, we cannot state that it meets all the requirements that the stakeholders might have. This research is based on theory and findings derived from the analysis. For future studies we would recommend interviewing different types of stakeholders to increase the overall reliability of their research.

Validity

The content validity of this research is high. According to the researched literature models such as *Goal Question Metrics model (GQM)* and *Create, Sustain, Benefit model (CSB)* help with establishing KPIs. Within this research we used both of these models to establish the new KPIs for the Dutch governmental dashboard. The models used within this research helped answering the main research question.

The internal validity of this research is high. Within this research we used qualitative and quantitative methods. The used qualitative methods were *analyzing the requirements given by the Dutch government within the letter* by R.W. Knops (2020) and *analyzing and comparing the existing governmental ICT dashboards from the USA, Australia and The Netherlands*.
The quantitative method used within this research was: *retrieving and analyzing the data of the current Dutch governmental dashboard* that was done by coding of the data.

The methodological triangulation is therefore high as the methods used to answer the main research question were data analyses (quantitative method) and desk research (qualitative method).

However, the internal validity of this research could have been even higher if we used the method of interviewing the stakeholders. The reason for it is that the stakeholders have been mentioned in the research questions. The answers concerning the stakeholders were now answered with the method of desk research. It would have been more valid if we would also interview the stakeholders and ask them the questions we needed to answer within this research. This was also described within the thesis proposal. The interviews were never taken, due to time constrains.

7.3 Limitations and Future work

The research within this thesis is done by researching existing data. With this existing data the requirements and then the KPIs were made. This could be improved in the future by adding by interviewing stakeholders. We did not have access and the resources to get information of enough stakeholders to get validated requirement. This is also the case for the validation of the new model we did not have access to all the stakeholders. This can be researched in future work.

Within in this research the raw data of the Dutch government is used. This has its limitations because of missing information within the data. There were a lot of irregularities. The data for the IT project also end after the projects are finished. To give more insight in to improvement for future projects valuable to have data about how the applications perform. Some metrics that could be easily obtained for example are: what is the maintenance cost, how many mutations did the application executed within a year of being in operational use, and what was the total downtime of the applications. Following our recommendation, we hope that the quality of the raw data improves so that in future research. The stakeholders can get better information on how IT projects are being executed within in the Dutch governments. This enables the stakeholders to deliberate decisions and gives them the capability to inspect the ICT Projects.

The cost-benefit analysis of the Dutch Government ICT dashboard is incomplete at the current moment. This is because of there is not information on the benefits and the sustain part of the CSB. There is not enough information on what maintenance cost are for application or information on the expected value the new application will give the government. To give the stakeholders a better insight on the IT projects within the government it would be good to have total overview of all the information.

The research did not focus on creating a working prototype of the governmental dashboard. It main focus was to give a visual indication on how the data can be presented. For future work this can be researched to create a working governmental dashboard.

As mentioned previously in Chapter 5.1, on the 23rd of May 2021 we downloaded the data from the Rijks- ICT-dashboard for the first time for research purposes. As the finalisation of the research was done in September 2022, by this time the data we worked with within this research was already a year old. On the 21st of September 2022 we downloaded the data again from the Rijks ICT-dashboard in the JSON file format, because we wanted to see if the developed code still worked on a new data set. We loaded the data into R Studio and made a few modifications to the code. We found out that the code worked on the majority of the new data file. There were 30 new projects within the data, but the format of how the data was saved before, stayed unchanged for what we could see. This means that the code we developed is still applicable to the data set from the Rijks ICT-dashboard. The code is included within Appendix I of this research and can be used for future researches, as well as be further developed.

BIBLIOGRAPHY

Adviescollege ICT-toetsing. (2022).

Allio, R. J., & Fahey, L. (2012). Joan Magretta: what executives can learn from revisiting Michael Porter. Strategy & Leadership, 40(2). https://doi.org/10.1108/10878571211209297

Chavez, M. D., Ceballos, H. G., & Cantu-Ortiz, F. J. (2020). A data analytics approach to contrast the performance of teaching (only) vs. research professors. International Journal on Interactive Design and Manufacturing, 14(4), 1577–1592. https://doi.org/10.1007/s12008-020-00713-5

Darch, C., & Underwood, P. G. (2010). Freedom of Information and the Developing World. In Freedom of Information and the Developing World. https://doi.org/10.1533/9781780630205

de Bruijn, H., & van Helden, G. J. (2006). a Plea for Dialogue Driven Performance-Based Management Systems: Evidence From the Dutch Public Sector. Financial Accountability and Management, 22(4), 405–423. https://doi.org/10.1111/j.1468-0408.2006.00432.x

European accessibility act - Employment, Social Affairs & Inclusion - European Commission. (2022). https://ec.europa.eu/social/main.jsp?catId=1202

Fuguitt, D., & Wilcox, S. J. (1999). Cost-Benefit Analysis For Public Sector Decision Makers. Praeger.

