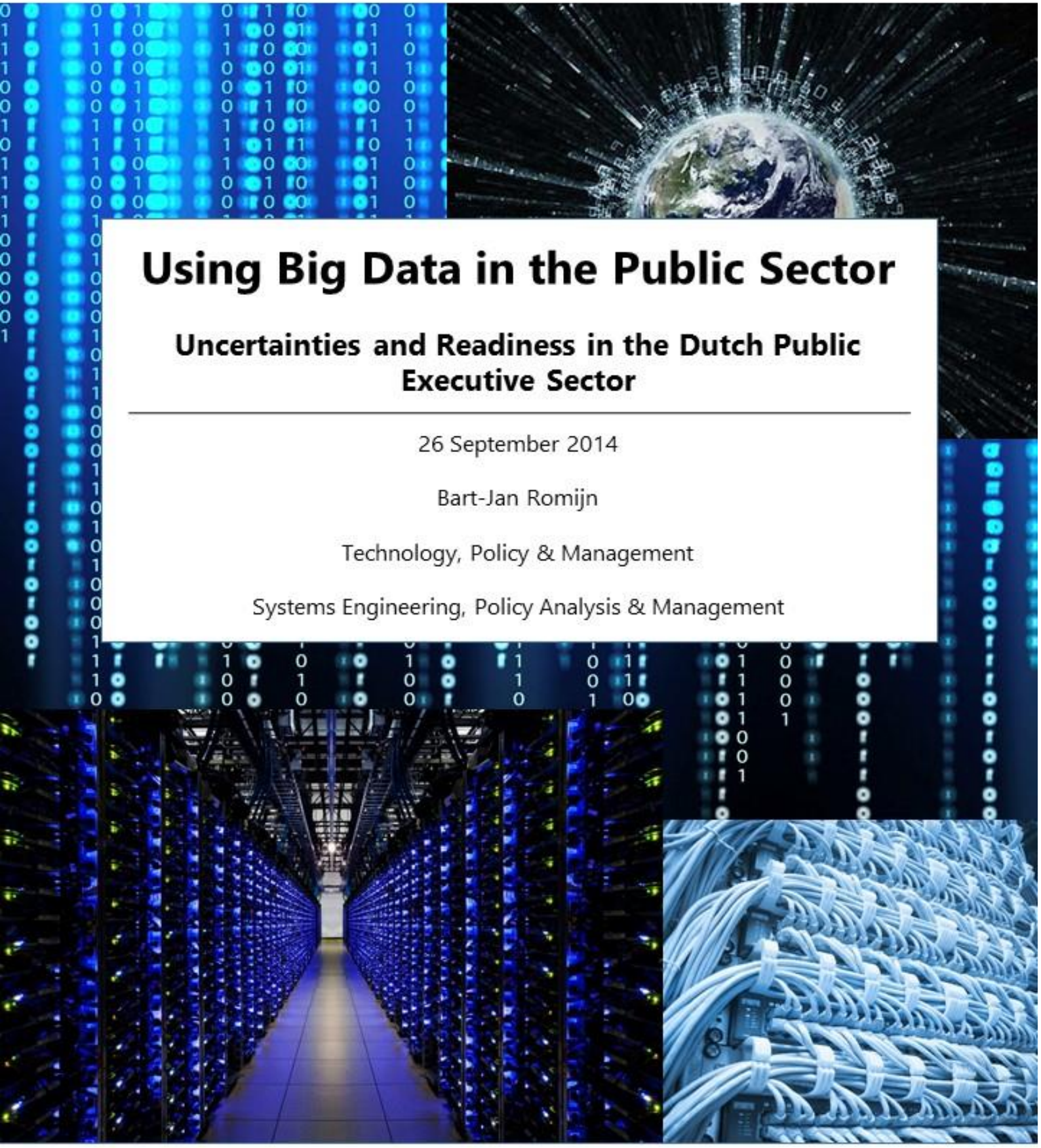
MSc Thesis Report



Using Big Data in the Public Sector

Uncertainties and Readiness in the Dutch Public Executive Sector

---

26 September 2014

Bart-Jan Romijn

Technology, Policy & Management

Systems Engineering, Policy Analysis & Management

TU Delft Delft University of Technology

Berenschot

Final Version Thesis Report

Using Big Data in the Public Sector

Uncertainties and Readiness in the Dutch Public Executive Sector

11 September 2014

Graduation Committee

Chairman:	J.A. de Bruijn
1st Supervisor	A.J. Klievink
2nd Supervisor	S.W. Cunningham
External supervisor:	G.A.J. Putman

Bart-Jan Romijn

1258958

Systems Engineering, Policy Analysis & Management

Faculty of Technology, Policy and Management

Delft University of Technology

Preface

The report *Using Big Data in the Public Sector* before you is the final product of a five month research project written for my graduation from Delft University of Technology. The Master Thesis Project represents the final part of my Master's degree Systems Engineering, Policy Analysis & Management at the Faculty of Technology, Policy and Management.

The assignment behind the project was created from a mutual interest in the development of Big Data in the Dutch public sector of Prof. Dr. J.A de Bruijn and Tanja Verheij at Berenschot. From this mutual interest an open and explorative graduation project was created including a graduation internship at Berenschot in Utrecht. In conjunction with Berenschot consultants multiple experts from practice were approached for this research project. Their input has been invaluable and I am therefore very thankful for the time and effort they have offered me and my graduation project.

Furthermore, I have received a lot of guidance, both practical and conceptual, from my graduation committee. I would therefore like to thank Bram Klievink, Scott Cunningham, Gerard Putman and Floris Bannink from Berenschot and Hans de Bruijn very much for the time, effort and attention they have given me during the last couple of months. You have helped me tremendously with my questions and difficulties throughout my project and have enabled me to write the report in the form it is now.

Executive Summary

The amount of digital data in the world is increasing very rapidly, but so are the power and possibilities of the technologies to handle all of that data. This development is being called the Age of Big Data and it is believed to have a lot of potential value for organizations from all sectors. The brief definition of Big Data used in this thesis report is that Big Data is data so large, varied and dynamic that it cannot be handled by conventional data processing technology.

Successful examples of Big Data initiatives are identified from both the private and the science sector, but the public sector seems to be falling behind. And this is strange, as the potential value of Big Data use for the public sector is estimated to be very high. Big Data use can have significant value for public sector organizations, as it can help them to improve efficiency, effectiveness and transparency, through better decision support information, faster and richer images from reality and better insights into citizen demands and needs.

Managers from the public sector are interested in Big Data, see its potential value, but so far, implementation of Big Data use is not taking place. Public officials are unsure whether their organizations are ready for the introduction of Big Data and if their organizations are able to take advantage of the opportunities offered by Big Data technologies.

The uncertainty around the organizational readiness for Big Data use is the main research problem addressed in this research project. Not being able to accurately assess the organizational readiness for Big Data use is problematic for public sector organizations, as this slows down development of potentially valuable Big Data use and increases the risk of premature Big Data use implementation that could cause significant problems for the future success of Big Data use in the public sector.

Interviews with public officials from the Dutch public executive sector are held to find the main uncertainties for practitioners around Big Data Readiness of their organizations. For the research project the Dutch public executive sector was chosen as research object, as this is assumed to be a leading edge of data use knowledge in the public sector and intensive work with substantive data is performed here every day for which the potential value of Big Data is expected to be very high.

From the interviews it became clear that the current situation in the Dutch public executive sector is that organizations are interested in using Big Data, are starting with projects for its development, have high expectations of Big Data use , but all have no clear path towards Big Data use implementation in their organization. The interviewed public officials indicate a number of uncertainties that exist on whether organizations in the sector are ready for Big Data use of not. The three main uncertainties indicated are:

1. Uncertainty on the form of Big Data use that is suited for the organization
2. Uncertainty on the maturity of the organization to change for Big Data implementation
3. Uncertainty on the capabilities of the organization to successfully use Big Data

The uncertainties will be used as the main starting points for a framework to accurately assess the Big Data Readiness of organizations in the Dutch public executive sector.

Furthermore, the interviews show that all organizations are struggling with the exact meaning of the concept of Big Data. The very dynamic and ambiguous definition from literature impedes understanding of Big Data and communication on it. In the thesis report an alternative method to describe Big Data is proposed; instead of trying to gauge what Big Data precisely is, the actual use of

Big Data in organizations is described. Firstly, the difference between conventional data use and Big Data use is described. Big Data use is different from conventional data use in five differentiating characteristics:

- a combination of internal and external data is used
- a combination of structured and unstructured data is used
- real-time streams of data are used
- advanced analytics and computing is used for the data
- data is used in new and innovative ways.

The Big Data use process is furthermore describe as a cyclical Big Data Value Chain with four distinct data use activities; Data Collection, Data Combination, Data Analytics and Data Use.

The five differentiating characteristics of Big Data use and the four Big Data use activities are used to create a Big Data Application Typology, based on the different knowledge the types are able to create, describing three Big Data application types; Object/Subject Evaluation, Research and Continuous Monitoring applications.

To help practitioners with the uncertainties they experience, a framework to assess the readiness of public organizations for Big Data use implementation is developed. The framework consists of three aspects, which are each based on existing models from academic literature, but adapted towards Big Data. The first aspect is Organizational Alignment and is based on the Strategic Alignment Model (Henderson & Venkatraman, 1993). The second aspect is Organizational Maturity and is based on e-government growth stage models. The model proposed by Klievink & Janssen (2009) is selected for the development of the framework. The third aspect of the framework is Organizational Capabilities and is based on the E-government Maturity Model (Valdès, Solar, Astudillo, Iribarren, Concha, & Visconti, 2011).

The organizational alignment is assessed by identifying the type of the public organization based on its main statutory tasks and current data activity. Then the planned Big Data application types are identified for each organization. Based on the current data activities of the organization and the distinct requirements for each of the Big Data application types, the alignment between the organization and the planned Big Data application(s) is established.

The organizational maturity of each organization is assessed by investigating their current information and activity sharing activities, IT facilities for that purpose and currently used data systems. The organizational capabilities of organizations are assessed for seven capabilities, specifically selected for their relevance for Big Data use. The seven capabilities are: IT Governance, IT Resources, Internal attitude towards Big Data, External attitude towards Big Data, Legal Compliance, Data Governance and Data Science Expertise.

The assessment of the Big Data Readiness of organizations in the Dutch public executive sector based on the developed framework indicates that the organizations currently are not ready to implement Big Data use. The Dutch public executive sector is not ready for Big Data, and additional development initiatives are required for all organization before implementation of Big Data use should be executed. Dutch public executive organizations have well developed capabilities to start the use of Big Data in their organizations, but lack understanding of the implications of Big Data applications on their organizations. Furthermore they lack structural data and information sharing arrangements and facilities, limiting the amount and variety of data available to them.

A main area for improvement for Dutch public executive organizations is to improve their understanding of the implications of Big Data applications on their organizations and when further developed, to start with the implementation of Big Data applications that fit their current activities and provide the type of knowledge they currently use in their daily operations. Another area for improvement is the development of structural information and data sharing arrangements and facilities with other organizations. This will enable the organizations to acquire the amount and variety of data needed for the input of their future Big Data applications, so that they can provide knowledge on the granular, individual level that is required to enable the public organizations to develop towards a more demand driven form of government.

Concluding, the Dutch public executive sector is currently not ready to take advantage of Big Data use and should first develop its organizations further, before Big Data use should be implemented. The Big Data Readiness Framework provides public officials from the Dutch public executive sector with a Big Data Readiness assessment and indications for areas for improvement. With this information practitioners will be better equipped to plan and take the road towards Big Data use implementation in their organizations and improve their organizations and society as a whole in the future.

Table of Content

List of Figures

List of Tables

1. Introduction

1.1 The Age of Big Data

With the technological and societal developments of the last decade, the amount of new digital applications and devices has sky-rocketed and with it the possibilities and value for its users. Digital devices and applications, such as mobile phones, websites, social media, smart household appliances, enterprise software packages, industrial machines and smart cars have the potential to be producers or sources of digital information. The produced information is very often in a very raw form and therefore referred to as data. The development that digital devices and applications are being used ever more extensively and that these devices and applications are (un)consciously producing more and more digital data is called datafication by Mayer-Schönberger & Cukier (2013).

Because of the datafication development, the amount of digital data produced is growing very rapidly. This trend is called the data deluge, as the vast and ever growing amount of digital data is flooding the world around us (The Economist, 2010). To give an indication of the growth of digital data Google's former CEO Eric Schmidt, stated in 2010 that we were already producing the same amount of data every two days as what we have done as a whole civilization up to 2003. He further added that that pace is speeding up (Kirkpatrick, 2010). Although the quote is refuted by several experts as an exaggeration, it does effectively show the exceptional growth of our data production in recent years. Currently the size of our total stored digital data is about 7,500 Exabytes, and data creation is expected to roughly double every year, which means we will go towards over 40,000 Exabytes (or 40 trillion Gigabytes) of total stored digital data by 2020 (IDC, 2011; IDC, 2012).

But what to do with all of that digital data? Fortunately, the exponential growth in data is parallel to the development in IT technologies, such as computational power and storage technology. The exponential development of computational power, described in Moore's law, provides us with the raw computing power to handle ever growing quantities of data. The cost of computing power and storage has developed too and has gone down significantly in the last decade, which makes dealing with the data deluge not only technologically feasible, but also economically viable (Bryant, Katz, & Lazowska, 2008; IDC, 2011). Next to the cost and performance developments of computer hardware, significant developments in the way to open, read, understand and interpret digital data and the many forms it comes in are made in the last decades. We now have smarter software tools, programs and algorithms to work with digital data. So, in short, next to the enormous flood of digital data, we now also have the means to deal with it (The Economist, 2010; KPMG, 2012; Almeida & Calistru, 2013).

The following graph (Figure 1) shows the previously described developments by depicting the amount of structured and unstructured data (for further explanation, see page 40) in the world between 2005 and 2022 in Exabyte and the hardware cost for consumers of 50Gb of storage technology and computational power comparable to a modern day, widely used high-end Intel I7 CPU unit. The graph (Figure 1) shows that we are currently in the period of time where the amount of digital data is really taking off and where the cost of IT is at a level that makes handling all of that data is becoming economically viable. In short, things are getting interesting.

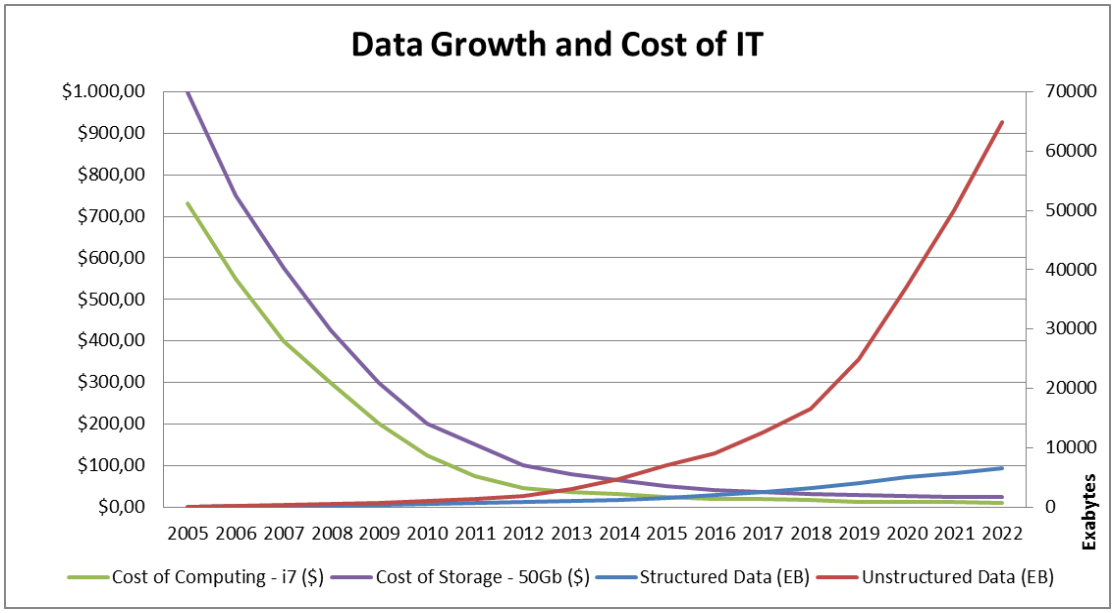


Figure 1: Data Growth and Cost of IT

The combination of the ever-growing, enormous volume of digital data and the technological advancements that enable us to economically handle that data provide us with a unique opportunity. Handling all of that data provides us with new insights through undiscovered patterns in the vast amounts of data, faster information discovery from additional data sources, and generating information by combining sets of data that were not compatible before (Simon, 2013). We are able to take advantage of all of that data and create value from it, rather than just seeing it as a burdening side-effect of technological progress. The co-existence of these two developments opens the road towards a revolutionary new paradigm of knowledge discovery (Mayer-Schönberger & Cukier, 2013). Some call it, perhaps somewhat dramatically, the dawn of the age of Big Data (Lohr, 2012).

Big Data is the term currently used to describe the enormous volumes of data that can now be stored, analyzed and used with the latest technologies. In this thesis the following definition of Big Data is used: Big Data is data so large, varied and dynamic that it cannot be handled by conventional data processing technology. An overview of the current definitions of Big Data from multiple authors is discussed in further detail in chapter four (page 35).

Big Data in practice

Although the promise of Big Data is almost unlimited according to some authors (Mayer-Schönberger & Cukier, 2013; McKinsey Global Institute, 2011), the benefit of Big Data has to be viewed with some reserve. But, there is little doubt that Big Data can have value for organizations in practice. How large that value will be however, remains to be seen and can only be established when more practical experience with Big Data is available.

Currently, a growing number of organizations in practice have started using Big Data and are (trying to) take advantage of it. According to a 2011 survey of The Data Warehouse Institute, 37% of surveyed private sector organizations are indicating they are using Big Data and even more are currently developing it (Russom, 2011). Big companies from retail, such as Walmart (Bryant, Katz, & Lazowska, 2008), Sears (Henschen, 2012), and Amazon (Kelly, 2013) are using Big Data to better

understand their customers and their buying choices. By combining customer information, transaction records of products, website behavior and geographical information they are trying to create a better understanding of customer needs, behavior and buying decisions. With these insights they hope to improve their operations, marketing and customer satisfaction ratings.

Financial institutions, such as Morgan Stanley (Groenfeldt, 2012) are using Big Data to better predict market behavior and investment opportunities to improve their performance. Online companies, such as Google, LinkedIn, Twitter and Facebook have created their entire business model on personalized advertising, and customer portfolios based on huge volumes of digital data on online behavior, expressed sentiments, information requests and stated preferences of individuals (Simon, 2013; Davenport & Dyché, 2013). Individuals are consciously and unconsciously increasingly sharing information about themselves online, feeding these organizations with more data every day.

Not only private sector organizations are starting to use Big Data. Big Data is enabling large science programs, such as the Large Hadron Collider in Geneva and the future Large Synoptic Survey Telescope in Chile to revolutionize physics and astronomy (Mayer-Schönberger & Cukier, 2013). Other examples where Big Data is important for science are biology research fields, such as genome analysis and neuroscience. Big Data technologies allow for fast and economic processing and analysis of the huge volumes of data involved, making these science programs feasible.

Big Data technologies have the potential to change the productivity and operations across industries, sectors and nations, in all different kinds of applications. Big Data can therefore be seen as a potential next General Purpose Technology, as it has the potential to develop all of its attributes. General Purpose Technologies are technologies that can be applied in almost all circumstances and over time increase productivity, possibilities and economic growth on a national or even global scale (Bresnahan & Trajtenberg, 1995). And Big Data is starting to do that in the private and science sector.

But, in contrast to the examples of Big Data use above, it seems that the public sector is currently not jumping onto the Big Data bandwagon. Although a 2013 Bloomberg survey shows that public officials have substantial interest in using Big Data in their organizations and see a lot of potential value (Mullich, 2013), the officials all indicate they are currently not using Big Data in their organizations. This image also seems to apply to the public sector in the Netherlands (Toorn, 2014). The public sector is looking into Big Data, but further steps towards implementation are currently not taken (Federal Big Data Commission, 2012; Ministerie van Binnenlandse Zaken en Koningsrijkrelaties, 2012).

Potential value of Big Data for the public sector

Although public sector organizations are currently not using Big Data, they do see a lot of potential value from using it. The value of Big Data for the public sector is therefore only theoretical at this moment. However, a number of authors describe the specific value of Big Data for the public sector in their work.

According to Bertot & Choi (2013), the value of Big Data use for the public sector comes from higher quality decision making and a better understanding of the environment. Citizen demands and developments in the environment are better understood through more information, which becomes available faster. With these benefits, knowledge discovery from Big Data may allow the public sector to make substantial progress in solving lingering societal problems, such as transport congestion, health care quality and costs, and the transition towards sustainable energy production (Scholl &

Scholl, 2014). Furthermore, it could allow government to structurally change and become much more demand driven in its service provision and in policy design (Policy Exchange, 2012). And this last change may just be the biggest benefit of Big Data for the public sector. By becoming more demand driven and better understanding that demand and the effects of government policies, Big Data can help government to become more effective, more efficient and more transparent, through more granular, precise information and new insights from combining data, which can now be done much faster and much easier (Joseph & Johnson, 2013). This structural change provides the public sector with a huge opportunity to make a big leap in its performance, which is strived for by many practitioners. Big data has the potential to be a stepping stone towards a new form of government, a form of government that sees IT not just as a complementary channel, but as a way to transform how government works (Dunleavy, Margetts, Bastow, & Tinkler, 2006).

The opportunity Big Data seems to offer to the public sector is very welcome, especially since government organizations across the globe have had to cut back in the past few years due to economic downfall and consequential austerity policies (Dunleavy & Margetts, 2010). Government organizations in the Netherlands are no exception to this and face substantial challenges because of it (Algemene Rekenkamer, 2013; Rijksoverheid, 2013). Next to cutting cost, higher demands for better performance are making these challenges even larger. In the last few years, citizen and national government demands on public sector organizations for higher quality, higher effectiveness and faster service and decision making are getting stronger (Scholl & Scholl, 2014). According to Mayer-Schönberger & Cukier (2013) Big Data is currently the best way to meet the challenges faced by public organizations today and in the future.

1.2 Background

As described, Big Data is credited it has the potential to improve and transform the way public organizations work, provide services and evaluate and design policies. Both the production of vast amounts of digital data and the tools to use that data are consequences from innovations in the IT around Big Data. Using Big Data can therefore be described as an IT innovation that enables organizations to do things they have never before been able to do, both in data collection and in knowledge creation from that data (Mayer-Schönberger & Cukier, 2013).

And that is no different for public organizations. Public organizations have been implementing and using new IT innovations to reform and improve their possibilities and performance for the last couple of decades, albeit with differing success, and Big Data is regarded as no exception to this trend. Public officials in the US (Executive Office of the President, 2014; Federal Big Data Commission, 2012), Australia (Department of Finance and Deregulation, 2013), and the Netherlands (Ministerie van Binnenlandse Zaken en Koningsrijkrelaties, 2012) are currently looking at the opportunities Big Data technologies have to offer.

Reforms that are supported or even enabled by IT have been researched and described extensively in academic literature. And reforms enabled by IT in the public sector are the focal point of e-government literature (Cordella & Bonina, 2012; Heeks, 1999). It is therefore not surprising authors have begun to connect Big Data with e-government theory (Bertot & Choi, 2013). This connection is also made in this research project. Big Data is seen as an IT innovation that can enable public sector organizations to significantly develop and improve their possibilities and performance. Therefore, concepts and knowledge from this field of research will be used in this research project to explain

the connection between Big Data and public sector organizations and the development of these organizations through the implementation of Big Data IT.

However, as we have seen, the development of Big Data has the potential to go further than just developing and improving the possibilities and performance of public organizations. According to a number of authors, the potential impact of Big Data is so large that Big Data should not only be seen as an enabler for reform and performance improvement, but also as an enabler for much more radical and structural change in the public sector. Because of its potential, Big Data can be regarded as an IT development that helps government organizations substantially improve and structurally reform towards the end-state of e-government development, so-called digital-era or transformational government (Margetts & Sutcliffe, 2013; Joseph & Johnson, 2013; Rajagopalan & Solaimurugan, 2013). This statement is supported by the potential of Big Data use technologies to radically advance government performance through lower cost (Milakovich, 2012), higher quality and faster insights in the environment and citizens demands (Policy Exchange, 2012), resulting in better and faster decision making (Puron-Cid, Ramon Gil-Garcia, & Luna-Reyes, 2012), policy evaluation, design and execution (Bertot & Choi, 2013) and an increase in transparency (Dawes, 2010). Furthermore, Big Data use has the potential to improve effectiveness, efficiency and transparency in public organizations (Joseph & Johnson, 2013) and that is in line with the goals of t-government development (Weerakkody, Janssen, & Dwivedi, 2011).

T-government is seen as the final stage of e-government development and is a reaction to the limited success of e-government initiatives, as they have not always delivered on their potential. E-government initiatives have helped the development of government organizations by providing new instruments and possibilities for their operations and communications. E-government initiatives are undertaken as they are believed to increase the efficiency, effectiveness and transparency of government organizations (Bertot, Jaeger, & McClure, 2008; Gilbert, Balestrini, & Littleboy, 2004; Scholl & Klischewski, 2007). However, most e-government programs have only delivered partially on these goals (Andersen, Henriksen, & Medaglia, 2010). Actual structural changes towards states similar to t-government, are required for further development of governmental organizations, in order for them to be ready for their changing environments (Weerakkody & Dhillon, 2008). Through these structural changes, the initial goals of e-government initiatives can be better achieved.

Because Big Data could have the potential to change organizations, it is considered to be not only an enabler of e-government development, but also an enabler of the development towards t-government, which could radically change public sector organizational structure, focus and performance. This consideration shows the significant value Big Data technologies potentially have for public sector organizations and shows why successful implementation of these technologies can be very important for the public sector. Lastly, the goals of e-government, enforced by t-government, improvements in efficiency, effectiveness and transparency will be used as starting points in the remainder of this research project to express the potential benefits of Big Data technology implementation.

1.3 Research Problem

So, although Big Data holds substantial potential value for public sector organizations and public sector managers are interested in using Big Data, implementation of Big Data use is currently not taking place (Toorn, 2014; Mullich, 2013). Of course, immediately the questions rises why Big Data is currently not used in the public sector. Especially since organizations in the private and science

sector are using Big Data and seem to have success with it. Malik (2013) and the Dutch Ministry of Internal Affairs (2012) provide a clue why public organizations are currently hesitant to start using Big Data. Malik states that many organizations, in both the public and private sector, are currently adopting policies that promote data intensive decision making in their organizations, but that the step towards using large and more varied Big Data is still contemplated by many as they are unsure if their organizations are ready for it. The Dutch Ministry of Internal Affairs describes that a lot of work still has to be done, before organizations in the Dutch public sector are ready to start using Big Data. The Dutch Ministry furthermore states that further development and experimenting is needed before organizations under the Ministry are ready to start using Big Data.

So, while organizations in all sectors seem to struggle with the question whether they are ready to start using Big Data or not, a very different approach is taken in the different sectors. While private sector organizations are taking the leap and start with using Big Data, public organizations are more careful and believe they should develop their organizations further, before Big Data use can be implemented. And not without good reason. Big Data has significant potential for society, but it is not a miracle cure that can be instantly applied when and where needed. Implementation of Big Data requires substantial effort, time and investment to change the organization (Brown, Court, & McGuire, 2014). And that can be very difficult to do. For example, recent enquiries into public sector IT project performance shows that the public sector, at least in the Netherlands, has a lot of trouble to effectively implement promising IT solutions (Algemene Rekenkamer, 2007; ANP, 2014) and there is no reason to believe Big Data technologies are an exemption to this.

Careful planning and effective measures for implementation are therefore crucial for public organizations to really be able to take advantage of the promises Big Data offers. Public sector organizations should not dive head first into the flood of data that is starting to become available to them. They should make sure they are ready to use Big Data, before they decide to start with using it. The two Dutch Ministries mentioned earlier state that they observe that much work has to be done before organizations can start using Big Data. They think public organizations in the Netherlands are not ready yet and should develop further, before Big Data can be used by them.

Although this may be a fair assessment, it remains unclear just how far developed the organizations in the Dutch public sector actually are in regard to using Big Data. And, perhaps even more importantly, it is unclear what still has to be done in development, before organizations in the Dutch public sector can start using Big Data. This problem is further enhanced by the fact that Big Data is a very new concept that is still under development. Scientific and practical knowledge on Big Data is still under development and no Big Data specific information is available that can help public organizations in the assessment of their readiness for using Big Data and to identify the areas in which they should improve. This brings us to the main research problem that is the motivation for this research project.

> *Main Research problem:*
>
> *It is unclear how ready public organizations are to start using Big Data and in what areas they should improve before Big Data can be used by them to help them develop their performance and help them change structurally to be able to successfully meet their (future) demands.*

The research problem stated above is a practical problem for which currently no clear cut solution is available. The solution of this problem is to develop a framework to help assess the readiness of public organizations to start using Big Data and identify the areas in which the organizations should

improve. However, the readiness assessment framework cannot be developed without addressing two other, sub research problems.

The first sub-research problem is that it is unclear what the current situation regarding Big Data in the public sector exactly is. An overview of current public sector Big Data views, experiences and activities is not available and is a vital part of the assessment of the readiness for Big Data of public organizations, as it will show current knowledge, issues, uncertainties and goals regarding Big Data in the public sector. The second sub-research problem is that Big Data still is quite an ambiguous concept. While the definition of Big Data itself may be reasonably described and is receiving attention by a swiftly growing number of authors, the description of actually using Big Data in practice currently is not. What using Big Data in practice entails exactly is still ill defined and, while partly described by several authors and practitioners, not much useable knowledge is available.

Solving the main research problem will require solutions to the two sub-research problems first, as an accurate assessment method can only be developed if it is based on current information from practice and a thorough understanding of Big Data use. As the research problems introduced in the section all deal with a large number of uncertainties and open questions, the nature of the research project is exploratory. This approach allows for information found in earlier research steps to shape the following steps, which is, as described above, necessary to solve the main research problem. The structure of the research problems is presented in Figure 2.

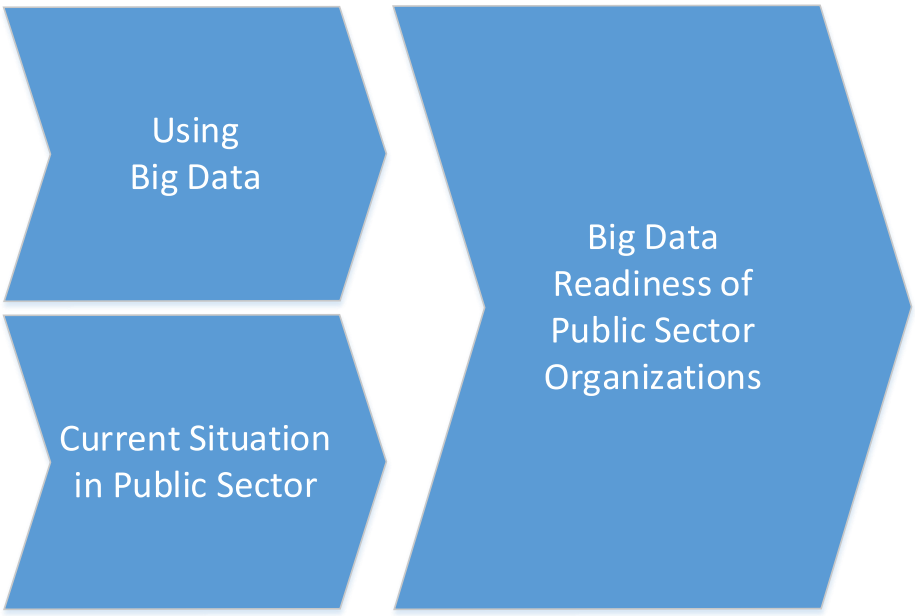


Figure 2: Research Problems

Added value of the research project

The research project presented in this thesis has added value in two distinct directions. First, the practical value of the research project is described, followed by the academic value.

The added value of the research project for practice has three distinct parts. Of course, the first part is that public sector officials will gain insights on how far the readiness of public sector organizations has developed and what the main areas of improvement are. Additionally, the overview of current Big Data activities and uncertainties at public organizations will provide a first benchmark for other

public organizations to compare themselves with. Secondly, this research project will provide a framework that will help practitioners in the public sector to assess whether their own organizations are ready to take steps towards the use of Big Data. The framework will furthermore provide information on which areas of their organizations should be developed further to improve the readiness of the organization for the use of Big Data. Thirdly, the research project will provide practitioners with knowledge of what using Big Data exactly entails for their organizations, which will greatly enhance understanding and communication of the Big Data concept in their organizations. With these three additions, the research project will enable practitioners to contribute to a more successful and fastened implementation of Big Data use and its potential value in the public sector, resulting in better performing and more demand driven government organizations.

The second direction in which this research project adds value is in academic knowledge on Big Data. Many articles and reports on Big Data are currently being published and the field seems to be developing fast. However, very few of these additions to the field are based on broad experience from practice and even fewer on experience from the public sector. By connecting knowledge from practical experience in the public sector and existing academic knowledge, the concept of Big Data and more specifically the use of it in organizations is demystified further. A new approach to describe both the potential value and the organizational impact of Big Data use is proposed here. This will contribute to the existing knowledge and understanding in the field. Furthermore, by coupling existing theoretical frameworks in literature, not initially intended to support Big Data, to the concept, Big Data theory will be grounded stronger within academic knowledge. Lastly, the framework that is proposed to assess the readiness of public organizations to start using Big Data is will provide further academic knowledge on how public organizations can be assessed in light of technological innovations. The Readiness assessment is a contribution to literature on Big Data, as the assessment provides an overview of important aspects related to Big Data and their position in regard to it. This overview provides additional insight into the prerequisites and implications Big Data brings to organizations and society as a whole. Also, this research project further clarifies the role Big Data can play in the development towards the t-government state in public organizations, by presenting in more detail the added value applications of Big Data can bring these organizations.

1.4 Structure of the Thesis

After the introduction that is presented in the chapter above, the thesis report continues with the research chapter. In this second chapter, the research objectives, objects, used research methods, and approach is presented. In chapter three, the current situation regarding Big Data in the Dutch public executive sector is described. Specifically the uncertainties that practitioners face when trying to assess if their organizations are ready for Big Data or not are established. Chapter four contains an explanation of the use of Big Data within organizations. The characteristics of Big Data use, its activities and three Big Data application types including specific added value are presented here.

In chapter five the framework to assess Big Data Readiness in public organizations is presented. The three parts of the framework, Organizational Alignment, Organizational Maturity and Organizational Capabilities will be further explained here. Chapter six continues with the results of the application of the framework to organizations from the Dutch public executive sector. The results are followed by a discussion in chapter seven. Chapter eight provides concluding remarks and recommendations on Big Data Readiness and its assessment. The thesis report is concluded by a reflection on the research project in chapter nine.

2. Research

In the following chapter the characteristics of the research project are presented. The main research objectives are derived from the described main research problem and two sub-research problems from the previous chapter and are presented in the first section of the research chapter. Secondly, the research objects that were examined for the research project are described and its selection explained. Thirdly, the research questions that guide the research project are formulated and presented. Finally, the research approach and research methods are presented in the final section of this chapter.

2.1 Research Objectives

The first research objective of this research project is to establish whether insufficient readiness for Big Data use in organizations in the public sector is a reason why Big Data is currently not used in the public sector. Additionally, the research project is conducted to identify the possible areas for improvement. From this overview new insights on the readiness and use of Big Data in the public sector can be established that can help practitioners in the public sector to be more successful in their attempts to develop their organizations towards using Big Data and to take advantage of its potential value in the future.

As described in the section on the research problem on page 15, the main research problem and the corresponding first research objective are primarily focused on practice. However, a second research objective is set that focusses on the academic added value of the research project. The second research objective is to demystify the concept of Big Data further, by describing Big Data and its use in (public) organizations. By adding further handles to the concept of Big Data and its use, this research project aims to create further insights into how Big Data can be used in organizations in the public sector and what implications this will have. This will allow both practitioners and academics to create a better understanding of the role of Big Data in organizations and the public sector in general.

2.2 Research Object

This section of the research chapter describes the research object that is investigated to achieve the research objectives set for the project. Firstly, the planned research object is described, followed by the resulting conduct on the selection of the research object.

The planned research object for this research project will be the public sector. Of course, it is not possible to investigate the whole global public sector and therefore a selection is made from the public sector. The selection is made by choosing one country to investigate in more detail. This is done to minimize communication issues, especially in light of such a complicated and new concept as Big Data, and to have similar technological, administrative and political circumstances for all respondents from the public sector of that nation. For this research project the Netherlands was selected, as it provides a very developed and stable public sector and is one of the leading countries in the world in IT facilities and usage. Also, the Netherlands scores very high in e-government indexes, such as the UN e-government survey 2014 (United Nations Department of Economic and Social Affairs, 2014). It is assumed that because of these attributes the Dutch public sector is one of the most advanced public sector regarding the development and interest in Big Data technologies.

To create a representative image of the interests, knowledge and experience regarding Big Data in the Netherlands, a group of public officials from different organizations within the public sector in the Netherlands is chosen as respondents for the research project. It was furthermore chosen to approach public officials from organizations that possibly have an advantage in expertise and knowledge on Big Data and data use in their organizations. Additionally, these organizations are likely to be able to contribute the most valuable information to the research project and are most likely to currently encounter uncertainties regarding Big Data readiness. In short, the research project would most benefit from information from organizations in the Dutch public sector that represent the leading edge of Big Data knowledge. It is assumed these organizations would be most likely to be found in a branch of the public sector where organizations are structurally working with a large amount of substantive data. This assumption is made on the basis that Big Data technologies are possibly able to provide the most value with the input from large amounts of data on substantive objects or subjects. And that therefore organizations working with a lot of substantive data, are most interested in Big Data use implementation and are most knowledgeable on it.

Organizations working with large amounts of substantive data can best be found in the executive branch of the public sector, as they are working directly with citizens, private sector, environment and other objects and subjects in reality. They therefore have a large amount of substantive data available to them and have experience and knowledge on working with that data. It is therefore assumed organizations in the Dutch public executive sector are most likely to be working on Big Data use implementation and can contribute valuable information to this research project. Another criteria that was considered in the selection of Dutch public organizations to approach for this research project is that the participating organizations should be (part of) larger organizations, with substantial internal IT expertise and facilities, as they are assumed to have a lead in developing towards Big Data implementation.

The planned selection criteria mentioned above were used to find and select public executive organizations in the Netherlands. These organizations were approached on the basis of the criteria described above and the selection was added upon with public organizations that were mentioned in multiple media outlets as organizations working on very advanced data applications. Also, some organizations were approached as they fall under the coordination of Ministries that are actively working on Big Data orientation activities for their agencies (Ministerie van Binnenlandse Zaken en Koningsrijkrelaties, 2012; Ministerie van Infrastructuur en Milieu, 2014). Not all approached organizations were willing to cooperate in the research project, but a sufficient number of public officials was able to participate.

The process of finding, selecting and contacting led to that representatives of eleven organizations from the Dutch public executive sector participated in the research project. The participating representatives come from the organizations presented in Table 1. More information on the public officials and their organizations can be found in Appendix I on page 129. Combined, the eleven respondents grant us a representative image of the current situation in the Dutch public executive sector as a whole. This will allow that the results of this research project can give a clear view of the Big Data Readiness of organization in the Dutch public executive sector and that conclusions and recommendations can be formulated that apply to the entire sector.

Organizations of participating representatives	
Zorginstituut Nederland	Kadaster
Gemeente Utrecht	Rijkswaterstaat
Gemeente Dordrecht	Planbureau voor de leefomgeving
Dienst Wegverkeer (RDW)	Centraal bureau voor de statistiek
Brandweer Amsterdam-Amstelland	Landelijk informatiecentrum voor voertuigcriminaliteit
Koninklijk Nederlands meteorologisch instituut	

Table 1: Organizations of participating representatives

2.3 Research Questions

In this section of the research chapter the research questions are presented that are answered in this research project. The questions formulated here guide the content that is addressed in the research project. As the research project uses an exploratory approach, the sub-research questions are sequentially answered and each answer provides input for the following research questions and their respective answers. The answers to the three sub-research questions will provide a framework that is used to analyze the situation is practice and will help to answer the main research question.

Main Research Question:

> *What is the organizational readiness for Big Data use in the Dutch public executive sector and where can it be improved?*

Sub-Research Questions:

1. What is the current situation in the Dutch public sector regarding Big Data?
    a. What are organizations in the Dutch public sector currently doing regarding Big Data?
    b. Which uncertainties are experienced by organizations in the Dutch public executive sector in assessing their organizational readiness for using Big Data?

2. What does using Big Data entail for organizations in the Dutch public executive sector?
    a. What makes using Big Data different from using more conventional data in Dutch public executive organizations?
    b. How does the process of using Big Data in organizations in the Dutch public executive sector look like?
    c. For which applications can Big Data be used by organizations in the Dutch public executive sector and what is their specific added value?

3. How can the organizational readiness for using Big Data of organizations in the Dutch public executive sector and its potential areas of improvement be assessed?
    a. How can the uncertainties in assessing organizational readiness for using Big Data experienced in practice be addressed is the assessment method?
    b. Is the assessment method suited to answer the main question of the research project?

2.4 Research Plan, Approach & Methods

In the following section the research approach and the respective research methods are explained. First the initial research plan is briefly discussed, followed by a more detailed description of the actual conduct during the research project. The conduct will consist of a brief general overview of the research approach, followed by a more detailed description of the three different research steps. The research project was ultimately completed in three steps, as knowledge developed in previous steps was used as important input for the following steps.

The initial research plan is presented in blue in Figure 3. It is a consecutive research plan, as the research project is set up as an exploratory research project. Firstly, it was planned to conduct a literature research on the concept of Big Data and its implications. However, it soon became clear that the usefulness of Big Data specific literature was quite limited for the scope of this research project. The literature lacked public sector specific Big Data contributions and was describing Big Data at a very conceptual level. The second planned step was to hold interviews with public officials from the Dutch public executive sector. The interviews further confirmed the view that Big Data literature had limited usefulness for the research project. However, the interviews also gave indications to other literature that may be useful and showed a number of uncertainties in practice that were not anticipated earlier. Thirdly, the research plan required the development of an assessment method for Big Data Readiness in the Dutch public sector. Now, as anticipated, Big Data literature did not prove to be helpful in this regard, but due to the unexpected discovered uncertainties, additional literature was required to formulate the assessment method for the readiness for Big Data use in the Dutch public executive sector. The following planned step was to use a questionnaire to acquire the information from practice needed in the assessment method to establish the Big Data Readiness.

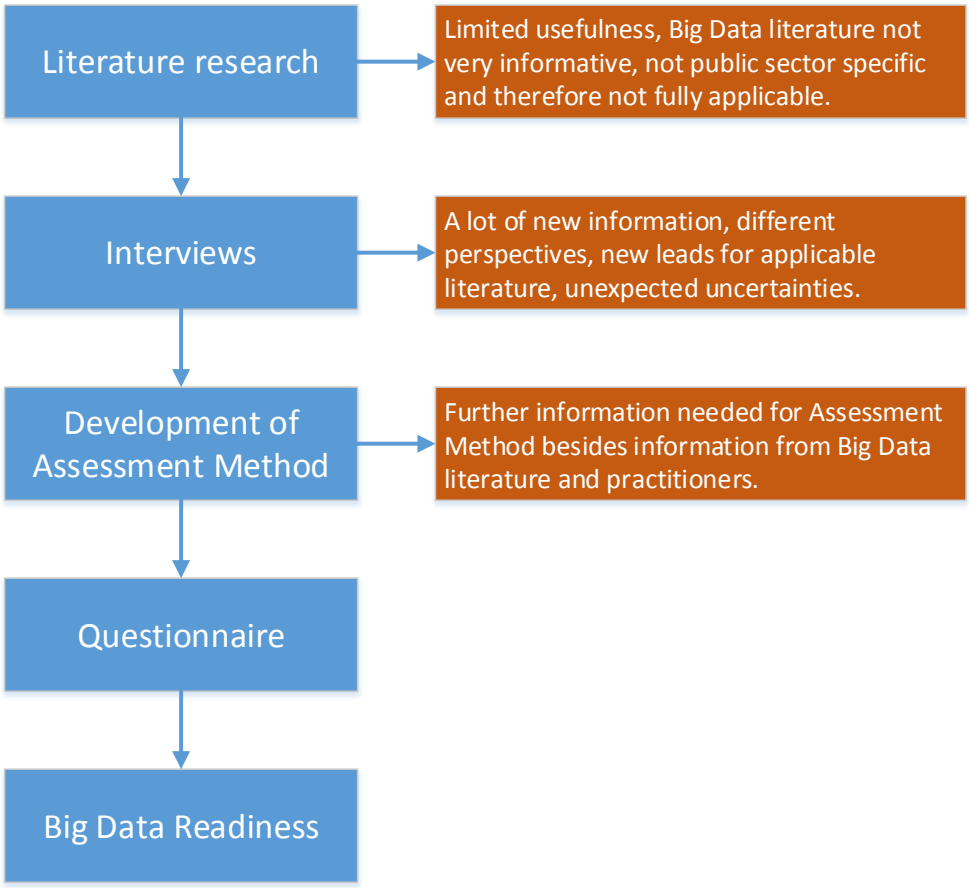


Figure 3: Initial Research Plan and Experienced Problems

The first research step conducted in the research project was an exploratory step, in which very diverse information was gathered on Big Data use and the current situation regarding Big Data in the Dutch public executive sector. In this research step new insights were used to gather additional information in new avenues of exploration, until the required insights for the next research step were obtained and sub-research questions one and two could be answered.