Ganapati, S. (2011). Use of Dashboards in Government. www.businessofgovernment.org

Graham, C. B. (2012). Book Review: Governmental Transparency in the Path of Administrative Reform. The American Review of Public Administration, 42(1), 125–127. https://doi.org/10.1177/0275074011421628

Grimmelikhuijsen, S. G., & Meijer, A. J. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. Journal of Public Administration Research and Theory, 24(1), 137–157. https://doi.org/10.1093/jopart/mus048

Guillaume Lafortune, S. G. & Z. L. (2018). The Palgrave Handbook of Indicators in Global Governance. In The Palgrave Handbook of Indicators in Global Governance. https://doi.org/10.1007/978-3-319-62707-6

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly: Management Information Systems, 28(1). https://doi.org/10.2307/25148625

Hevner Alan, R. (2007). A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems, 19(2), 87–92. https://www.researchgate.net/publication/254804390_A_Three_Cycle_View_of_Design_Science_Research

Jaarrapportage 2019Bureau ICT-toetsing. (2020).

Jan Van den Broeck, & Jonathan R Brestoff. (2013). Epidemiology: Principles and Practical Guidelines. In J. van den Broeck & J. R. Brestoff (Eds.), Epidemiology: Principles and Practical Guidelines. Springer Netherlands. https://doi.org/10.1007/978-94-007-5989-3

Kloeckl, K. (2019). Bringing Clarity to Transparency Transparency Clarity to Clarity to Bringing Bringing Bringing Bringing.

Laurent, W. (2007). Dashboarding Governance. DM Review, 17(April).

Liu, Y.-C., & Lin, C. (2008). How Are Public Sector Organizations Assessing their IT Investments and Benefits - An Understanding of Issues For Benchmarking. International Journal of Advanced Information Technologies, 86–104.

Maheshwari, D., & Janssen, M. (2014). Dashboards for supporting organizational development: Principles for the design and development of public sector performance dashboards.

ACM International Conference Proceeding Series, 2014-January. https://doi.org/10.1145/2691195.2691224
Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680
Parmenter, D. (2015). Key Performance Indicators - Developing, Implementing, and Using Winning KPIs - Third Edition. In SpringerBriefs in Applied Sciences and Technology.
Rijks ICT Dashboard. (2021). https://www.rijksictdashboard.nl/
Rowley, J. (2011). E-Government stakeholders - Who are they and what do they want? International Journal of Information Management, 31(1), 53–62. https://doi.org/10.1016/j.ijinfomgt.2010.05.005
R.W. Knops. (2020). Informatie- en communicatietechnologie (ICT); Brief regering; Doorontwikkeling Rijks ICT-dashboard (Issue 18, p. 2).
Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019a). What do we talk about when we talk about dashboards? IEEE Transactions on Visualization and Computer Graphics, 25(1), 682–692. https://doi.org/10.1109/TVCG.2018.2864903
Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019b). What do we talk about when we talk about dashboards? IEEE Transactions on Visualization and Computer Graphics, 25(1), 682–692. https://doi.org/10.1109/TVCG.2018.2864903
van Solingen, R., & Berghout, E. (1999). The Goal/Question/Metric Method: a practical guide for quality improvement of software development. McGraw Hill, ISBN, 7(January), 7.
Visser, J. (2013). Complexity and Success of Projects
Wirth, R., & Hipp, J. (2000). CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 24959, 29–39.