The second research step was the formulation of a framework that can be used to assess the organizational readiness to start using Big Data in the Dutch public executive sector. The framework was developed with input from the knowledge developed in the first exploratory step of the research project.

The third research step was the actual assessment of the Big Data Readiness of organizations in the Dutch public executive sector with the help of the framework that was formulated in step two. By gathering information from practice and analyzing that with the help of the framework, the organizational readiness for Big Data use and possible areas for improvement were established in this research step. The findings from this final research step will form the answer to the main research question of the research project.

In Figure 4 a graphical representation of the conducted research approach is presented. The research approach is described in more detail at each of the three research steps on the following pages. The different research methods are also described at their respective research steps.

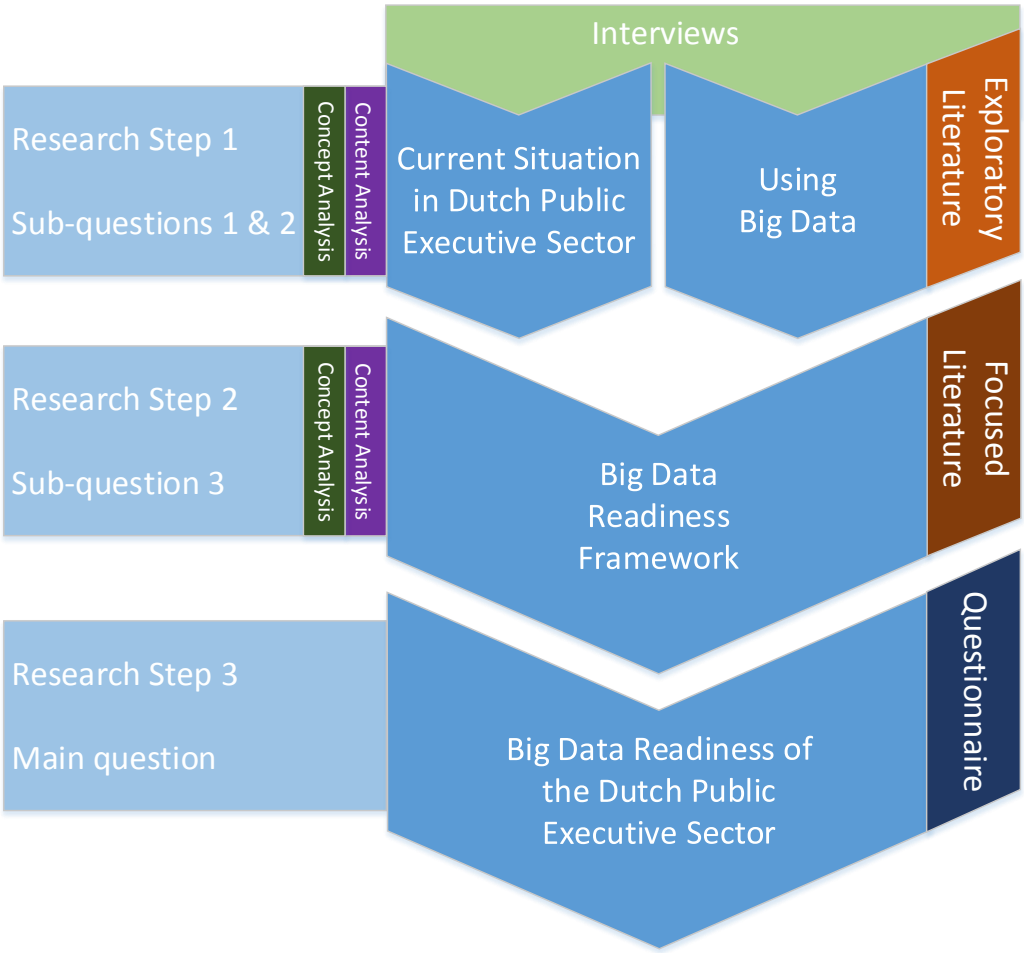


Figure 4: Research Approach

2.4.1 Research Step 1: Exploration

For the first, exploratory part of the research project, an exploratory literature research was conducted during the first research step. Firstly, the findings from the exploratory literature research were used to establish a conversation topic list for the unstructured interviews. Also, the first findings from literature on Big Data were used to establish a preliminary view of the concept and its consequences for public sector organizations. Secondly, as the exploratory interviews were held and new avenues of exploration were identified, corresponding literature was researched to further enhance insight and knowledge of Big Data and more specifically the use of it by organization in the public sector. As stated earlier, the exploratory literature research was conducted with input from the interviews, as the preliminary literature research on Big Data literature was limited in its usefulness for the research project. By using leads from the interviews the information from literature could be extended and became more useful and applicable to the research project.

Parallel and intertwined with the exploratory literature research, unstructured interviews with the representatives of the eleven participating organizations were held. The interviews were based on findings of the initial literature research on Big Data specific literature. These interviews were aimed to gather information on the current situation regarding Big Data in the Dutch public sector and to find the main uncertainties that were impeding the decision to start using Big Data in the participating organizations. Secondly, the interviews were used to discuss the actual use of Big Data with practitioners in the public sector. By discussing differentiating factors between using Big Data and conventional data, data activities important for Big Data use, applications of Big Data and its

potential value for the organization, an image of what using Big Data means for public organizations can be created.

For the formulation of the different products of the first research step, which are presented in chapter three and chapter four, the transcripts of the interviews and found literature was analyzed using content and concept analysis techniques. Content analysis, the process of counting the occurrence of certain terms, words, or concepts in articles and in this case also interview transcripts was used to establish an overview of the activities and interests regarding Big Data at Dutch public executive organizations. This analysis method was also used on the interview transcripts to identify the main uncertainties in the Dutch public sector on Big Data readiness. Furthermore, content analysis was used on Big Data specific literature to identify the most commonly presented Big Data related activities and differentiating characteristics. In combination with a content analysis on the interview transcripts this information was used to formulate five Big Data characteristics and four Big Data activities in chapter four. Finally, three types of Big Data applications are formulated with the use of concept analysis, the process of creating classifications of objects through differences in its attributes. By looking at all different planned Big Data applications discussed in the interviews, a classification, or typology of Big Data applications could be formulated. It is presented in chapter four.

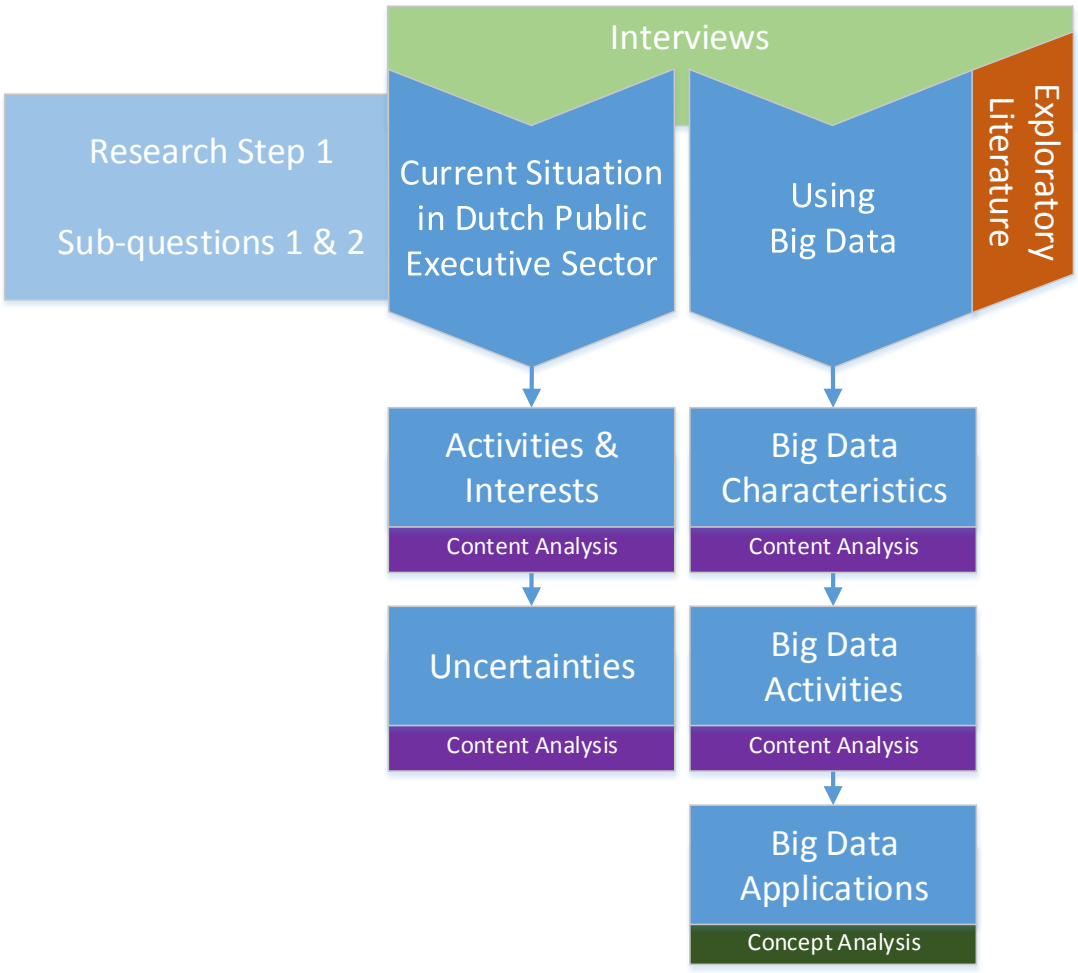


Figure 5: Research Approach – Research Step 1

Unstructured interviews

For the exploratory interviews, unstructured interviews with staff members of the described organizations in the Dutch public executive sector were conducted. These interviews were unstructured, as they were only guided by a brief set of conversation topics, which can be found in Appendix II on page 132. This way of interviews allowed sufficient leeway for exploratory conversation and specific topics per organization, which was chosen to be able to create rich images of the current perspectives, expectations, views, activities and uncertainties regarding Big Data from the representatives of the participating organizations.

Drawbacks of the unstructured interview research method are that the interviews allow too much room for different interpretation of conversation topics, questions and answers and that it is very hard to check whether all relevant information available from the interviewees is indeed discussed and recorded. Also, the comparability of the interview results is limited, as each interview in conducted differently. This, however, is not a problem, as the interviews are conducted as part of the exploratory research step and the gathering of information and development of insights is the priority in this research step.

Exploratory literature research

Next to the unstructured interviews, an exploratory literature research is conducted to provide an overview of the current relevant literature on Big Data and its use in organizations. The literature review was conducted with the help of search tools, such as Scopus, Google Scholar and Web of Science. Search terms used are all variations with the following terms*: big data, big data analytics, internet of things, data science, data mining, business intelligence, analytics, data driven decision making and datafication*. Furthermore, the reference lists of relevant articles and reports were used to further identify relevant and interesting literature.

All found articles and reports were assessed on relevance by scanning title, keywords, year of publication and abstract. Relevance of found literature was positively assessed if they were recent (post 2004) contained descriptions of the concept of Big Data or related concepts and that focused on the use and development of the concept of Big Data in organizations. Also articles handling the potential and the challenges of Big Data were considered to be relevant. More technical articles, focused on Big Data hardware and software or Big Data in specific fields completely unrelated to public sector organizations were not considered relevant for this research project. All relevant articles were thoroughly read to establish the conversation topic list for the unstructured interviews and to establish an overview of Big Data use in organizations.

One of the drawbacks of the literature review research method is that it is very difficult to establish whether all relevant literature has been found or not. Furthermore, a literature review only provides a snapshot of currently available and published knowledge, which possibly leaves out the latest insights and information.

2.4.2 Research Step 2: Formulation

For the second research step, the information gathered in the first, exploratory research step is used as an input to formulate a new framework that can help to assess the readiness of public organizations to start using Big Data. The two main inputs for the new framework from the first research step are the current uncertainties on readiness for Big Data use in the Dutch public executive sector and the description of the use of Big Data in public organizations. However, these

two inputs do not provide sufficient insight to formulate a framework that can help assess Big Data Readiness of public organizations. Therefore, additional information is needed. To gather this information, a more focused literature research was conducted to find literature that could help to mitigate the uncertainties on Big Data readiness assessment in the first research step and that could accommodate the concepts that are established for using Big Data.

With the help of the focused literature research suitable, established frameworks were found in literature that could be adapted towards Big Data readiness assessment and that provided the needed insights to be able to formulate a comprehensive framework that can help in the assessment of Big Data readiness in public organizations. The alignment of Big Data applications with organizations assessment aspect of the Big Data Readiness Framework was formulated with the help of an optimal alignment setup, that was formulated with the help of a concept analysis of the planned Big Data applications and organizational attributes on data activities and strategy discussed in the interviews. The maturity assessment aspect was also formulated with the help of concept analysis. By analyzing the attributes of the maturity stages from literature and the attributes of the use of data systems in practice discussed in the interviews, a connection between the two concepts could be formulated. The third assessment aspect of the framework, the capabilities are selected with the use of a content analysis of both the interview transcripts and on the found relevant literature.

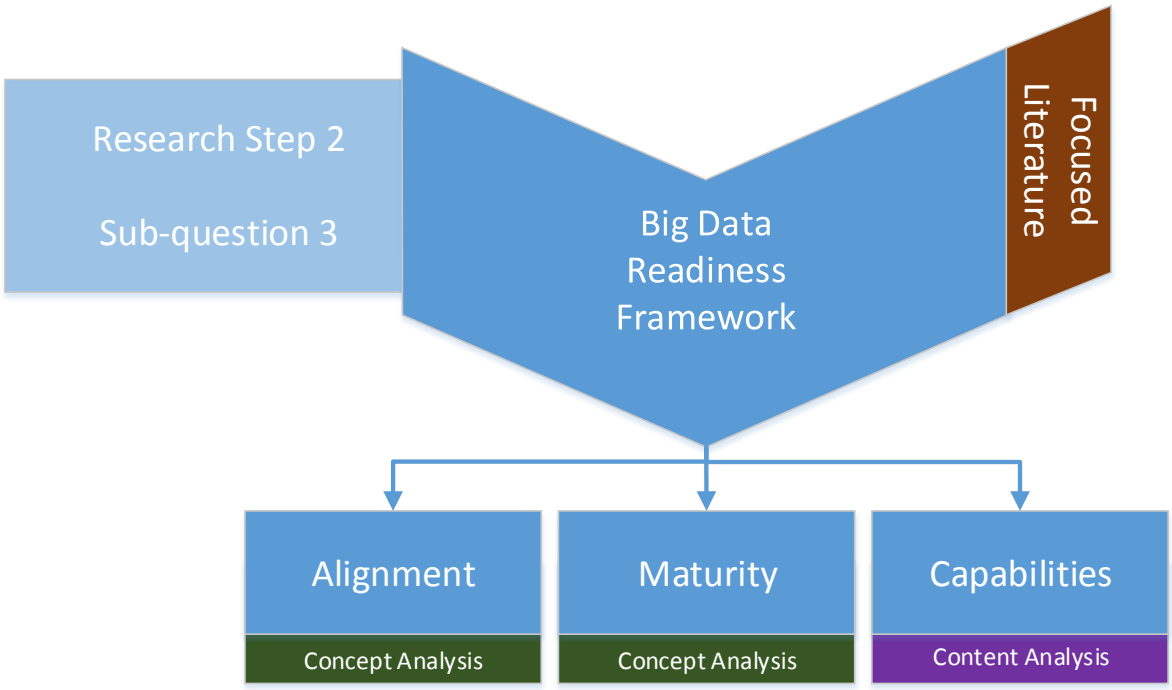


Figure 6: Research Approach - Research Step 2

Focused Literature Research

Just as the first exploratory literature research, the second, more focused literature research was conducted with the help of search tools, such as Scopus, Google Scholar and Web of Science. This time, the search terms used were not Big Data related, but more geared towards related, but more established concepts in literature. The search terms used were combinations of related theoretical concepts and technological terms, such as IT, IS or innovative IT solutions that are more established

than Big Data. The theoretical concepts are connected to the scope and perspective of the research project, and were inspired by remarks from practitioners in the interviews. The first set of technological terms (1) are broad terms, describing information technologies in general, the second set of technological terms (2) are technologies that share certain characteristics with Big Data technologies in areas such as knowledge discovery, connecting and sharing information, bridging information between separate information silos and handling data.

Theoretical concepts: *(organizational) readiness, (strategic/organizational/IT) alignment, (organizational) maturity, e-government maturity, t-government, (dynamic/core) capabilities.*

Technological terms*: (1) IT, IS, ICT, information technology, information systems, and (2) business intelligence, database management, EDI, ERP, DBMS*

Found articles were assessed on relevance by scanning the titles, keywords and abstracts. Articles with very similar theoretical bases were taken together and the initial articles proposing that theory were selected. Relevance was based on if articles had perspectives that were compatible with the perspective of this research project. The relevant articles had to describe organizational aspects, not individual or technical, had to describe the interaction between technology and organizations and had to consider that the technology was new and radically different from what the organization was currently doing.

The frameworks from the relevant articles were then compared with the established uncertainties on Big Data readiness from the unstructured interviews to see if they were compatible. Also, the relevant framework were checked for compatibility with the description of Big Data use in public sector organizations established in the first research step. Frameworks that did not survive both checks were discarded. Finally, the remaining frameworks were assessed on their usability and adaptability towards the assessment of Big Data readiness of public organizations and the insight they could provide on this subject.

Lastly, the focused literature research method was used to find additional information for the formulation of a comprehensive list of capabilities important for using Big Data in a public organization. Search terms such as *data governance, IT governance, IT systems adoption and IS adoption, implementation and innovation diffusion* were used to find empirically focused articles that described the most important factors, capabilities and characteristics of organizations that best predicted the successful adoption/implementation/diffusion of a new IT innovation in organizations.

2.4.3 Research Step 3: Assessment

In the third step of the research the developed framework from step two is used to establish the Big Data Readiness of the eleven participating organizations, provide an overview of the Big Data Readiness in the Dutch public executive sector and formulate recommendations to improve Big Data Readiness in the organizations. For the assessment of the Big Data Readiness of the eleven organizations, the three different aspects of the framework are used for the assessment first and then combined to create a full image of the readiness per organization. By combining the eleven readiness assessments an overview of the readiness in the Dutch public executive sector can be established.

The information required for the assessments based on the developed framework is gathered from the participating representatives with the use of a questionnaire. An overview of the research approach of the third part of the research project is presented in Figure 7.

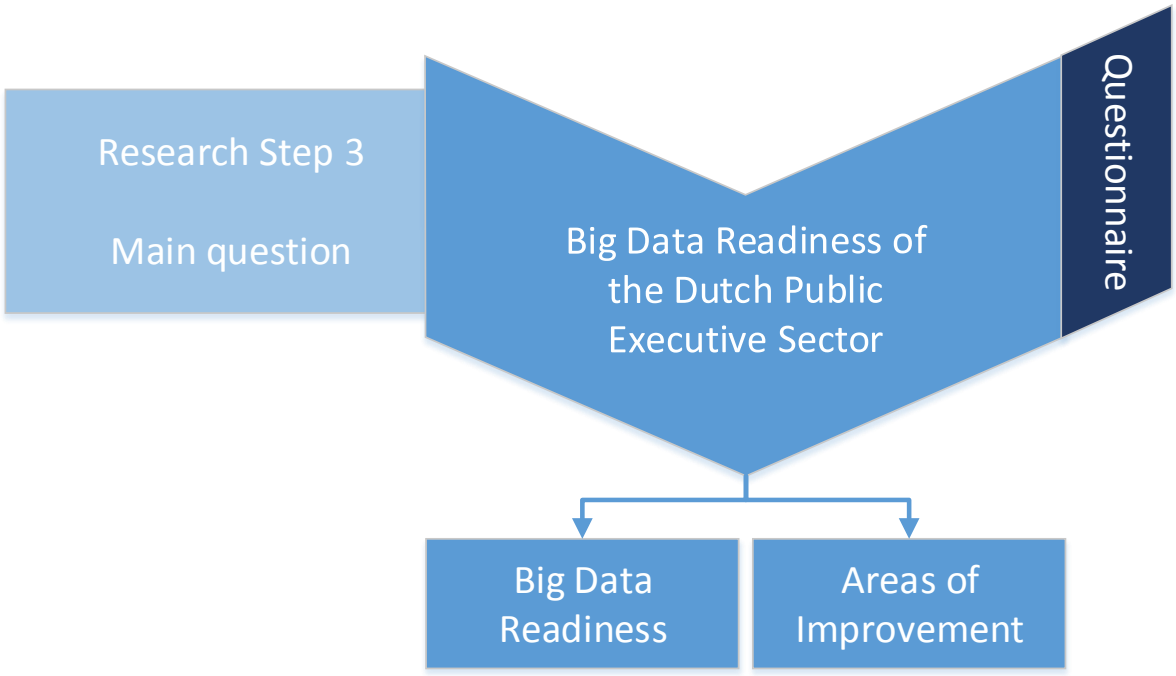


Figure 7: Research Approach – Research step 3

Questionnaire

The third and last research method used is a comprehensive questionnaire . The questionnaire was provided online via the tool Google Forms. This free Google service provides all the required functionality and has no limitations in number of respondents and number of questions. Results are automatically generated in an excel format. The link to the questionnaire is provided via email to all representatives.

The questionnaire consists of 39 questions, of which the majority has sub-questions and contains open questions, multiple choice questions and 7 point scale questions. The full questionnaire is presented in Appendix III on page 134.

A large drawback of the questionnaire research method is that the interpretation of the questions used can be very different between respondents. This can decrease the comparability of the answers given. Secondly, the response time to the questionnaire cannot be controlled by the researcher. The arrival of the results is completely dependent on the availability and willingness of the respondent. This also applies to the attention and quality of the answers given. Although these drawbacks partially remain, the extent to which they influenced the research project has been severely limited by the approach that all respondent were first interviewed in person for research step 1. This allowed the questionnaire to be partially designed to accommodate respondent communication styles and take into account the time available to each of the respondents to complete the questionnaire.

3. Current situation in the Dutch public executive sector

In the following chapter the current situation regarding Big Data in the Dutch public executive sector is described. The current situation is described as it will provide vital input for the next research steps of the project. Firstly, the current activities and interest in Big Data of organizations in the Dutch public executive sector will be presented, followed by the identified uncertainties that impede the assessment of the organizational readiness to start using Big Data in public organizations.

3.1 Activities and Interests

The first step of this empirical part of the research project is to establish an overview of the current situation regarding Big Data in the Dutch public executive sector. For this purpose the current interest in Big Data, current Big Data activities and future expectations regarding Big Data are discussed with the representatives in the interviews. From the responses it can be concluded that the present situation in the Dutch public executive sector is very comparable to the situation sketched in the introduction of the thesis report, which provided the motivation for this research project.

Firstly, the representatives of the participating organizations indicate that their organizations are indeed very interested in using Big Data and that they have high expectations of the value Big Data could bring their organizations in the future. However, the interest in using Big Data still seems to be not very specific at a number of organizations. Possible applications are not identified yet and exact forms of Big Data processes and its implications for the organization seem to be largely unknown. However, some of the organizations have much more specific interests in Big Data use and see a large spectrum of possibilities and benefits that Big Data could bring in the future.

The interest in using Big Data of the eleven organizations manifests itself in the fact that nearly all organizations are currently busy with some form of R&D activities around Big Data use in their organizations. The scale, scope and maturity of these R&D activities varies significantly across the eleven organizations, but almost all are developing their knowledge on Big Data in one way or another. These R&D activities are mostly concentrated on researching the possibilities of Big Data for the organization and developing future Big Data applications. Also, some ministries and other groups are cooperating in this development by organizing meetings and expert sessions on the topic of Big Data. However, most R&D efforts seem to be focused internally. The graph in Figure 8 depicts which sort of R&D activities the eleven organizations are currently undertaking regarding Big Data use. The categories in the graph are based on the scope of answers given in the unstructured interviews to questions regarding the current R&D activities in Big Data related topics. Organizations are classified based on the representatives description of the most dominant R&D activity in their organization.

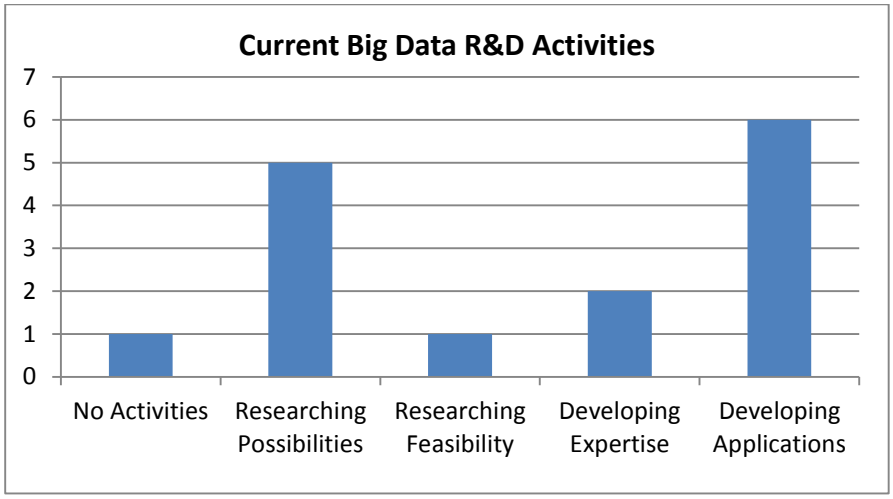


Figure 8: Current Big Data R&D Activities

The described R&D activities are performed for good reason, as ten of the eleven organizations expect their organizations to start using Big Data in their operations in the near future. However, as stated earlier, many of the representatives cannot give a full report on the exact applications and processes their organizations will be using Big Data in.

It is then no surprise that the other element of the current situation regarding Big Data derived from literature, that Big Data is currently not used in the public sector, is also recognized in the Dutch public executive sector. Only three of the eleven organizations are said to be currently using Big Data in their operational processes. However, when discussed in more detail, it became clear that the data usage stated to be Big Data does not fully qualify as such when a more strict definition of Big Data is applied. An overview of the current situation regarding Big Data in the Dutch public executive sector described above is presented in the graph in Figure 9. Each bar represents a question from the unstructured interviews and whether the representatives of the organizations gave a predominantly positive (yes) or negative (no) answer.

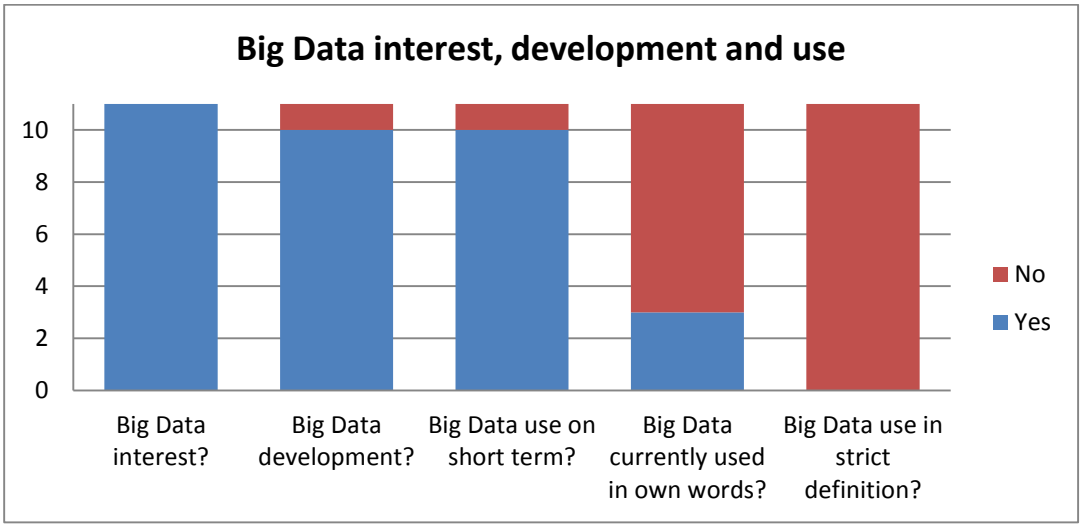


Figure 9: Big Data interest, development and use

The different interpretation of the term Big Data that lead to the difference between the fourth and fifth bar of the graph in Figure 9 is an indication of a bigger issue that became apparent during the unstructured interviews. Almost all representatives appear to use a different interpretation of the term Big Data. Some use interpretations that are specific to the area in which their organizations are active, others focus on different aspects of the definition of Big Data. A number of organizations see Big Data as a new data source, or form of data, others use the term to describe the new technology to handle data and some use it as a term to describe intensive data use. This very diverse use of the term Big Data is making communication about it much more difficult and even seems to hamper the cooperation in the development of knowledge and understanding about Big Data. It is therefore important to create a mutual understanding of Big Data and especially its implications for organizations that is usable and understandable for all (public) organizations. A proposal for such an explanation of Big Data is presented in chapter four on page 35.

So, although almost all representatives state that they are interested in Big Data, are developing it and expect that their organizations will be using Big Data in the near future, they all indicate that they currently have no fixed plan to get to actual Big Data use implementation in their organizations. Some of the representatives state that they are currently starting to develop these plans, but that they are not sure about the form and quality of these plans yet and that they are not sure if these plans will be put into action in the near future. All representatives indicate that they have many questions and uncertainties to answer before they can start using Big Data, and that these uncertainties are very hard to address, even within their current R&D activities for Big Data. An overview of the main uncertainties indicated by the representatives is presented in the next section of this chapter.

3.2 Uncertainties

When the representatives of the Dutch public executive organizations were asked about the main uncertainties that are impeding their progress towards Big Data use implementation, they almost all replied with a version of the same question:

Are we ready for Big Data use and how do we know if we are able to use it now?

Indeed the same main uncertainty as identified from literature and discussed in the introduction. As this main question is very broad, the representatives were asked if they could indicate which more detailed uncertainties were underlying their main question. From the answers three distinct underlying questions could be formulated that apply to almost all of the eleven participating organizations. The formulated uncertainties are presented in Figure 10. The three uncertainties are identified with the use of content analysis on the interview transcripts on questions relating experienced problems in assessing and deciding whether the organizations should start using Big Data or not.

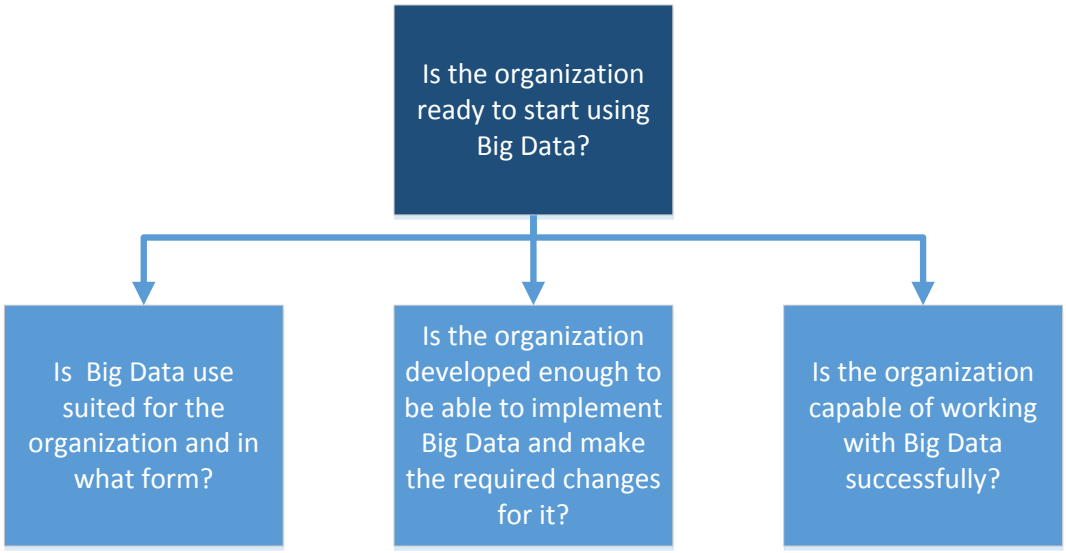


Figure 10: Main uncertainties for Big Data use implementation

A first uncertainty that is impeding the planning and undertaking of Big Data use implementation, according to the interviewed representatives, is that organizations question whether Big Data is suited for the organizations or not. And if Big Data is suited for their organization, in what form they could or should use Big Data to support their primary activities and strategy. Often a number of application ideas are being developed, but no structural information as to what will be applicable in the organization and what value that will add to it is currently available to the organizations.

Secondly, representatives are wondering if their organizations are ready to make the step towards Big Data use implementation or not. They wonder if their organizations are mature enough to successfully make the organizational changes that will probably be required for Big Data use implementation. They question whether they have developed sufficiently as an IT-supported organization to be able to take advantage of Big Data use technologies. And according to the representatives of the participating organizations, the organizations have difficulty in answering these questions.

Thirdly, the representatives question whether their organizations have the right capabilities and organizational conditions to use Big Data effectively and appropriately in the near future. They are not sure if the organization is currently capable of successfully using Big Data in the organization and think that further development of Big Data related knowledge and skills may be required before Big Data use should be implemented. The number of fields to address however varies wildly between the different participating organizations and representatives indicate their organizations have difficulty identifying the exact areas that need to be improved in their capabilities.

To summarize, the main reason why public sector organizations are currently having trouble to decide whether they should implement Big Data use or not, is that they are unsure if the organization is ready for it or not. The uncertainty on organizational readiness for Big Data use is recognized by virtually all Dutch public executive organizations participating in this research project. The representatives of these organizations furthermore indicate that the uncertainty on organizational readiness come from uncertainties regarding the following three points:

1. Uncertainty on the form of Big Data that is suited for the organization
2. Uncertainty on the maturity of the organization to change for Big Data implementation
3. Uncertainty on the capabilities of the organization to successfully use Big Data

The three uncertainties above are the main points that seem to be currently impeding Big Data implementation decisions in the Dutch public executive sector and will therefore be used in the remainder of the research project as the three focal points of the framework to assess organizational readiness for Big Data that is developed and presented in the following chapters.

4. Using Big Data in Public Organizations

As described in chapter three, the interpretation of the term Big Data used in practice differs greatly between the representatives that were interviewed in research step 1. As stated, the different interpretations are causing problems for clear communication about the development of Big Data and make it harder for public organizations to discuss Big Data and its consequences with other public organizations, national government bodies, employees, private organizations and civilians. And this of course has a negative effect on the successful development of Big Data use in public organizations.

A big part of the problem is that the current definitions of Big Data used in literature are not very sharp or specific. A brief description of the current widely used definition of Big Data in literature is given in the following section. Also, the following section will show why a clear and comprehensive definition of Big Data itself firstly is not entirely possible and secondly is not very useful for practitioners.

However, as stated earlier, a clear, understandable and communicable interpretation of the term Big Data is needed. Therefore an alternative way to interpret Big Data is presented in the remainder of this chapter. The proposed solution is here to describe and explain the use of Big Data in the organization, rather than to try to comprehensively explain what Big Data itself is exactly. The new explanation of Big Data use includes its differentiating characteristics from more conventional data use, a description of the activities of the Big Data use process and a typology for possible Big Data applications and its specific benefits for the public sector.

4.1 Current Big Data Definition in literature

The term Big Data was first used by two NASA aerodynamics researchers to describe datasets so large that specialized data analysis solutions had to be developed before they could be processed (Cox & Ellsworth, 1997). From this notion the first definitions for Big Data were derived, which had at least one element in common. Big Data was described as data so large that it did not fit and could not be handled by conventional database solutions. Big Data was data for which specific database solutions and data management strategies had to be designed and developed.

In 2001, data management expert Doug Laney introduced a first, more detailed framework to describe the problems that this Big Data (although he did not call it that) posed for data management because of the developments in digital data that needed to be handled (Laney, 2001). Laney describes three dimensions in which Big Data could challenge data management at organizations. The first dimension is volume, it describes the capacity of database systems that need to store ever larger quantities of data. The second dimension is velocity, it describes the need for very fast processing of data, as data is needed ever faster in organizations. The third and last dimension is variety and describes the fact that databases should be able to handle ever more diverse formats in which the data is introduced to the database.

Although this 3V model is still used today by many authors and practitioners to (partially) describe Big Data, the meaning of some of the terms used has been severally adapted since then. The 3V's are now used to describe the data itself. Now volume is the size of the datasets itself. Velocity is the speed at which this data is being refreshed or updated, in other words, the intervals at which new data is streaming into the data systems. Variety is now not only the different formats in which data

is provided and requested, but also how structured that data is and the variety of sources the data is coming from (Sagiroglu & Sinanc, 2013).

Although the 3V model provides some pointers as to what Big Data may look like, it is far from a complete definition. For example, it does not specify how large data should be before it can be classified as Big Data, nor does it do that in terms of velocity and especially not on variety. So, in short the 3V model is not very useful in practice if you want to draw the line between what is Big Data and what is not, it merely provides a glance at the end of the scale.

The suggestion of Cox & Ellsworth in 1997 that Big Data is data that cannot be handled by conventional data processing systems provides some indication as to where the line between Big Data and conventional data may be drawn. It is therefore commonly added to the 3V model to come to the following widely used definition:

> *Big data is a blanket term for any collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.* (Wikipedia, 2014)

Although this definition is understandable and neatly combines the notions of both Cox & Ellsworth and Laney, it still lacks specific information on what is Big Data and what is not. Of course, a line between the two is given. Data becomes Big Data when it can no longer be handled by conventional and available processing applications. The problem here however is that is it very hard to define what is conventional and what is not. It is a matter of perspective and a matter of timing. What is a conventional and available data processing application for a very large organization in 2010 may be very futuristic for a much smaller organization in 2020. If this definition of Big Data is applied in practice, what is Big Data and what is not will be different for every other organization out there.

Although it can be debated whether a clear line between Big Data and 'normal' data is specifically needed in practice, the currently used definition is so unspecific that it makes communication and understanding Big Data is practice much harder.

Apart from the blurriness of the line between Big Data and 'normal' data, there is another problem with the currently used definition. Because of the very fast (technological) advancements in IT the IT landscape changes very rapidly. And so will a line between conventional data and Big Data drawn on the basis of that landscape. What is experimental today, will be conventional tomorrow, old by the day after tomorrow and obsolete in a week. This makes that the line between Big Data and 'normal' data is not only blurry, but also very dynamic. The line between Big Data and conventional data shifts almost continuously with the introduction and adoption of new technology and is therefore very hard to define and perhaps even harder to interpret. Even further, it seems that the line can also be different for every other organization, or at least every industry, geographical area, and organization size and its technological capabilities. Because Big Data technology can be seen as a general purpose technology, trying to describe all of the applications and the data used by the technology is virtually impossible.

In short, the currently used definition of Big Data itself is not very usable in practice at all. And, because of the very specific and dynamic nature of Big Data it is unlikely a better definition will become available. And even if this is the case, it will remain uncertain how useful that definition will be to practitioners, as they have to deal will the very specific nature of the data that is available to their own organization and its specific corresponding complexities and dynamics. For these reasons, another approach should be chosen to describe Big Data and its consequences for organizations in practice. Furthermore, both practitioners and academia would benefit from an interpretation of the

term Big Data that is more specific, but also more widely applicable, to be able to better understand and communicate about Big Data.

In the following section the phenomenon of Big Data will be explained from the perspective of actually using all of that data by an organization.

By describing the use of Big Data instead of Big Data itself, the complexity of its dynamic and blurry nature can be overcome by describing the way these dynamics are handled, as the characteristics, activities and applications of using Big Data are much more similar between organizations than the data they are using itself. In this way all aspects of Big Data use can be described, which will make the phenomenon of Big Data more understandable for practitioners and academia, which will hopefully improve communication, cooperation and knowledge creation in the field.

4.2 Big Data Use Characteristics

The first step in describing the use of Big Data is clarifying the difference between conventional use of digital data and the use of Big Data. The five characteristics that are presented at the end of this section are formulated on the basis of insight from both literature and practice. In literature, several authors have mentioned one or more differentiating characteristics of Big Data use, compared to more conventional data use. A brief description of their insights is given below.

Mayer-Schönberger & Cukier (2013) see that Big Data is the result of the development in the world that data is no longer static, but speeding up, is becoming messier and that more and more data and data sources are becoming available. By creatively combining this data and using the new technologies as tools to process and analyze it, Big Data can be used to provide new insights and more value.

Adrian (2011) derives a number of characteristics that differentiate Big Data from conventional data when he describes the recent developments in data systems in firms. He uses the three V model to describe how the introduction of Big Data influences data systems. He explains that the high level of variety of Big Data means that organizations will receive data from a multitude of sources, in very differing formats and in both structured and unstructured forms. Furthermore he states that the increasing volume and velocity of Big Data demands investments in new and advanced hardware and analytical tools to collect, store and analyze all of that data. The increased velocity of Big Data also means that intervals between incoming data and changes in data are decreasing, leading to radically different methods of using and dealing with data.

Chen, Mao, & Liu (2014) describe elements of Big Data in their paper and, among other aspects, describe that Big Data use is different from conventional data use, as it requires very specialized hardware and software in the form of analytical tools and data management applications. They also mention that Big Data systems are able to analyze unstructured data as compared to more conventional data systems, which cannot do this. Finally, they indicate that Big Data use allows organizations to do real-time monitoring and receiving real-time feedback from a variety of new data sources.

Davenport, Barth, & Bean (2012) discuss how Big Data is different from more conventional data. They write that Big Data should be seen rather as a constant stream of external data, compared to the stock of (small) data that an organization stores internally. Combining these two is the power of Big Data use according to Davenport et al. They furthermore state that the main part of the stream

of Big Data is unstructured and that this stream of data is increasingly coming from real-time sensor networks.

The International Data Corporation sees Big Data as the product of the development of new and very extensive (networks of) data sources (IDC, 2011). They furthermore recognize that the increasing amount of digital data in the world is largely coming from a very steep increase in the amount of unstructured data and that the new Big Data technologies are allowing us to take advantage of this type of data.

The characteristics that describe Big Data use mentioned in the descriptions above are also present in the more than thirty presented definitions of Big Data on the OpenTracker website (OpenTracker, 2013). However, some of the presented definitions on the website also mention that Big Data use is about finding and using new data sources and innovatively using existing data sources for Big Data applications.

Simon describes a number of characteristics that make Big Data different from more conventional data in his book (Simon, 2013). He states that Big Data is different from regular, small data, because Big Data is much more fragmented and diverse and much more dynamic. He furthermore states that Big Data is unstructured and comes from outside of the organization and can only be used when it is combined with internal, structured, small data. He also recognizes that Big Data can only be used when new and specialized tools are used.

Next to the mentioned insights from literature, the representatives of the eleven participating organizations were asked in the interviews about what they thought Big Data constituted and what are differentiating characteristics of Big data use in comparison to the use of more conventional data. With the help of content analysis on the interview transcripts, the following differentiating factors could be identified and presented in the graph in Figure 11.

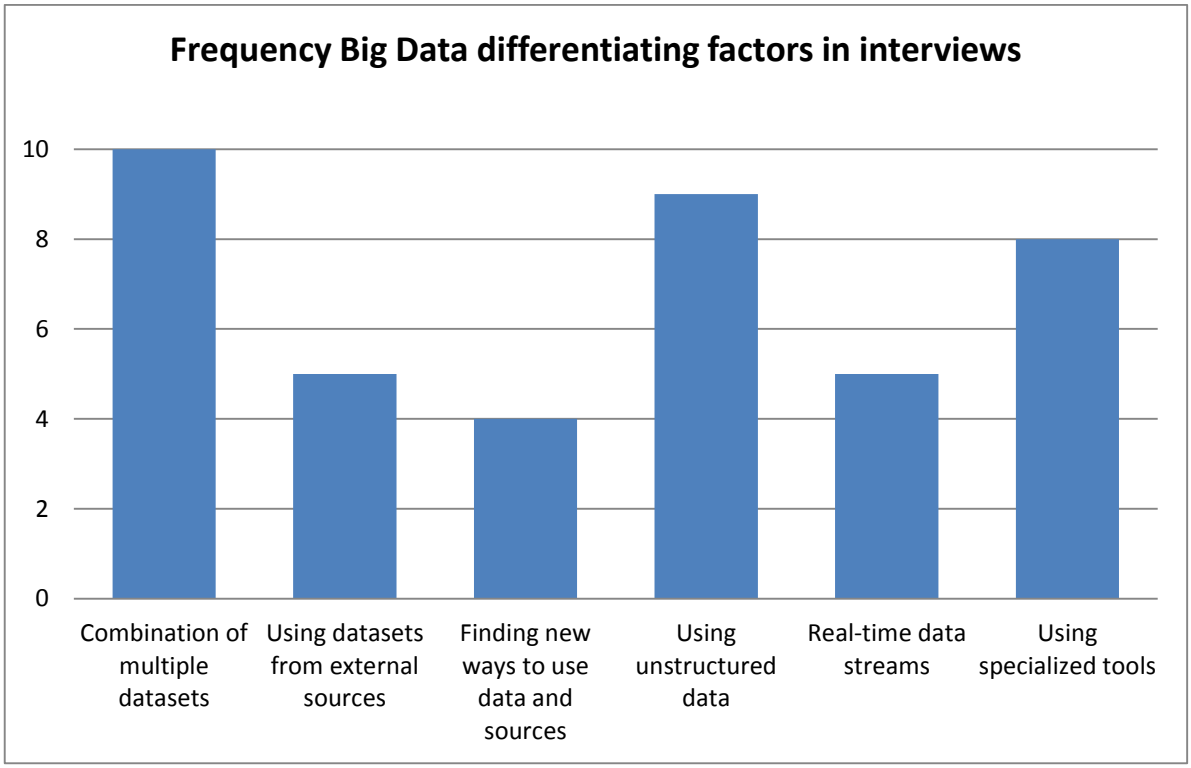


Figure 11: Frequency Big Data differentiating factors in interviews

After the organization of all differentiating factors of Big Data and the combination of the insights from literature and from the interviews five characteristics that differentiate Big Data use from the use of more conventional data could be formulated. The notion made by Phil Simon (Simon, 2013) that Big Data use will bring value especially if Big Data is combined with conventional data is specifically included in the characteristics. Through translation and interpretation of all the notions on differentiating factors between Big Data and conventional data described earlier, and by combining related notions into more comprehensive and complete factors, five characteristics are proposed in this research project that comprehensively and distinctively describe the differences between the use of conventional data and Big Data.