Appendix I Code

```
library(tidyverse)
library(jsonlite)
library(kableExtra)
library(dplyr)
library(purrr)
library(bizdays)

rm(list = ls())
#cleaning the project

library(jsonlite)
ICT_projects <- fromJSON("/Users/jaychin/Dropbox/Dropbox/thesis
documents/thesis.Data/23may.json", flatten = T)
#loading in the JSON data using jsonlite
list_list <- lapply (ICT_projects, "[[", c("herijkingen"))
nrows <- lapply(list_list, nrow)
df_ict_projects <- dplyr::bind_rows(list_list)
#creating a list of all the herijkingen data
ids <- lapply(ICT_projects, "[", c("id", "peildatum"))
ids <- dplyr::bind_rows(ids)
#connecting ID's of the project to the data frame
name <- lapply(ICT_projects, "[", c("id", "naam", "peildatum"))
name <- dplyr::bind_rows(name)
#connecting names and id's of the project to the data frame
verwachtekosten <- lapply(ICT_projects, "[", c("id", "naam",
"verwachte_kosten_beheer_onderhoud", "peildatum"))
verwachtekosten <- dplyr::bind_rows(verwachtekosten)
startdatum <- lapply(ICT_projects, "[", c("id", "naam",
"startdatum"))
startdatum <- dplyr::bind_rows(startdatum)
ministerie <- lapply(ICT_projects, "[", c("id", "naam", "startdatum",
"ministerie"))
ministerie <- dplyr::bind_rows(ministerie)
id_vector <- numeric()
for (i in 1:length(nrows)) {
 id_vector <- append(id_vector, rep(names(nrows)[i], nrows[i]))
}
df_ict_projects <- cbind(id_vector, df_ict_projects)
df_ict_projects <- df_ict_projects %>%
 arrange(id_vector, peildatum)
df_ict_projects <- left_join(df_ict_projects, ministerie, by =
c("id_vector" = "id"))
#joining startdatum to df_ict_projects
df_ict_projects <- df_ict_projects %>%
 arrange(id_vector, peildatum, naam, startdatum)
#change data type for char to date
df_ict_projects <- df_ict_projects %>%
 mutate( peildatum = as.Date( peildatum , format= "%Y-%m-%d"),
```

```
 aanmaakdatum = as.Date( aanmaakdatum , format= "%Y-%m-%d"),
 actueel_einddatum = as.Date( actueel_einddatum , format=
"%Y-%m-%d"),
 startdatum = as.Date( startdatum , format= "%Y-%m-%d"),
 actueel_overige_projectkosten = as.numeric
( actueel_overige_projectkosten),
 daadwerkelijk_hardware_software = as.numeric
( daadwerkelijk_hardware_software ))
#change data type from char to date and from char to numeric
df_ict_projects <- df_ict_projects %>%
 mutate( daadwerkelijk_hardware_software = as.numeric
( daadwerkelijk_hardware_software ))
#change data type for char to numeric
#df_ict_projectstest <- df_ict_projects %>%
 #mutate( actueel_overige_projectkosten = as.numeric
(actueel_overige_projectkosten) * 1000000)
# if needed to change to mln
df_ict_projectstest <- df_ict_projects
#connecting ID's of the project to the data frame
df_ict_projects_uncleaned <-df_ict_projects
df_ict_projects <- df_ict_projects %>%
 mutate(werkdagen = bizdays(startdatum, actueel_einddatum))
#%>%
 #select((-daadwerkelijk_dataverbindingen),
 #(-actueel_hardware_software))
 names(dfwerkdagen)[2] <- "dagen"
#naam veranderen van collum
dfwerkdagen <- data.frame(bizdays(from = df_ict_projects$startdatum ,
to = df_ict_projects$actueel_einddatum, ))
dfwerkdagen <- cbind(id_vector, dfwerkdagen)
df_ict_projects <-df_ict_projects %>%
 group_by(id_vector) %>%
 mutate(start_end = ifelse(peildatum == first(peildatum),
 1,
 ifelse(peildatum == last(peildatum),
 2,
 NA)))
#creating a extra collum called start_end with first date in
peildatum called 1 last date in peildatum ccalled 2 and evreything
inbetween NA
df_ict_projects %>% filter(start_end %in% c(1,2))%>%
 select(peildatum, id_vector, actueel_totaal_projectkosten,
start_end)%>%
 distinct(id_vector,start_end, .keep_all = TRUE)%>%
 pivot_wider(id_cols = c("id_vector", "start_end"), names_from =
start_end, values_from = actueel_totaal_projectkosten)%>%
#distinct is een functie waar alleen de rijen geeft met unieke datum)
 mutate(geld_verschil = as.numeric(`2`) - as.numeric(`1`))

df_ict_projects <- df_ict_projects %>%
 mutate (brun_rate = as.numeric(actueel_totaal_projectkosten) /
```

```
as.numeric(werkdagen)*1000000)
#calculating the burn rate for evrey working day
# weg halen ggplot
(df_ict_projects$actueel_einddatum - df_ict_projects$startdatum)
#if datum > begin datum and datum < actuele eindatum
df_ict_projectstest <- df_ict_projects %>%
 mutate(year_month = seq(df_ict_projects$startdatum, to
=df_ict_projects$actueel_einddatum, by = "month"))
df_ict_projects <- Months_projects =
seq(df_ict_projects$startdatum[1], to
=df_ict_projects$actueel_einddatum[1], by = "month")
#df_ict_projects <-

 #df_ict_projects %>%
 #mutate(year_month = format(as.Date(startdatum), "%Y-%m"))%>%
 #group_by(year_month, ministerie)%>%
 #summarise(sum_burn = sum(brun_rate))
#apply(df_ict_projects, 1, function(x) seq
((x[df_ict_projects$startdatum]),
(x[df_ict_projects$actuel_einddatum]), "month"))
#df_ict_projects %>% group_by(year_month) %>% summarise(sum_burn =
sum(brun_rate))
#f_ict_projects %>% group_by(year_month, ministerie) %>%
summarise(sum_burn = sum(brun_rate))
#werkt niet
empty_list <- list()
id_vectors <- unique(df_ict_projects$id_vector)
for (i in 1:length(id_vectors)) {

 print(i)
 x <- df_ict_projects %>%
 filter(id_vector == id_vectors[i],
 start_end %in% c(1,2)) %>%
 distinct(startdatum, actueel_einddatum, .keep_all = T)

 if(nrow(x) > 1) {

 empty_list[[i]] <- data.frame(months = seq(from =
as.Date(x$startdatum[1]),
 to =
as.Date(x$actueel_einddatum[x$start_end == 2]),
 by = "months"),
 burn_rate = x$brun_rate[2],
 ministerie = x$ministerie[1],
 id = x$id_vector[1])
 } else next

}
df_joined <- lapply(empty_list, rbind)
bind_rows(empty_list) %>%
 mutate(year_month = as.factor(substr(months, 1, 7))) %>%
```

```
 group_by(year_month) %>%
 summarise(sum_burn = sum(burn_rate)) %>%
 ggplot(aes(x = year_month, y = sum_burn, group = 1)) +
 geom_line() +
 theme(axis.text.x = element_text(angle = 90))
df_BZK <- df_ict_projects[df_ict_projects$ministerie == "Binnenlandse
Zaken en Koninkrijksrelaties",]
df_AZ <- df_ict_projects[df_ict_projects$ministerie == "Algemene
Zaken",]
df_BZ <- df_ict_projects[df_ict_projects$ministerie == "Buitenlandse
Zaken",]
df_Mindef <- df_ict_projects[df_ict_projects$ministerie ==
"Defensie",]
df_EZK <- df_ict_projects[df_ict_projects$ministerie == "Economische
Zaken en Klimaat",]
df_Minfin <- df_ict_projects[df_ict_projects$ministerie ==
"Financiën",]
df_IenW <- df_ict_projects[df_ict_projects$ministerie ==
"Infrastructuur en Waterstaat",]
df_JenV <- df_ict_projects[df_ict_projects$ministerie == "Justitie en
Veiligheid",]
df_LNV <- df_ict_projects[df_ict_projects$ministerie == "Landbouw,
Natuur en Voedselkwaliteit",]
df_OCW <- df_ict_projects[df_ict_projects$ministerie == "Onderwijs,
Cultuur en Wetenschap",]
df_SZW <- df_ict_projects[df_ict_projects$ministerie == "Sociale
Zaken en Werkgelegenheid",]
df_VWZ <- df_ict_projects[df_ict_projects$ministerie ==
"Volksgezondheid, Welzijn en Sport",]
#creating Dataframes for each ministery of the government
 df_ict_projects
#dfwerkdagen <-
data.frame(difftime(df_ict_projects$actueel_einddatum,
df_ict_projects$startdatum , units = c("days")))

#dfwerkdagen <- cbind(id_vector, dfwerkdagen)

colnames(dfwerkdagen)
#namen van de collum opvragen
```