The following five differentiating characteristics describe the difference between Big Data use and conventional data use:

1. The use and combination of multiple large datasets, from both external and internal sources.
2. The use and combination of structured and unstructured data in analysis activities.
3. Real-time or near-real-time streams of incoming data which are structurally handled and analyzed.
4. The use of advanced analytics and algorithms, distributed computing and/or advanced technology to handle very large and complex computing tasks.
5. Innovative use of (existing) datasets and/or data sources for new and radically different applications.

In the following sections all five characteristics are briefly described.

Characteristic 1: Combination of Internal & External Datasets

Many organizations are currently using their own internal datasets to create information needed for their main activities and processes. Some organizations also use their internal datasets to evaluate their own performance. With the introduction of Big Data technologies, the insight from the internal datasets can be enhanced with the introduction of datasets from external data sources on a large scale. Big Data technologies enable organizations to find valuable information from the combination of multiple datasets much faster and much easier, which was much more difficult to do without Big Data technologies. In the past, only internal dataset could be easily used, as they could be generated in the desired formats for the organization. External datasets are more often than not, not available in the desired formats and transforming them was very difficult and time consuming to do. With the introduction of Big Data technologies as data and text mining algorithms, these barriers are now lifted and use of external datasets is now much more feasible.

From the combination of data from both sets of data sources, new and more reliable insights can be created from patterns and correlations (or the lack thereof) in the combined dataset. With the introduction of Linked and Open Data programs several years ago, large data sets of other government organizations have become freely available online for everyone. Although these datasets are usually not exceptionally rich and exhaustive, they can provide very valuable information when combined with the internal datasets of an organization.

Through data partnerships, where selected data is shared between multiple data generating organizations, targeted data acquisition programs and cooperation between private and public organizations, even more datasets can be combined with each other. Data acquisition initiatives by

both private organizations have even started to develop a new market for datasets in society, worth several billions US dollars per year (IDC, 2012).

Being able to combine and analyze both internal and external datasets easily is one of the differentiating characteristics of Big Data use.

Characteristic 2: Combination of Structured & Unstructured Datasets

Besides enabling easier and faster use and combination of internal and external datasets, Big Data technologies allow for the use and combination of both structured and unstructured data. Structured data is data that can be loaded into relational databases with fixed labels for both rows and columns. This kind of data can be read by machines, as the location of the requested information can be logically deduced. Another kind of data is semi-structured data. Although information contained in this kind of data is no longer loadable in relational databases, or other forms of data tables, the information contained within these data files is readable by machines through tags and metadata additions to the data. Lastly, there is also unstructured data. This type of data contains information that cannot be found in obvious, logical places within the files and can therefore not be read by machines without special applications. Examples of these types of data are images, video files and audio clips.

Due to evolutions in text mining, data mining and specialized algorithms associated with Big Data technologies, machines and computers are now capable of retrieving information form these types of files. And because of this, unstructured data can now be combined with structured data and analyzed to generate new information without the need for the semantic capabilities of humans.

The evolution of these kind of technologies enabling the use of both structured and unstructured types of data are another differentiating characteristic of Big Data use. It also enables data systems to make sense of machine logs and other data formats, which may enable Internet of Things developments is the near future.

Characteristic 3: Real-time data streams

As described before, the velocity of data is speeding up. What this means is that the intervals between the moment data is updated or coming into the data systems is getting smaller and smaller. This happens for several reasons. One, communication channels are getting faster and faster and tow, more and more data sources are connected to them. Furthermore, demands of data for being 100% up-to-date are getting higher, as data is being used more intensively. This results is demands for faster data collection and analysis, which in turn will results in smaller intervals and a higher data velocity.

When the intervals between data updates or changes are getting sufficiently small that they are becoming close to the speed of change in reality, the incoming data can be regarded as real-time data. The (near) real-time incoming data is sometimes referred to as streams or flows of data.

Important aspects of real-time streams of data are that all aspects of the data system of organizations should be able to process data at that same speed. This includes collection of data, in some cases by specialized sensor networks, the communication channels that provide

transportation of the data, storage and analysis of the incoming data and creating an overview of the data that allows either machines or humans to use the information that is created from the data.

Even further, automation of all of these processes will allow applications of data use that automatically collect, process, analyze and use the data and for example only provide information to users about changes in the streams of data, the patterns in the data or values in the data that surpass preset limits. These kind of signaling systems already exist, but only use conventional and a very limited range of data. Big Data technologies allows these kind of systems to use much more and much more varied data, creating much richer information from the streams of data.

Being able to effectively use a very large portion of the (near) real-time data that flows into the organization is a differentiating characteristic of Big Data use, as this cannot be done without Big Data technologies.

Characteristic 4: Advanced Analytics and Computing

In the previous descriptions of the differentiating characteristics, it was mentioned several times that Big Data technologies enable more possibilities for the input of data systems. Of course, the processing and analyzing of more datasets, less structured data and incoming streams of data require additional computational power and more advanced analytical software. So, the use of these technologies is a fourth differentiating characteristic of Big Data use in comparison with conventional data use.

Advanced computational power for Big Data use is found in specialized distributed computing cluster set-ups, or Cloud computing solutions, where analytics are offered As A Service to organizations with data use activities, without advanced internal IT facilities. Specialized and advanced analytical software applications with advanced mathematical and statistical functionality provide further support for Big Data analytics. These software applications are capable of more than descriptive analytics, as they also provide support for prescriptive and predictive analytical models.

Using both advanced analytics and advanced computing solutions is another differentiating characteristic between Big Data use and conventional data use.

Characteristic 5: Innovative Data Use

The final differentiating characteristic of Big Data use is that existing data sources and data sets are used for completely different ends than for what the data was originally collected and by developing innovative data collection methods, data sources and uses of data. Because of the much improved ease to combine datasets, experimentation with multiple datasets can lead to radically new uses of data. Data collected for a specific goal and used for it is no longer just valuable for that specific goal. With the aid of Big Data technologies these data sets can now be used for all sorts of other applications, further enhancing the development of new knowledge from data.

This way of out-of-the-box thinking regarding data has led to some remarkable applications, such as crowd-sourcing and traffic information from mobile phone network data. And this trend will only increase, as more sources and forms of data can be used and combined supported by Big Data technologies. The increase in Open Data being offered by government organizations provides further data that can be used for this creative and innovative process in search of new and valuable datasets.

4.3 Big Data Use Process

The second step in describing Big Data Use is the formulation of the main activities in the Big Data use process. By describing these activities, the unique consequences and aspects of Big Data use can be coupled to their specific data use activities and processes. Many of these data activities and processes are most likely already present within the public executive organizations, but probably not (completely) geared towards Big Data use. The activities of the Big Data use process are derived from literature, and especially from data value chain and Big Data specific literature and from the remarks of the representatives of the eleven participating in the unstructured interviews.

Big Data use process and activities in literature

From the additions on data value chains and knowledge creation processes from data of a large number of authors, technology consultants and vendors, four Big Data Use activities are formulated. The four activities are formulated with the help of a content analysis on activities discussed in the literature presented in Table 2. The multitude of found activities is reduced by finding higher level activities that entail a large number of more detailed activities described in literature. The four data use activities are: Collection, Combination, Analytics and Use. An overview of the authors and their contributions on Big Data process activities is presented in Table 2.

By formulating the four activities, Big Data use and all of its associated aspects can be made more specific and understandable for both the academic world and for practitioners. In this section a process containing the activities is proposed for the description of the connections between the main activities of Big Data use. The processes of how Big Data is and can be used is described in the several scientific articles, empirical research and Big Data system supplier information presented in Table 2.

	Agrawal	Brohman	Miller	Cumbley	Innosight	Oracle	TechAmerica	Tekiner	KVC	Bryant	Chen et al.	
	2011	2000/2006	2013	2013	2013	2012	2012	2013	2011	2008	2014	Total
Collection - category												**13**
Collection/annotate	0	0	1	1	0	0	0	0	0	1	1	4
Aquisition/Recording	1	0	0	0	0	1	1	0	1	0	1	5
generate	0	0	0	0	1	0	0	0	0	0	1	2
choosing data	0	0	0	0	0	0	0	1	0	0	0	1
sense	0	0	0	0	0	0	0	0	0	1	0	1
Combination - category												**14**
Extract/clean/prepare/process	1	0	1	0	0	0	1	0	1	0	1	5
Combination	0	0	0	1	0	0	0	0	0	0	0	1
Organize	0	0	1	0	0	1	0	0	0	0	0	2
store	0	0	0	0	1	0	1	0	0	1	1	4
Integration/Representation	1	0	1	0	0	0	0	0	0	0	0	2
Analytics - category												**15**
Analysis/Modeling	1	1	1	1	1	1	1	1	1	1	1	11
vizualize	0	0	1	0	0	0	0	0	0	0	0	1
communicate	0	0	0	0	0	0	0	0	1	0	0	1
interpretation	1	0	0	0	0	0	0	1	0	0	0	2
Use - category												**9**
Initiation	0	1	0	0	0	0	0	0	0	0	0	1
evaluation	0	1	0	0	0	0	0	0	0	0	0	1
deployment	0	1	0	0	0	0	0	0	0	0	0	1
make decision	0	0	1	0	0	1	0	0	0	0	0	2
Use	0	0	0	1	0	0	1	0	1	0	0	3
insights	0	0	0	0	1	0	0	0	0	0	0	1

Table 2: Big Data activities in literature

As mentioned, the activities mentioned in the articles and listed in Table 2, are also supported by the representatives of the eleven participating organizations in the interviews. The representatives indicated that they recognized the activities as important in the Big Data use process and furthermore indicated that these activities were often already part of the current data use processes in their organizations. In Figure 12, the frequency of the identified Big Data activities that are currently being performed in the participating organizations is presented.

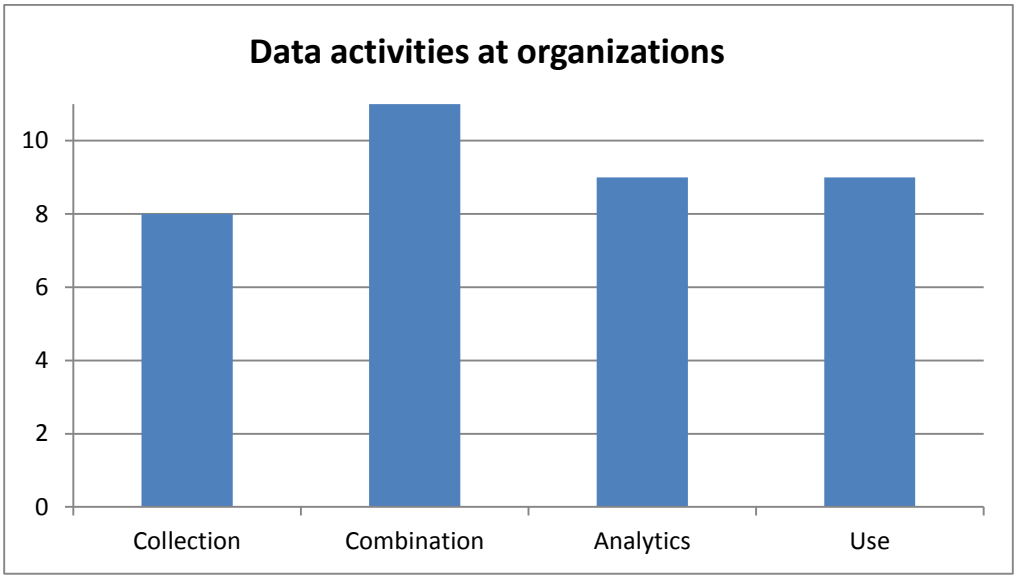


Figure 12: Data activities at organizations

Now that the four Big Data use process activities are identified, they can be connected to each other. Following the concept of the data value chain, the Big Data use activities are connected with each other based on the flow of data within the organization. Each step in the Big Data use process adds value to the data for the organization and if executed properly will go from the collection of raw data to actionable knowledge that is useful to the organization at the end of the process.

Furthermore, the last activity of the Big Data value chain, Big Data use, is connected to the first activity, Big Data collection, providing for a cyclical Big Data use process. The connection between Big Data use and Big Data collection is made because the Big Data use processes is a continuous process in which previous iterations influence following cycles. The Big Data use process is cyclical, as it is a knowledge discovery process. The process is driven by questions, a demand for information. And newly created information will lead to new questions and an adapted demand for new information. This iterative process is taken into account in the Big Data Value Chain through the feedback loop from Data Use to Data Collection and therefore the process is cyclical. The total Big Data value chain is presented graphically in Figure 13.

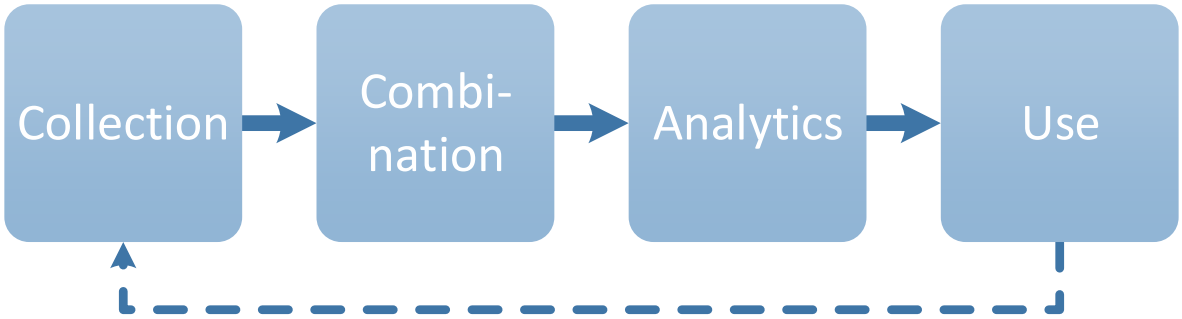


Figure 13: Big Data Value Chain

The Big Data value chain is proposed in this research project to give a clear and understandable overview of the process of using Big Data in an organization. The value chain is formulated based on notions from previously presented literature and added upon with interview from interviews with practitioners. Now that the Big Data value chain is presented, the individual Big Data activities are briefly discussed in the following sections. In the descriptions of the four Big Data activities the content and more detailed and lower level activities within the four higher level activities are discussed.

Data collection

The activity Data Collection involves of course the collection of the data used in the other data activities of the organization. Collecting data here is seen as the capture, both automatically and manually, of raw data from operational processes, from the environment, from handled objects and subjects and archived hard copy documents. In one form or another every organization is currently employing data collection in their day-to-day operations.

However, with the implementation of Big Data use in an organization, the collection of data not only becomes more important, it will change significantly. Once Big Data is used in the organization, the scope of the activity data collection will become substantially bigger. Now that the organization can make use of many different forms of data, the amount of available data and demanded data will grow substantially.

New potential sources of data with new data formats will create a need for a more elaborate execution of data collection, with more focus on the speed with which new data is being collected. Secondly, more focus for the question which types of data (structured, semi-structured and unstructured) will provide the best effort-information value ratio at the other three data activities will also become more important. Lastly, data collection will have to become a much more creative and innovative process. Big Data technologies enable organizations to look for new data sources from which they can retrieve data. These new data sources could provide additional and/or richer data adding to the data which is currently collected. But the new data sources could also be substitutes for the currently used data, and if they are easier, more secure and/or cheaper could make currently used data sources obsolete.

Data Combination

The second data activity that will be affected by the implementation of Big Data is the combination of data. Currently, data combination is a process in which newly collected datasets are combined with already stored datasets and/or other incoming datasets. These combinations, even without analyses, can create significant added value. The added value can come from richer images from data combinations, being able to interpret the new dataset with the values derived from older datasets and updating the old datasets with the new data. While data combination has many valuable outputs, it is only feasible when the old and new datasets fit together very well. They have to be quite similar, otherwise combination is very hard, or even impossible. So, thus far, data combination activities have always been quite limited and very dependent on the compatibility of new and old data.

With the introduction of Big Data technologies, this limitation has become much smaller. By using much more data, compatibility issues between the datasets will be much smaller. Furthermore, because Big Data technology allows the automated use of semi-structured and unstructured data, much more data will be available to combine with each other. Now text files, traditional databases and even real-time video files can be combined with each other to provide much more added value.

Data Analytics

Data analytics activities have been performed in organizations for a number of decades now. However, the use of more advanced, statistical and mathematical modelling techniques was, until very recently, not very widely spread. Mainly because the input of such modelling techniques had many strict requirements, the usefulness of such techniques in practice was very limited. Furthermore, interpretation of the results of such models was quite difficult, so the added value of such techniques in practice was very limited.

With the introduction of Big Data technologies, advanced analytical models can now support much more variety and diversity as input, which severely improves its usability. Lately, even unstructured data is becoming suited for input for these kind of analytical techniques. Secondly, as visualization techniques are also improving, interpretation of the results of such models is becoming much more accessible, further improving the added value of data analytics.

Now that computational power and distributed computing technologies have developed significantly over the last few years, advanced analytics can now even be applied to real-time data, to provide near real-time insights.

With the introduction of Big Data use in organizations, data analytics will become a very important part of the data value chain in organizations. Data analytics will possible be the activity which will added the largest amount of value to data going through the chain.

Data Use

The last data activity described in this chapter is Data Use. Although it seems logical that data is used at the end of the data value chain, it requires more attention than one may think. Actually using data, even if it is made substantially more understandable (and therefore more valuable) through combination and analytics of that data remains challenging for many organizations. For many decades, decisions and designs were based primarily on the experience and intuition of managers,

as very limited data was available. This culture has not changed significantly, although relevant data for decision making and designing is now much more available. So-called Data Driven Decision Making is needed for data to be able to create value for an organization. For effective use of data, not only the willingness of for using data is necessary, but also the expertise on how to interpret it, so that it can support the decision making process.

Making Data Driven Decision Making part of the decision process is important to let data help the organization. It becomes even more important when Big Data use is implemented in the organization. Now decision support information will also be coming from new and innovative sources of data and from outside of the organization. This requires both trust and clear guidelines on how to use and assess the (information from) data that is made available. Furthermore, when the information from data is coming available in real-time and decisions need to be made virtually instantly based upon that information, management of the actual use of that data becomes even more important. Big Data use implementation will affect the use of data in the organization and will require changes in it for organization to effectively use Big Data.

Connections between the Big Data use process activities.

To complete the description of the Big Data value chain and the Big Data use activities, the connections between the Big Data activities are described in this section. In these connections two widely used and more conventional data systems can be positioned compared to the use of Big Data. The positioning between Big Data and more conventional data systems is described here to show how Big Data technology relates to other systems with a similar knowledge creation function. Furthermore, the connection shows that Big Data technology is much larger in scope than other, more conventional systems and will therefore have an extensive impact on organizations and their data activities. Lastly, the conventional data systems described in this section are used in chapter 5.2 on page 63 in the organizational maturity assessment.

The connection between the Big Data activities Analytics and Use is described in literature as Business Intelligence (Chen, Chang, & Storey, 2012; Negash & Gray, 2003). Some people see Big Data as the next development and new name of Business Intelligence, but it is not (Simon, 2013). Business Intelligence is only designed to handle small and structured data, while Big Data systems are designed for much more, namely combined that small and structured data with big and unstructured data. However, experience with Business Intelligence systems can have very positive effects on the development of Big Data activities in organizations, as data analytics and data use activities are partially already being executed in those organizations.

The connection between the activities Big Data Collection and Big Data Combination within organizations is described in literature as Data Management (Ramakrishnan & Gehrke, 2003; DAMA International, 2009). For organizations that already employ Data Management systems, the same applies as for organizations that use Business Intelligence systems. Big Data activity development, especially on Big Data Collection and Big Data Combination can be positively influenced by experience gained on data use in the processes of the currently used Data management systems.

Both concepts have become conventional data systems in practice over the previous decade and are recognized and discussed in most interviews. Most representatives state that these systems are used in their organizations and that they therefore have developed substantial experience in some of the aspects of the Big Data use process. The use of Business Intelligence and Data management system

may therefore be a useful indicator for the assessment of Big Data Readiness of organizations and is thus incorporated in the Big Data Readiness Framework, which is presented in chapter six on page 55. Both conventional data systems are graphically represented in green in Figure 14 on page 48.

Now, two of the four connections between the Big Data Use process activities have been identified as known and existing concepts. The other two connections however are not clearly identified as such and require some explanation. As stated earlier in this chapter, the connection between Big Data Use and Big Data Collection is a connection that allows for iterative development of the entire Big Data use process. The information from the analyzed Big Data is used in the organization which will often lead to new, or at least adapted needs for information. These needs are then translated to requirements for which data to collect, so that the process can start at the beginning of the Big Data value chain again for another iterative cycle.

As the organization will adapt its needs for information to better be able to fulfill its organizational strategy, the connection between Big Data Use and Big Data collection is placed under the label of Strategic Data activities. The two strategic data activities, Big Data Use and Big Data Collection represent the strategic choices an organization can make in its data activities. By choosing how to use (Big) Data in the organization, the information need for the data activities is determined. From this the requirements for the collection of data can be formulated, which in turn will determine for a large part the further steps of the Big Data use process.

In this regard, the other two Big Data use process activities, Big Data Combination and Big Data Analytics are the supporting data activities for the choices made in the two strategic data activities. Once the requirements for the data to be collected are fulfilled, the combination of that newly collected data with existing data and the sequential analysis of the combination of that data will support the choices made in the earlier activities. Big Data Combination and Big Data Analytics and their connection are therefore labeled as Data Resources. Both new connections Strategic Data Activities and Data Resources are presented in purple in Figure 14. Choices made for the two strategic data activities, Data Collection and Data Use determine in a large part the strategy that organizations will follow in their data use processes. The two data activities are therefore used in the remainder of this research project as an indicator of current data strategies at the participating organizations. How the two data activities are used in the research project is explained in chapter five on page 56**Fout! Bladwijzer niet gedefinieerd.**.

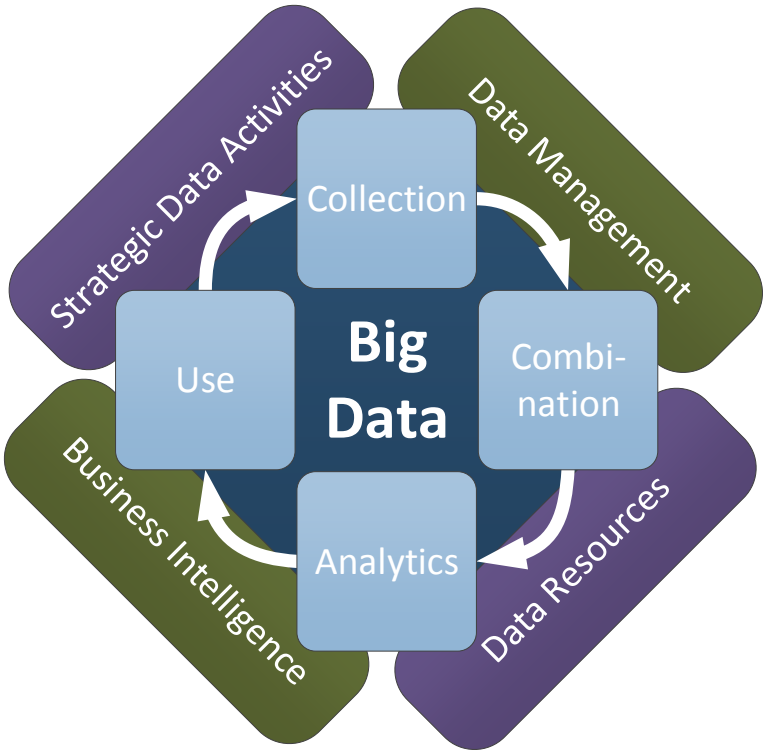


Figure 14: Big Data Value Cycle

4.4 Big Data Use Applications

Now that the Big Data use process and the differentiating characteristics of Big Data use have been explained, the next step is to describe the different applications of Big Data use. As there are as many examples of Big Data applications possible as there are ideas for them, a typology of Big Data applications is needed to give a clear description. A number of authors have tried to create typologies for Big Data applications, however, their usefulness for practitioners is doubtful at best. Chen, Mao, & Liu (2014) propose a Big Data application typology based on the various data formats that are used for the applications. As the biggest value of Big Data comes from combining different data formats, this typology is not particularly suited to explain Big Data applications to practitioners. Secondly, as data formats vary greatly across organizations and industries, applicability of such a typology will be severely limited. Another typology of Big Data applications is based on the type of analytics used to analyze the data itself (Gartner, 2012). Although this may be convenient for analytics gurus, it remains a rather limited typology for practitioners that are much more interested in the practical effects of the applications on their organizations and main activities. Other typologies are (Shiri, 2014; Milakovich, 2012) more industry or sector based, but lack sufficient detail to provide additional insight to practitioners.

So, a new typology for Big Data applications is needed for this research project to be able to effectively describe the differences and similarities of different Big Data applications and connect them to the proposed explanation of Big Data use previously presented in this thesis report. The Big Data application typology presented in this section is made on the basis of the questions practitioners, the representatives of the eleven participating organizations, had on how Big Data applications for their organizations would look like. They questioned what would be the starting

points in the data use process of Big Data applications, which aspects of Big Data would play the most important role in possible Big Data applications and especially which value applications of Big Data would add to the organizations. By answered these questions for all of the Big Data applications that were discussed during the interviews, an overview of the range of Big Data applications and their relevant characteristics was formulated. By using concept analysis on the discussed planned Big Data applications in the interviews with public officials, an overview of Big Data applications and their attributes was formulated for this research project

From the overview a typology for Big Data applications could be determined. The proposed typology is primarily based on the differing added value the different types of Big Data applications would preferably produce and what the starting points of those types of applications would be in the Big Data use process. In other words, the three types of Big Data applications differ from each other in the type of knowledge they create for their users. And this is not strange, as the main function of Big Data applications will always be knowledge creation. By using different inputs of information and different analyses methods, different types of knowledge are created, for different purposes, adding value to their organizations in different ways.

After further examining the examples of Big Data applications for each type of applications, the three most prominent differentiating characteristics were identified for each of the application types. With that information, the Big Data application types presented in Table 3 could be formulated.

Detailed descriptions and an example per Big Data application type are presented in the following sections.

Application type	**Object/Subject evaluation**	**Research**	**Continues Monitoring**
Prominent Big Data Use Characteristics	- Internal & External Datasets - Innovative use of existing data - Advanced Analytics & Algorithms	- Internal & External Datasets - Structured & Unstructured data - Advanced Analytics & Algorithms	- Real-time or near real-time - Advanced Analytics & Algorithms - Innovative use of existing data
Initial Big Data Use activity	Combination	Analytics	Collection
Added value	Decision support information	New insights	Richer Image of Reality

Table 3: Big Data Application Types

Object/Subject Evaluation Applications

The first Big Data application type to be described is the Object or Subject Evaluation application. This type of application is used to evaluate, rank or classify large sets of objects or subjects. The evaluation is based on a large number of different attributes of each of the objects for which data is available. By analyzing all the values of each of the attributes of all of the objects, the objects can be evaluated individually, or ranked or classified in comparison with each other based on pre-set decision rules. The decision rules can be based on anything from legislation, to policies rules, to outcomes of predictive models or even expert opinions. With this application, very specific cases can be identified literally out of millions of objects or subjects. It could be that objects with very special attributes are sought after as they will receive a special status for future reference, or that the top 1% in a certain combination of attributes has to be known, or that a set of objects needs to be classified in specific categories as they will receive tailored services for their specific category, or any other advanced method of selection, evaluation, ranking or classification.

Although this application of Big Data might not seem to be extremely advanced at first glance, as it could all be done by hand or with less sophisticated technologies, it becomes a very different story when the sets of objects or subjects number in the millions and each record has over 50 different attributes. Add to this that the data is combined from multiple datasets from different sources, with different formats, different intervals of updates or refreshes, different labels for objects or subjects and hundreds of different decision rules including mathematical combinations of object attributes and Big Data technologies are the only tool that can be chosen.

Big Data Use characteristics

Three of the previously described Big Data Use characteristics are important parts of Object/Subject evaluation Big Data applications. The first characteristics is the combination of both internal and external datasets for the evaluation. The evaluation of objects or subjects will become more reliable, more precise and richer if more attributes can be used in the evaluation by the preset decision rules. By combining the attributes from multiple datasets, both from within the own organization and from other organizations and/or data sources the quality of the outcome of the object/subject evaluation application can be further improved.

The second Big Data characteristic important for object/subject evaluation applications is the innovative use of existing data for the evaluation. By re-using data for new and innovative purposes, the information on objects or subjects may be enhanced even further by additional attributes that were not available previously, resulting in even better evaluations based on even more specific decision rules.

The last Big Data characteristic important for object/subject evaluation applications is the use of advanced analytics and algorithms. When datasets become very large and decision rules become very complex, more advanced algorithms will be needed to be able to automatically and correctly apply all decision rules on the data for the applications.

Initial activity/starting point

The most important Big Data activity for the object/subject evaluation applications in the Big Data use process is the combination activity. The vital starting point for these kind of Big Data applications

is the combination of multiple datasets to allow these applications to build rich images of attributes for each object or subject to base their evaluations on.

Added value

The main added value the object/subject evaluation applications offer to organizations is decision support information. The evaluations, rankings and/or classifications of objects or subjects provide detailed input information for decision making processes within organizations. This can be in both policy and management decision making, or in every day processes, dependent on the goal of the application.

Example

An example of a possible object/subject evaluation Big Data application is the classification of all cars in a fleet of rental cars in maintenance intervals. By classifying each car in a category with a certain service level of maintenance, the cost of maintenance and the risk of breakdown can both be reduced. Each car is put into a category by using decision rules that evaluate attributes of the cars, such as age, mileage, size, brand, number of trips, rental history. The classification can, for example, be further improved by adding data from the car maintenance companies on maintenance history, used spare parts, maintenance cost per defect, etc. Furthermore, the classification may be improved by adding a dataset of the national weather service on road conditions on which the cars were used. Adding and combining more datasets will require more advanced algorithms and analytics as decision rules will become more complex, but will improve evaluation quality and therefore reduce cost and risk of breakdowns.

Research Applications

The second Big Data application type described is the Research Big Data application type . In this type of application Big Data is used to find new information and create new insights. By for the first time combining a very large amount of data from various data sources and with very different formats and structures, and using a very large number of different and advanced analytical methods, new relations and connections can be found, patterns can be identified and never before observed behavior can be recognized. The new connections and patterns in the data can be used as clues to further investigate those specific combinations of data. From these radically new correlations can be detected that were unimaginable and untraceable without Big Data technologies, leading to new insights into areas that have yet to be explored. Whether these areas are unknown areas of physics for scientists, new customer segments for private firms, or unknown effects of present policies for public organizations, the new insights will always add value.

Big Data Use characteristics

Again, three Big Data use characteristics are important parts of this Big Data application type. First of all, the combination of multiple datasets, from both internal and external data sources are important for research Big Data applications. By combining more datasets and more data, more connections and patterns are likely to be discovered, which can be investigated in more detail.

The second Big Data use characteristic important for research Big Data application types is the combination and use of both structured and unstructured data. Structured data is the type of data that can be read by computers, can be found in relational databases and has been used for research for decades. Unstructured data however can only recently be used in analysis by computers and is therefore a largely untapped, albeit very difficult to handle, source of new connections and insights for research.

Thirdly, advanced analytical methods and algorithms are needed in research Big Data applications to be able to find new connections and patterns and to handle and combine structured and unstructured data.

Initial activity/starting point

The starting point in the Big Data use process of research Big Data applications is at the analytics activity. Without making sure the proper analytical tools and methods are available and the right algorithms are present to handle unstructured data, research Big Data applications will not be able to function.

Important to note is that solid analytical skills and knowledge is vital for organizations, as this very complex discipline can lead non-experts to conclusions and interpretations from the data that do not necessarily are correct. Not only the difference correlation and causality is important here, but also the selection of correct and appropriate analytical methods, as they are not all suited for all types of input.

Added value

The added value of research Big Data applications is that it can offer new insights into the world around us. And this can even be done by analyzing already collected and existing data. This makes research both easier and cheaper and the new insights can provide significant value for all types of organizations in the form of opportunity of threat identification, cost reduction, quality improvements and so on.

Example

An example of a research Big Data application is research into cancer treatment. Previously these kind of research projects would include carefully choosing 200 qualifying people, collection all of the required information from them one by one, following them for ten years and finally observing if treatment was successful or not and drawing conclusions which attributes of the patients and the treatments correlate with successful and with unsuccessful treatment, if any can be found. And then the process would start again with a slightly different scope to verify results and find new connections. A slow and very intensive research project, with in the end a limited applicability, as treatments, patients and other circumstances vary greatly. With Big Data technologies a very different approach would be possible. With Big Data all (anonymized) digital patient records of cancer patients of a certain period in a certain area or nation could be combined into a very large dataset, along with hospital records, health insurance data and other similar datasets. With the latest Big Data technologies, even unstructured data, such as x-ray photos and MRI scans could be

included. From these combined datasets, advanced analytical tools can then find new connections, correlations and patterns that would never have been visible in conventional research. These new insights could be very valuable as they could advance the field and help future cancer patients.

Continues Monitoring Applications

The last Big Data application type described is the Continues Monitoring Big Data application type. This type of Big Data application makes use of data collection through extensive sensor networks that send data real-time or near real-time to a central point for it to be analyzed as quickly as possible. By providing the resulting information on for example dashboards in quickly understandable visualizations, human decision makers have the necessary information at hand to quickly react to certain situations is reality and take the required action. Of course, automated reactions and subsequent action by computers is also possible. Continuous monitoring Big Data applications have a signaling function that identifies if preset conditions are taking place in reality or not and notifies other systems, or humans that corresponding action is required.

When the amount of incoming data, its variety and its velocity are becoming very large Big Data technologies are needed to support these application. Especially when also (near) real-time analytics are required to analyze the incoming data to make sense of it for other systems or humans, these type of Big Data applications become indispensable.

Finally, this type of Big Data application also allows for experimentation. Because objects are monitored in real-time, their behavior or reaction to new signals can be recorded. By sequentially introducing various signals to the monitored objects, the signals with the behavior closest to the desired result can be identified.

Big Data Use characteristics

There are three important Big Data use characteristics for continuous monitoring applications. Firstly, these type of Big Data applications need to be able to handle data that is coming into the system in (near) real-time. This means data is streaming into the system at intervals shorter than one second, or even continues, and with minimum lag. Important to note is that not only should the hardware be able to transport and store these streams of incoming data, the system should also be able to handle, combine, analyze and visualize all of the incoming data in (near) real-time as well. Especially when advanced analytical methods are used, this will require an enormous amount of computational power.

The second big data use characteristic important for continuous monitoring Big Data applications is advanced analytics & algorithms. Although it is very difficult and possible not needed to use very advanced and complex analytical methods for analysis in this type of applications, the analytics and algorithms do need to be able to work very fast and very efficiently, which does mean they need to be very advanced in that respect.

Finally, the third important big data use characteristic for continuous monitoring applications is the innovative use of existing data. Often exact measurement of a certain object, subject or circumstance in reality is very difficult and very expensive to do. And this is enhanced if measurements need to be done in (near) real-time. These challenges can be overcome by finding proxies for the object that needs to be measured, for example in currently existing datasets that are

already collected for a different purpose, or new data that can be collected much easier and cheaper. Through analytics the proxies can be adapted towards the desired information.

Initial activity/starting point

The most important Big Data activity for continuous monitoring applications is data collection. This is the activity where these applications are initialized, as the available (near) real-time data are vital for the nature of these applications.

Added value

The added value of continuous monitoring Big Data applications is that these applications provide decision makers, policy designers and/or automated reactionary systems with a richer image of reality. As Big Data systems can process more data more quickly a multitude of the data processed by conventional systems can be used to create a richer and more precise image of objects, subjects and circumstances in reality, which will allow for better and faster decision making based on that image.

Example

An example of a successful application of Big Data continuous monitoring technology can be found in neonatal care at the Children's hospital in Toronto (Proffitt, 2012). Premature born babies receive intensive care right after birth, but a large percentage of them still get a life-threatening infection in the first few weeks. Doctors and nurses monitor the babies vital signs every hour, but are sometimes too late in detecting the starting infection. The equipment used to monitor the babies is actually doing this very precisely and continuously, but until now this data was not used. By digitally collection and combining the data on all vital signs of the babies, the big data application can now detect starting infections in babies over 24 hours before the regular manual checks of doctors and nurses would have detected the infection. This makes the infections more treatable and less dangerous for the babies health.

4.5 Value of Big Data Use Applications for the public sector

The three different Big Data application types and their main added values are specified in light of the three main goals for public sector organizations of e-government development, as was established in chapter one on page 14. The three goals are transparency, effectiveness and efficiency (Weerakkody, Janssen, & Dwivedi, 2011; Joseph & Johnson, 2013). This is done to connect the three Big Data application types and their specific added value to the contributions Big Data is claimed to have to the development of e-government and t-government statuses of public organizations. By using the three main goals of e-government and combining them with the specific added values of the three Big Data application types, the overview in Table 4 can be formulated. This overview is proposed in the research project to further specify the added value of the three Big Data application types and to make them more insightful for practitioners. The more detailed overview of the added value of the three Big Data application types show the potential impact and benefits of the application types on public organizations better.

The added value described in Table 4 is based on notions from both the interviews, and from literature. Some of the representatives gave very specific answers when asked for desired and planned added value of future Big Data applications in their organizations, which proved to be very helpful in formulating the added values in Table 4. Notions from literature also provided some very useful insights. The US Federal Big Data Commission and Milakovich identify substantial benefits of Big Data use in multiple fields for the US government to improve its efficiency, effectiveness and transparency, such as better decision support information, better insights into reality and more transparency in decision making processes (Federal Big Data Commission, 2012; Milakovich, 2012). Specific benefits of Big Data for the public sector are: more effective policies, faster signals of the effects of policies, so faster adaption of them and working more efficiently, as more information for their operations is made insightful and available (Toorn, 2014).

Application type	Added Value	Transparency	Effectiveness	Efficiency
Object evaluation	Decision support informa-tion	A. More transparent public sector decision making as information used in DM process is more fact-based	D. Better decision support information will create higher quality and therefore more effective decisions	E. Better decision support information enables faster and even automated decision making, increasing efficiency in operations
Research	New insights	B. New insight in true effects of current and past policies and processes can be demonstrated clearer and more fact based	F. Policy and process evaluation is richer and better with new insights, leading to better understanding for future policy and process (re)designs, making them more effective	H. New insights will provide additional information which will speed up policy and process evaluation and (re)design, increasing efficiency.
Continuous monitoring	Richer image of reality	C. Current situations in reality can be made more transparent by monitoring more variables with smaller intervals.	G. Continuous monitoring allows for real time experimenting, which, through testing, will improve policy effectiveness	I. Continues monitoring leads to higher efficiency, as non-performing policies and bad decisions are identified sooner.

Table 4: Added value of Big Data applications

The specific added value per Big Data application type are used in the Big Data Readiness assessment to establish the interest in and need for each of the three types of Big Data applications for each of the assessed organizations in the organizational alignment assessment described in chapter 5.1 on page 56. This is done, as practitioners had very specific, but not very consistent desires and expectations of benefits of Big Data applications for their organizations.

5. Big Data Readiness Framework

In this chapter the framework needed to assess the readiness for Big Data use of public organizations is developed and presented. As stated in chapter two, about the research approach, the framework is based on the three main uncertainties on Big Data readiness assessment identified in the interviews with representatives of organizations from the Dutch public executive sector. The three uncertainties identified are uncertainty on the best suited form of Big Data use, uncertainty on the maturity of the organization needed to implement Big Data and uncertainty on the capabilities of the organization to successfully use Big Data. The three uncertainties are translated into three separate parts of the Big Data Readiness Framework which will together make up the framework. The three separate parts are organizational alignment, organizational maturity and organizational capabilities. A graphical representation of the framework is given in Figure 15 below.

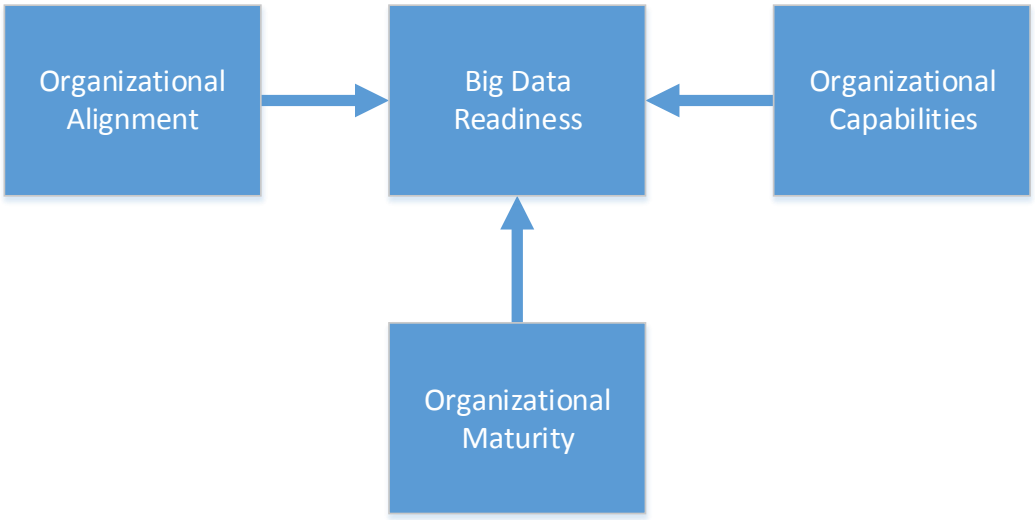


Figure 15: Big Data Readiness Framework

The three parts of the framework are each explained in detail in the following section in this chapter. For each of the parts of the framework established scientific frameworks on the interaction between organization and technology is used as a basis to create an assessment method for the readiness of public organizations to start using Big Data. However, none of the used frameworks was designed with Big Data in mind, so each of the frameworks had to be adapted towards it. The adaptions are explained in further detail at each of the three parts in this chapter.

The overall assessment of Big Data Readiness of an organization is the sum of the scores of the three parts of the framework and will be expressed in both a Big Data Readiness Score and a Big Data Readiness Assessment. A detailed description of the assessment decision rules and scorecards can be found in Appendix IV on page 140.

5.1 Organizational alignment

The first uncertainty that is addressed as a starting point for the new framework is:

Is Big Data use suited for the organization and in what form?

The elements of this uncertainty can best be caught in the term organizational alignment. The uncertainty revolves around questions if Big Data use can be aligned with the current organization and its main activities and strategy and if so, in what form Big Data use is best aligned with the organization.

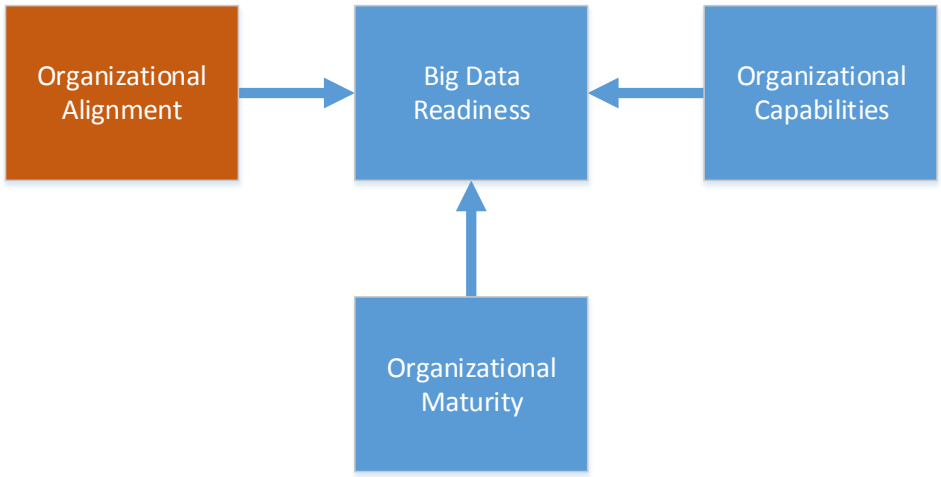


Figure 16: Organizational Alignment

A survey from NewVantage Partners shows that this issue is at the forefront of Big Data implementation issues at a very large number of organizations and that it is vital for successful use of Big Data that its alignment is properly managed (Kiron, 2013). Some Big Data specific literature provides some clues on this topic. Malik (2013) couples assessment of Big Data readiness with IT-business alignment assessment. However, currently no Big Data alignment theory is formulated. So, the next step is to look towards other fields of IT research and see if theory on alignment from those fields could be relevant to Big Data alignment issues.

One existing framework from IT implementation research was found to be relevant and usable in the context of the alignment of Big Data use with the organizational activities and strategy. The framework used as a basis for the Big Data organizational alignment assessment is the strategic alignment model, formulated by Henderson & Venkatraman (1993). The main notion taken from the model is that business strategy, IT strategy, organizational infrastructure and IT infrastructure should be aligned with each other, as they are interconnected within the organization and have to support each other.

However, these factors of the model are formulated well before the term Big Data was ever introduced, and they are therefore not surprisingly not formulated with Big Data applications or similar concepts in mind. To make the strategic alignment model, which is shown in Figure 17 on page 58, more geared towards Big Data applications in the public sector, the four aspects of the model are adapted to more Big Data and public sector specific and relevant aspects.

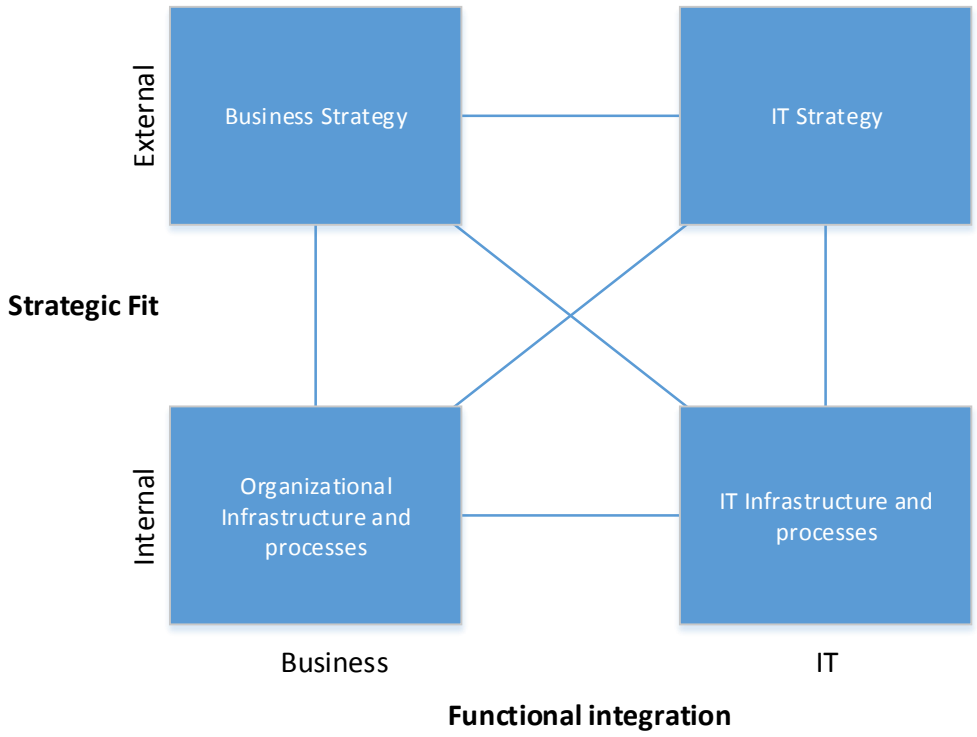


Figure 17: Strategic Alignment Model

The adaptions to make the model more Big Data and public sector specific, are based on the Big Data use concepts proposed in chapter four on page 35. They are furthermore based on insights from discussions on alignment of Big Data use in the unstructured interviews

The first factor of the strategic alignment model that is adapted to make it more Big Data and public sector specific is Business Strategy. The factor Business strategy can be adapted quite easily towards the main statutory tasks of public organizations, as these tasks largely determine what activities are performed and what choices are made by the public organizations. The statutory tasks of a public executive organizations are tasks set by law that these organizations are obliged to perform. In return for performing these tasks, the public organizations have the right to receive, in various formats, funds from the public to finance themselves. Furthermore, public organizations are often also limited to only perform their statutory tasks and activities that support it and not more. As these statutory tasks are the main focus of all activities and major decisions in public executive organizations, they can be seen as the organizations main strategy to continue its current existence. So, adaption of the first factor of the strategic alignment model was straight forward, however, the second factor, organizational infrastructure, is more difficult to adapt.

Discussions on the uncertainties around the alignment of Big Data application types and the organization in the interviews gave indications of a connection between the statutory tasks or main activities of the participating organizations and the data activities that the organization currently employs. This connection can be explained by identifying that the main statutory tasks of the organizations largely determine the main activities of the organizations and that these activities are supported by the data activities of the organizations. The connection between the statutory tasks of

the organizations and their data activities is therefore not surprising. Performance of certain tasks will lead to a need to perform certain data activities to support that task, creating a distinct set of data activities per type of statutory task the public organizations has to perform.

Earlier in chapter four (page 46), the Big Data process activities were described and two of the four activities, Big Data Use and Big Data Collection were labeled as the strategic data activities of the organization. By choosing how and especially how intensive these activities are structurally performed, the entire data use process has to be aimed towards the choices made for these two data activities. Therefore, the organizational infrastructure factor of the strategic alignment model can be adapted towards the intensity with which the strategic data activities Data Collection and Data Use are being performed in the organization.

Now the next step is how to determine how these two adapted factors of the strategic alignment model align with each other. To determine this alignment the main statutory tasks of the eleven participating organizations were compared with each other and from this comparison four types of statutory tasks were formulated as a typology for the organizational strategies. This was done with the help of a concept analysis on an overview of the main statutory tasks created from the interview transcripts. By comparing attributes of the different organizations and the described main statutory tasks, four main types could be formulated. These four types of statutory tasks are presented in Table 5.

Organization type	**1**	**2**	**3**	**4**
Main Statutory Task	Coordination, Project based	Research, Evaluation	Registration, Documentation	Administration, Management

Table 5: Main statutory task types

Now that a typology of the organizational strategies is formulated, a connection between the statutory task types and strategic data activities is investigated. After an overview of the intensity of the strategic data activities and the statutory task types per organization is made, a clear connection can be identified. The four types of statutory tasks all have a different configuration of intensity of the two strategic data activities. The connection between the two factors of the strategic alignment model adapted first is presented in Figure 18 and in Table 6.

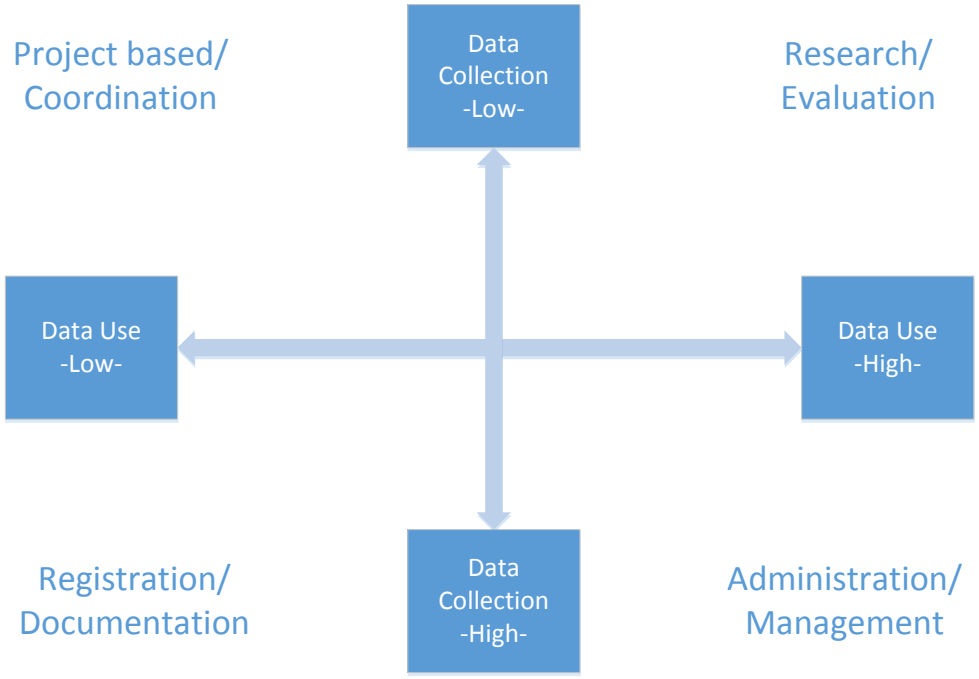


Figure 18: Statutory tasks and data activity intensity

Organization type	1	2	3	4
Main Statutory Task	Coordination, project-based task, no structural data use	Research, Evaluation	Registration, Documentation	Administration, Management
Data Collection Activity Intensity	Low	Low	High	High
Data Use Activity Intensity	Low	High	Low	High

Table 6: Statutory task types and data activities

Now that the first two factors of the strategic alignment model are connected with each other, the adaption of the third factor is discussed. The IT strategy factor is translated to the type of Big Data application that the organizations that will be assessed are most interested in. The type of Big Data application chosen will in a large part determine the IT strategy used in the organization to support the organizations Big Data activities.

The fourth and final factor of the strategic alignment model, IT infrastructure, is adapted towards the five differentiating characteristics of Big Data use described earlier in chapter four, on page 37. The differentiating characteristics of Big Data use are already connected to the three different types of Big Data applications in chapter four and presented in Table 3 on page 49. By assessing to what

extent the five Big Data use characteristics are currently already part of the data activities of the organization, the alignment between the current IT infrastructure and required IT infrastructure for the planned type of Big Data application can be determined, which is an important part of the alignment between the current situation in the organization and the planned future situation including Big Data use.

Optimal Alignment

From comparing Big Data characteristics, Big Data application plans, current data use and activities and current strategies and statutory tasks derived from the representatives of the eleven organizations in the interviews in a concept analysis, Table 7 with the optimal organizational alignment of Big Data application types can be formulated. The current alignment of the planned Big Data applications at the organizations, that will be assessed, with the characteristics of current data use, and the main statutory task will be compared to the optimal alignment in Table 7.

Organization type	**1**	**2**	**3**	**4**
Main Statutory Task	Coordination, project-based task, no data use	Research, Evaluation	Registration, Documentation	Administration, Management
Data Collection Activity Intensity	Low	Low	High	High
Data Use Activity Intensity	Low	High	Low	High
Most present Big Data Use Characteristics	-	- Internal & External Datasets - Structured & Unstructured data - Advanced Analytics & Algorithms	- Internal & External Datasets - Innovative use of existing data - Advanced Analytics & Algorithms	- Real-time or near real-time - Advanced Analytics & Algorithms - Innovative use of existing data
Best Aligned Big Data Application type	-	Research Applications	Object/Subject Evaluation Applications	Continues Monitoring Applications

Table 7: Optimal Organizational Alignment for Big Data

Alignment assessment

The first part of the Big Data readiness assessment can be performed by evaluating if the main statutory task (its organizational strategy) and strategic data activities (its organizational infrastructure) are in alignment with the planned Big Data application type (IT strategy) and supporting Big Data Use characteristics (IT infrastructure). The assessment of the organizational aspects is done based on the information from both the interviews and in the questionnaire. The assessment of the IT aspects is done by questioning the three Big Data application types on four dimensions: interest, need for specific benefits, applicability and feasibility. The exact assessment decision rules of the organizational alignment assessment can be found in Appendix IV on page 140.

Validity of Organizational Alignment

In this section some remarks will briefly be made on the validity of the previously described assessment method for the organizational alignment. Firstly, it must me noted that the assessment method proposed below is primarily suited to public executive organizations with clear and intensive data activities. Public organizations with a much lower intensity of working with data in their day to day operations may not be able to express their main statutory tasks in data activities and it may therefore not be possible to assess these organizations with the proposed assessment method. However, if intensive data activities are currently not part of day to day operations in public organizations it remains questionable if Big Data use is the right choice for these organizations. In the current assessment method these organizations are assessed with a lower organizational alignment score and that may be quite accurate.

The second remark on the validity of the assessment method is that the organizational alignment is partially based on the comparison between the characteristics of the current data use and the planned use of Big Data in future Big Data applications. Although an organization will be more ready for Big Data if current data use is more similar to future Big Data use (which the assessment method currently supports) it somewhat neglects the possibility that organizations make a larger and more discrete step in their data use when they implement Big Data use, instead of a more incremental development. This omission in the assessment method results in a lower assessment of Big Data Readiness on organizational alignment for public organizations that employ a larger and more discrete step between their current data use and their future Big Data use, even when the planned Big Data applications are fully suited with the organizations main data activities and statutory task. Now, during the interviews it became apparent that almost all organizations are employing the more incremental approach, but this is not necessarily the case for other public organizations, so this fact should be taken into account when the framework is used on additional public organizations.

5.2 Organizational maturity

The second uncertainty that is addressed as a starting point for the new framework is:

> *Is the organization developed enough to be able to implement Big Data and make the required changes for it?*

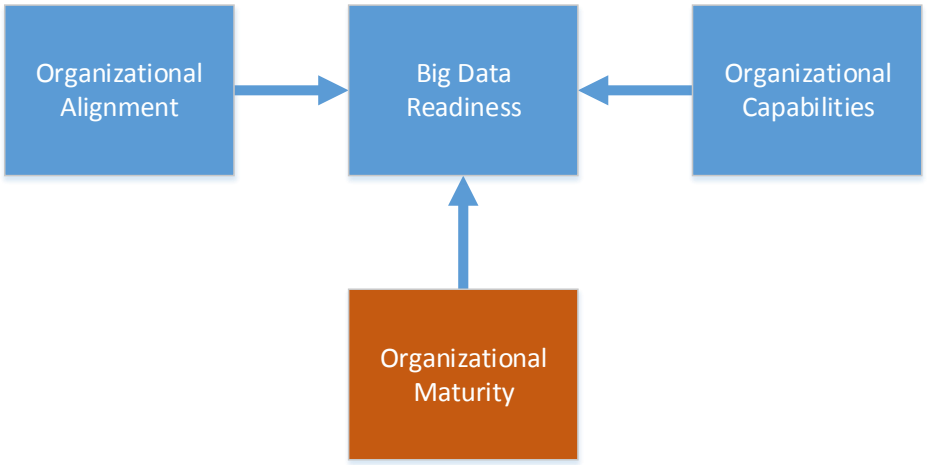


Figure 19: Organizational Maturity

How far an organization is on it development path is often described as the maturity of the organization. If we look at this notion and combine it with the research perspective of Big Data use as an enabler for e-government development, the second uncertainty can be translated into a question on the assessment of the maturity of e-government initiatives in the organization. So, the second set of theories, used as a basis to develop the new framework, is a theory to establish e-government maturity. In these models IT-enabled reforms of government are described to usually follow a certain growth path which are presented in growth stage or maturity models with multiple distinct steps (Layne & Lee, 2001; Andersen & Henriksen, 2006; Klievink & Janssen, 2009; Gottschalk, 2009). Although full assessment of e-government maturity using these framework falls outside of the scope of this research project, the five described stages of the Klievink & Janssen (2009) model and their description are used as a basis to assess in which stage of e-government development public organizations currently are. The Klievink & Janssen (2009) model is chosen, as this model recognizes the inter-organizational aspects of e-government development, which also plays a very significant role in Big Data development.

It is to be noted that development through the different stages of e-government maturity is not necessarily sequential. Steps can also be skipped during the development, for example because organizations make radical changes based on knowledge or experience acquired from outside of the organization.

The e-government maturity of an organization expresses to what extent the organization has developed its e-government initiatives and the organizational characteristics that support these initiatives. In the assessment of Big Data readiness of organizations, the e-government maturity can be used to express the current position of the organization in the development towards the end-state of transformational government. The current position will give a first indication on how much

more organizations will have to develop, before a transformation enabled by Big Data Use implementation can be initiated and is therefore an important part of the assessment of Big Data Readiness.

The e-government growth stages used in this part of the Big Data Readiness Framework are based on the five growth stages identified and described by Klievink & Janssen (2009). This growth stage model was chosen, because each of the five stages was extensively described in the article and because each of the five growth stage were very understandable and recognizable for the researcher. Secondly, the growth stages seemed to resemble the observations made and language used by the interviewed practitioners. And this is important, as practitioners need to be able to understand the framework in order to be able to use it themselves to assess the Big Data Readiness of their own organizations.

The five growth stages are expressed in the Big Data Readiness Framework on three different aspects. This is done to ensure that the scores given by the respondents in the questionnaire provide a rich image on which the assessment of organizational maturity can be based. The three aspects on which the organizational maturity level is scored by respondent are: *activities and information sharing*, *IT facilities* and *data systems*. The aspects *activities and information sharing* and *IT facilities* were taken from the description of the five growth stages in the article of Klievink & Janssen (2009). These two aspects were choses to depict the five stages as these aspects made for understandable and clear differences between the different growth stages that also connected with the language used, expertise and perspectives of the practitioners that were interviewed in the first research step of this research project. Other aspects of the growth stages from the article were not used, as they were less recognizable for the public officials of the public executive sector organizations.

The third aspect that was used to assess in which e-government growth stage the public organizations could be classified did not come from the initial article, but is added to adapt the growth stage model towards the introduction of Big Data use in public organizations. From the interviews a development- or time-path in data systems/concepts within the organizations was created using concept analysis that has a resemblance to the development of e-government growth stages and its aspects in the organizations of the interviewed practitioners. The development path of data systems used in the organizations is presented in Figure 20.

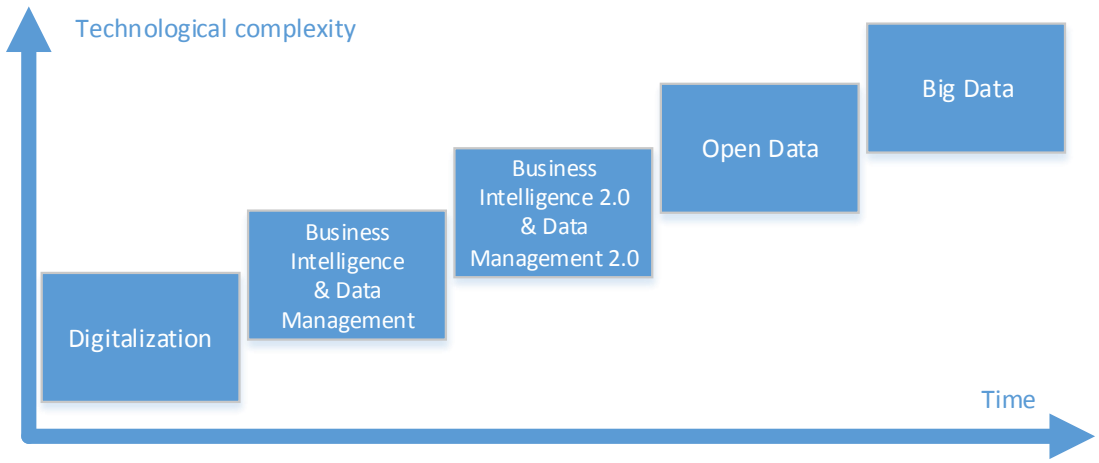


Figure 20: Development of Data Systems

By tying the development of the data systems used in public organizations in the last couple of decades to the five different growth stages from the e-government maturity model, the connection between the maturity of the organization and the corresponding data systems is made visible. In this way the assessment of the organizational maturity is more recognizable for practitioners and better connected to the introduction of Big Data Use in public organizations.

The connection between the five growth stages and the five data systems comes from remarks from practitioners in the interviews when discussing the characteristics and use of more conventional data systems in the organizations and their recent and future development towards Open Data and Big Data. Combined, the remarks from practitioners provided an image of the characteristics of five stages in the development of data systems that is quite similar to the development in the growth stages discussed by Klievink & Janssen (2009). From the remarks in the interviews, the following table with data systems, a short description and its primary use is described.

Data Systems	Explanation	Main Use
Digitalization of processes	Providing digital tools for operational processes in data entry and data search	Automated data creation and storage processes
Business Intelligence & Data Management	Bundling of information storage across the organization and automated analysis tools that create information from internal data in the organization.	Bundle and visualize information from the whole organization
Business Intelligence 2.0 & Data Management 2.0	BI & DM with options to access and change data and information from outside of the organization	Sharing specific data and information from within the organization with selected applications and stakeholders.
Open Data	Providing as much anonymized data as possible from the organization as public data, in a standardized format.	Sharing and exchanging data with many other organizations.
Big Data	Collecting, combining and analyzing large, complex datasets with unconventional technologies to create new knowledge for the organization.	Handling data overload and creating new and granular insights into wide variety of objects and subjects

Table 8: Data Systems

Now the five data systems can be connected to the five growth stages. Practitioners remarked that digital processes were primarily focused on improving and streamlining single processes within single activities and departments of organizations. This is consistent with the stove-piped organizations growth stage. The second data system, Business Intelligence & Data Management, is used by practitioners to bundle and use information from the whole organization to create information for performance optimization and decision making. Bundling and sharing information across the organization is consistent with the integrated organizations growth stage. The development towards Business Intelligence 2.0 and Data Management 2.0 allowed practitioners to

use these data systems not only within the organizations, but also to access these systems from outside of the organization. Also, with this development, the data systems could be structurally connected with selected outside applications in other organizations. The connection with outside application however need to be tailored to each application and are hard to establish successfully and therefore sparse. From the accessible Business Intelligence and Data management systems, the next evolution according to practitioners is Open Data. The Open Data concept is the policy for public organizations to provide as much anonymized data as possible, so that other organizations and individuals, from all sectors can access and use it. However, for practitioners, Open Data in this form is the public tip of the iceberg. The development is much broader than only the provision of anonymized datasets to the public in standardized formats. According to practitioners it is the development of extensive facilities and protocols to share as much information and data as possible between public organizations. In this perspective, Open Data is a collection of data systems that facilitates extensive inter-organizational information and data sharing and therefore fits very well with the Inter-organizational integration growth stage. Lastly, Big Data is a data system that allows the combination and analysis of huge volumes and very complex data. This allows organizations to handle much more data and on a much more detailed, granular level. Big Data technology allows organizations to process and analyze data on the level of individual subjects and objects. And that is precisely what is needed for effective and economically viable demand driven service provision. Big Data allows public organizations to identify the needs and demand of individual citizens and to tailor their service provision to them accordingly. Therefore Big Data can be coupled to the fifth and final growth stage, Demand driven, joined-up government. An overview of the combination of the five growth stages their aspects and the five data systems used for the assessment of organization maturity is presented in Table 9.

	E-government growth stage	**Activities and Information sharing**	**IT facilities**	**Data systems**
1	Stove-piped organizations	Information and activities are bounded within separate departments within the organization	All activities are, where possible, supported by IT	Digitalization of processes
2	Integrated organizations	Activities transcend the separate departments and information is shared throughout the organization	A organization-wide IT infrastructure in which information is accessible throughout the organization	Business Intelligence & Data Management
3	Nationwide portal	The activities go beyond the boundaries of the organization and information is accessible from outside of the organization	IT infrastructure suited for external access of applications and information within the organization	Business Intelligence 2.0 & Data Management 2.0

4	Inter-organizational integration	Activities and information is extensively sharing with other organizations	IT infrastructure suited for full access of applications and information for other organizations	Open Data
5	Demand driven, joined-up government	Activities and information is organized centrally and made available to all relevant organizations and stakeholders	Centrally Built IT facilities supporting all information and applications and fully accessible for all involved stakeholders	Big Data

Table 9: E-government growth stages and Big Data

Maturity assessment

By questioning representatives of the organizations how strong each of the factors is present in their organization a first indication of the most dominant e-government growth stage can be developed. This will give important insight in how mature the organization is in respect to e-government initiatives and Big Data enabled transformational government in particular. The assessment decision rules of the organizational maturity assessment can be found in Appendix IV on page 140.

Validity of Organizational Maturity

As for the first aspect of the Big Data Readiness framework, a number of remarks on the validity of this aspect of the framework are made. In the case of the organizational maturity aspect, two remarks on it validity should be taken into account when the framework is used. Firstly, the e-government growth stage model used is primarily aimed at public organizations with a service provision function towards citizens. However, this function is not always present in public executive organizations, as they do not necessarily primarily provide services to citizens directly. These organizations are also often supporting organizations, that provide tasks that support service provision by other public organizations. So, while the growth stages used in the framework seem to be applicable to the eleven organizations that participate in this research project, they do not necessarily have to do so for other public executive agencies. Even in the assessment presented in chapter six it becomes evident that the assessments on organizational maturity are quite low. This could be because not all the organizations assessed have a primary service provision function to citizens and have therefore not developed their organizations in the direction recognized by the e-government growth stages used in the framework. And although, the results of the assessment on organizational maturity do not indicate a higher assessment for service provision aimed public organizations in this research project, this remark should be taken into account in the interpretation of the organizational maturity assessments of additional public organizations.

A second remark that needs to be made on the validity of the organizational maturity aspect of the Big Data Readiness Framework is that the perspective on the data systems is adapted from their academic concepts towards the perspectives practitioners take on them and how they are used in

practice. While this provides the framework possibly with a higher applicability for practitioners, it does somewhat decrease the academic correctness, understandability and communicability of this part of the framework. Furthermore, as the adaption is based on the interpretation of just eleven public officials, the adaption may not be broadly supported in a wider circle of practitioners. Users of the framework should regard this fact in their interpretation of the assessments of organizational maturity of other organizations.

Finally, the factor of security, privacy and legality of information and data sharing between organizations is not taken into account in this part of the Big Data Readiness assessment. It is part of the third aspect, organizational capabilities, but the factor also has consequence on the organizational maturity aspect. When organizations possess and handle sensitive information, that needs to be protected and that cannot be shared freely with other organizations and individuals without proper protection, their activities, facilities and data systems may not be geared towards data sharing and therefore their organizational maturity will receive a lower assessment than may be the case if less sensitive data was used. While a large amount of sensitive data will make working with Big Data more difficult to do successfully and fully protected, it does not necessarily affect the maturity of an organization for their development towards Big Data. The sensitivity of the data handled within organizations is not comprehensively addressed in this aspect of the Big Data Readiness assessment and that fact should be taken into account in the interpretation of its results.

5.3 Organizational Capabilities

The third uncertainty that is addressed as a starting point for the new framework is:

Is the organization capable of working with Big Data successfully?

The uncertainty is if the organization possessed the organizational capabilities need to use Big Data, create value for the organization from it and make sure no negative consequences of their Big Data use arise. So, to establish if public organizations are ready for Big Data use implementation, an overview of the required organizational capabilities for Big Data use and a method to assess them is required.

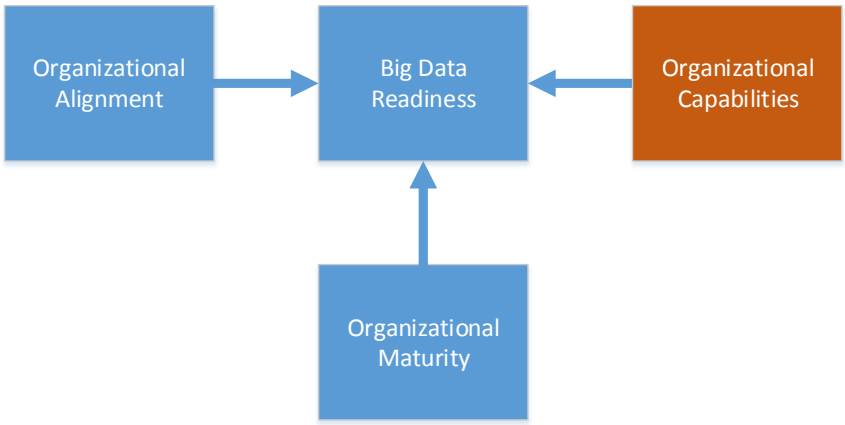


Figure 21: Organizational Capabilities

Klievink & Janssen (2009) further develop the e-government growth stage model concept by adding that advancement towards further stages of e-government development are based on the development of the required capabilities for those stages. The notion of capabilities enabling advancement brings us to a relevant assessment method from literature.

Valdès et al. (2011) developed an assessment method, named the e-government maturity model, based on capability levels relevant to e-government initiatives implementation and execution. The methods used in this maturity model can be adapted towards Big Data use implementation and execution to assess the specific readiness for Big Data within public organizations by using a different set of capabilities, specifically geared towards Big Data use.

The organizational capabilities considered vital for Big Data Use in the public sector organizations are established through analyses of capabilities mentioned in several, relevant research fields in literature. Capabilities from literature in IT Adoption (Jeyaraj, Rottman, & Lacity, 2006; Kamal, 2006; Ebrahim & Irani, 2005), IT Implementation (Ngai, Law, & Wat, 2008; Premkumar, 2003; Yeoh & Koronios, 2010; Finney & Corbett, 2007; Wixom & Watson, 2001), Innovation Adoption (Waarts, Everdingen, & Hilligersberg, 2002; Robey, Im, & Wareham, 2008), Dynamic and Core Capabilities in IT (Daniel & Wilson, 2003; Lin & Hsia, 2011; Klievink & Janssen, 2009; Wu & Hisa, 2008) and from Big Data specific literature (Chen, Mao, & Liu, 2014; Courtney, 2012; McAfee & Brynjolfsson, 2012; Milakovich, 2012; Tambe, 2014) are combined with repeatedly mentioned capabilities in the unstructured interviews.

First of all, all conditions important for IT and IS implementation from literature are identified from the articles in Table 10. The articles have different topics and theoretical backgrounds, but therefore provide a broad and diverse picture of all considered conditions for IT implementation in literature. Articles are selected on relevance of core topic (IT implementation/adoption by organizations), on number of citations and unique contribution of mentioned conditions that are considered applicable to Big Data use.

#	Author(s)	Year	Theory	Case(s)/Topic
1	**Kamal**	2006	IT innovation adoption	general IT innovations
2	**Zutshi & Sohal**	2004	Organizational adoption	EMS adoption
3	**Feeney & Willcocks**	1998	Core capabilities	General IT exploitation
4	**Finney & Corbett**	2007	IT implementation	ERP systems
5	**Daniel & Wilson**	2003	Dynamic capabilities	E-business
6	**Jeyaraj, Rottman & Lacity**	2006	IT innovation adoption	general IT innovations
7	**King & Burgess**	2005	IT implementation	ERP systems
8	**Lin & Hsia**	2011	Core capabilities	E-business innovation
9	**Klievink & Janssen**	2009	Dynamic capabilities	E-government
10	**Ngai, Law & Wat**	2008	IT implementation	ERP Systems
11	**Robey, Im & Wareham**	2008	Diffusion of innovation	IOS
12	**Premkumar**	2003	IT adoption	General IT
13	**Waarts, Van Everdingen & Van Hillegersberg**	2002	Adoption of innovations	ERP systems
14	**Wixom & Watson**	2001	IT implementation	Data Warehouse Systems
15	**Wu & Hisa**	2008	Dynamic capabilities	E-M-U-commerce
16	**Yeoh & Koronios**	2010	IT implementation	Business Intelligence
17	**Valdes et al.**	2011	IT governance	E-government
18	**Veenstra**	2013	Empirical	Big Data
19	**Ebrahim & Irani**	2006	IT adoption	E-government

Table 10: Organizational capabilities from literature

The capabilities are presented in Table 11 and selected after a content analysis on the conditions mentioned and found important and/or significant in the literature and interviews with practitioners. Alongside the conditions from literature, the most often mentioned and discussed conditions for Big Data use from the unstructured interviews are presented in the same table.

Capabilities most used in literature	Capabilities discussed in interviews
Relative advantage of new IT systems	IT management
IT expertise of organization	Using data in decision making and operations
IT facilities of organization	Data quality and availability
Presence of a champion for new IT systems	Internal and external attitude towards Big Data
The influence of a vendor	IT infrastructure
Ability to develop new business cases/strategy	Privacy and security issues
Ability to build internal and external commitment	Data Science Expertise
Top management support of implementation	Quality and transparency of analytics
Alignment of strategy/operations and new IT	Visualization of data and results
Ability to integrate new IT systems	Data use, re-use and ownership
Data management	

Table 11: Organizational Capabilities

When both sets of capabilities are selected on frequency and relevance, similar capabilities are combined, and formulated in light of Big Data use and its implementation, seven Organizational Capabilities can be established. The selection of the capabilities is based on whether it was mentioned in both literature and interviews, was applicable to Big Data in the public sector and adapted and combined towards a level of detail that was similar over all seven capabilities. Furthermore, the capabilities were formulated in such a way that they were understandable for both practitioners and academics, to promote the applicability of the Big Data Readiness Framework. The seven capabilities are presented in Table 12. Also a short explanation per capability is provided in the following section. Furthermore, the articles from literature that support each of the seven formulated capabilities are presented in the right column.

Capability	Explanation	Literature
IT Governance	Capability to design and develop IT strategy, decision making and responsibility structures, supporting the organization. This includes integration of new IT systems.	1-15, 17-19
IT Resources	Capability to design, develop and maintain suitable IT infrastructure and expertise to facilitate current and new IT systems.	1, 4, 6, 10
Internal attitude	Capability to develop internal commitment and vision for new processes and systems. Especially openness towards data driven decision making.	1-7, 9, 10, 12, 14, 16, 17, 19
External attitude	Capability to develop external commitment and support for new processes and systems with important stakeholders.	1, 4-9, 11-13, 15
Legal Compliance	Capability to design and develop a compliance strategy including process design, monitoring and redesign of processes. Specifically for data use on privacy protection, security and data ownership regulations.	6, 9, 11, 17-19
Data Governance	Capability to design and develop a data strategy including collection, acquisition, quality control and data partnerships.	14, 16-18
Data Science Expertise	Capability to bundle/acquire, develop and retain data science knowledge in the organization. Specifically bundling of IT, core business & mathematics knowledge.	-

Table 12: Organizational Capabilities

IT Governance

IT Governance is the purposeful design and execution of decision making processes and responsibilities concerning IT in the organization. This means that the organization has a clear plan how responsibilities and decision making is distributed in the organization around all present IT. Various forms of IT Governance are possible, from a very centralized approach, to a very fragmented and decentralized approach and organizations should use an approach that fits their specific needs and circumstances. The main point for Big Data use is however, that IT Governance should be well designed and very clear throughout the organizations, as the introduction of Big Data technologies will be a very large implementation project of a radically new IT system that will have an effect on almost all other aspects of the organizations IT facilities and governance.

Advantages in productivity from the diffusion of a new information technologies can, in most cases, only be achieved by organizations when business processes are redesigned to accommodate for the new possibilities (Grover, Teng, Segars, & Fiedler, 1998). And Big Data technologies will most likely be no exception. The new possibilities that the new Big Data technologies will offer the organization will most likely require (partial) redesign of existing systems and processes and this should be well understood and handled by IT decision makers for Big Data technologies to pay off for the organization. The organization has to be capable of creating and IT Governance program that can

accommodate the implementation of Big Data technologies and applications for it to be ready to start using Big Data in the organization.

IT Resources

The IT Resources of an organizations are all of its current IT facilities. This includes current IT infrastructure, such as its hardware, network, and enterprise systems. Furthermore, internal IT expertise on design, maintenance and development of the IT infrastructure is an important part of the IT resources of an organization.

The IT resources of an organization are important for the success of possible Big Data applications, as these applications will in a large part make use of the current IT resources and the expertise on them within the organization. If current IT resources in the organizations are not suited for the demands of Big Data applications, or not compatible with Big Data technologies, the implementation of Big Data applications will be much harder to execute successfully. The organizations should be capable to create IT Resources that are compatible with Big Data technologies and that can support the Big Data applications it plans to implement, before the organization can start using Big Data.

Internal Attitude

From the literature mentioned earlier in this section on organizational capabilities one very important critical success factor for successful IT implementation can be widely recognized and that is the support of top management in the organization for the implementation of the new IT system (Premkumar, 2003; Jeyaraj, Rottman, & Lacity, 2006; Ngai, Law, & Wat, 2008). There is no reason to believe that Big Data technologies should be an exception to this observation and so the Big Data Readiness Framework will also have this factor incorporated in its assessment method. Going further than just the support of top management, for a successful introduction of Big Data use in an organization, the support of other parts of the organization are also important. It is therefore important for organizations to be capable to create a broadly carried support for technology change and implementation projects in the organizations before the should start with implementing Big Data technologies.

Secondly, to really benefit from the introduction of Big Data use in the organization, the organization should adapt its decision making culture more towards a data driven decision making process. Only if data is structurally used in decision making processes will organizations truly use the new information that is created with the use of Big Data applications (McAfee & Brynjolfsson, 2012; Ross, Beath, & Quaadgras, 2013).

Combining the broad support for Big Data technology implementation and the adaption towards data driven decision making processes will create a new culture more oriented towards intensive data use in the organization. Data oriented cultures within organizations are beneficial to the value that big data has for these organizations. Important data oriented characteristics are use of analytics on data, top management support for analytics and widely available data and data insights (Kiron & Schockley, 2011). Organizations should be capable of changing their current culture towards a more data driven culture in order to take full advantage of Big Data applications.

External Attitude

The external attitude towards Big Data use in the organization is important for the successful introduction of Big Data technologies in the organization. If there is no successful cooperation with consultants, suppliers and experts on Big Data technologies and applications, successful implementation of Big Data applications will be much more difficult for the organization. Secondly, without support from higher level governmental bodies and with consent of citizens and other stakeholders, the successful use of Big Data in the organization will be much more difficult. The organization has to be capable of managing expectations and demands from external stakeholders before Big Data can be used successfully in the organization.

Legal Compliance

Being able to comply to all relevant legislation in the handling of Big Data is another important capability of organizations that plan to start using Big Data. Because data is often sensitive and valuable, sufficient security measures need to be in place to protect the data from unwarranted access by external and internal parties. Not only could this unwarranted access lead to undesired leaking information, it could also negatively affect data quality if undesired adaptions are made to the accessed data. Especially when data is owned by another party, unwarranted access and possible detriment effects to data quality are particularly painful. Using Big Data means security must be sufficient for more data, more varied data and more incoming and outgoing data streams. Organizations need to be capable of managing the required security before they start using Big Data.

Secondly, privacy legislation is of vital importance in the handling of Big Data. Especially in public organizations often data consists of a lot of personal information on citizens. Not only is the security of this data of importance, but also the protection of the privacy of the citizens itself. In privacy legislation the collection of data is controlled, data can only be collected for a predetermined legal purpose, or with consent of each person the data is collected on. Also storage of this data is restricted in time. Using the data again is only allowed after it has been stripped from all personal information and no identification can take place on the basis of the dataset. However, as Big Data mostly involves combining multiple datasets, the amount of data per person increases to such levels that identification is again possible, which entails a breach of privacy legislation. Mayer: Current legislation on privacy does not provide sufficient guidelines for re-use and combination of datasets, which makes legal compliance on privacy protection in Big Data applications much more difficult (Mayer-Schönberger & Cukier, 2013). Big Data makes protection of privacy much harder for organizations and they need to be aware and capable to manage sufficient protection of privacy before they can start to use Big Data.

Thirdly, with ever increasing amounts of data and possibilities to access and use that data, data ownership starts to play a larger role. A lot of data in our world is freely collectable, but some data is not. This data is owned by other organizations or citizens and cannot be freely collected and used for analysis. Not all data available to organizations is rightfully usable by them and while data ownership is currently hard to establish, check and enforce, it will become very important in the future and organizations should be capable of managing the ownership of the data available to them before they start with using Big Data.

Data Governance

Data governance is the term used to describe all activities related to successfully handling the data needed for the Big Data applications desired by the organization. Extensive, successful data activities in the organization require a strategy that makes all of the data in the organization known, available, accessible, combinable and of sufficient quality (Adelman, Moss, & Abai, 2006). For data governance to be successful, the organization should be capable of successfully designing, executing and improving this strategy for all available data for the organizations.

For Big Data applications to be successful within organizations, especially, the quality of the data available should be managed and optimized as much as possible (Chen, Mao, & Liu, 2014). That entails that both currently existing data and the collection of data should be optimized as much as possible. A higher data quality will lead to better quality information from the analysis of the data and of faster analysis, as data is more understandable for the analytical tools.

Lastly, data governance also involves the successful acquisition of external datasets, policies of sharing internal data with external data and the creation of data partnerships that will regulate data sharing between organizations for a longer term. The organization should be capable of successfully governing all of the aspect of its data before it starts with using Big Data, as the quality of the raw material used for the Big Data applications will, for a large part, determine the quality of the produced information from the Big Data applications. The importance of data governance of an organizations is further emphasized by the relationship between higher productivity gains from Big Data investments and the availability of high quality data assets in firms (Tambe, 2014).

Data Science Expertise

Data Science Expertise in the organizations is the expertise required to successfully design, build and use Big Data applications in the organization. The expertise in data science entails a combination of extensive knowledge of three separate fields of expertise. Firstly, expertise on IT and computer programming, to be able to use, design, build and adapt Big Data applications, the required specific software solutions for the organization and to design, write and adapt the algorithms required for the applications. Secondly expertise of mathematics and statistics is required to be able to understand, use, evaluate and interpret advanced analytical methods to analyze the data in the applications. Thirdly, a high degree of expertise is needed on the core business of the organization. This is required to make sure the Big Data applications are suited to help with the core activities of the organizations and provide useful and actionable information for the organization.

As people who are experts in all three areas will be very rare, extensive training programs will be required to acquire enough people with data science expertise. Of course, another solution could be to make use of multidisciplinary teams in which experts on all three areas work closely together. Also, data science expertise could be acquired by hiring external parties, such as consultants or independents. Which way data science expertise is acquired is of no consequence, however, organizations do have to have this expertise available to them to make effective use of Big Data technologies within their organization. Organizations have to be capable of bundling or acquiring data science expertise before it can start with using Big Data in the organization.

Capabilities assessment

The seven identified organizational capabilities are assessed on three dimensions; importance for Big Data success in the organization, possibility to develop capability in the organization and current presence of the capability within the organization. Based on the maturity assessment methods used by Valdès et al. (2011), the maturity on each capability is derived from comparing given scores to the maximum score. This will give a maturity level per capability for each organization, which can be used to compute overall maturity levels per organization and organization type. This will give further insight into the Big Data readiness of public sector organizations in the area of Organizational Capabilities for Big Data use. The decision rules of the organizational capabilities assessment can be found in Appendix IV on page 140.

Validity of Organizational Capabilities

As with the previous aspects of the Big Data Readiness framework, a remark can be made about the validity of the aspect of organizational capabilities. The primary remark on the validity of the organizational capabilities selected for the Big Data Readiness framework that are presented in the previous section is that the capabilities are all formulated without a proven connection with successful Big Data use. The capabilities from literature have, in most cases, a proven connection with the successful implementation and use of other IT systems and innovations, but not with Big Data, as empirical evidence of successful Big Data use is very sparse and nonexistent in regard to organizational capabilities in connection with it. Furthermore, the capabilities discussed in the interviews are based on assumptions from practitioners on capabilities important for future Big Data applications, but not actual experience with Big Data use in the organization. Therefore, the capabilities used in this aspect of the Big Data Readiness Framework are not based on empirical evidence on the successful use and implementation of Big Data in organizations, but on substantiated assumptions on which capabilities that are important for Big Data use and implementation. Practitioners and academics using the framework should consider this remark in their interpretation of the results of the assessment.

6. Findings

After the formulation of the Big Data Readiness framework and its three aspects, the framework is used to assess the Big Data Readiness of eleven organizations in the Dutch public executive sector. The interviewed representatives of the organizations from the Dutch public executive sector are asked to fill in a questionnaire designed to gather information on the organizations Big Data readiness. The Big Data readiness of each organization is determined for each of the three parts of the framework and combined to establish Big Data Readiness scores for each organization. Furthermore, the Big Data Readiness assessments are used to formulate potential areas for improvement of the Big Data Readiness in the participating organizations.

6.1 Big Data Readiness Assessment

In this section the Big Data Readiness is presented. First the findings on organizational alignment are described, followed by the findings on organizational maturity and then organizational capabilities.

6.1.1 Organizational Alignment

The framework and assessment method used to assess the organizational alignment of the organizations in the Dutch public executive sector are described in section 5.1 on page 56. In the following section the findings on the organizational alignment of the organizations in the Dutch public executive sector is presented

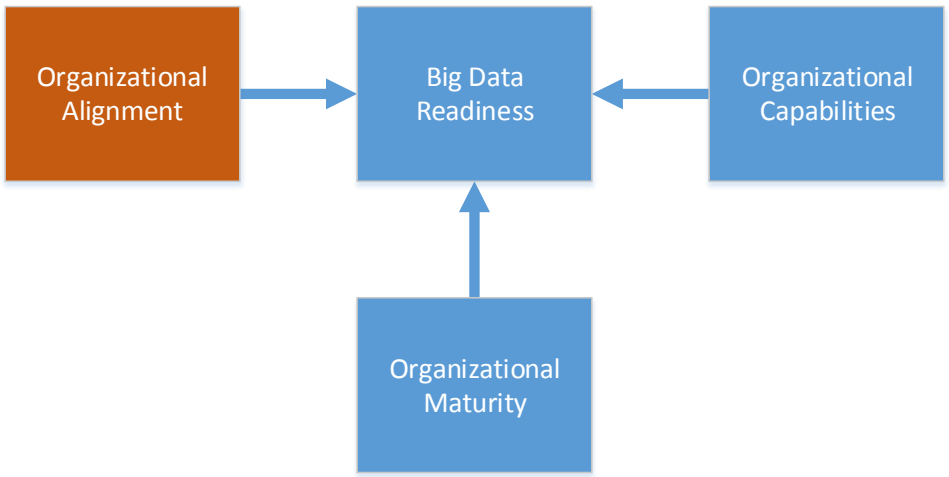


Figure 22: Organizational Alignment

As discussed in section 5.1 the organizational alignment is assessed by identifying if the organizations main statutory task and current data activities of the organizations are aligned with the planned Big Data application type(s) and corresponding Big Data use characteristics. The basis of the assessment is to compare the current alignment of the organizations with the proposed optimal alignment, which is presented in Table 7 in section 5.1 on page 61.

First, the results of the questionnaire are used to establish the main statutory task types of the public executive organizations of the eleven participating public officials. A fairly equal distribution of the eleven organizations on the four different types of statutory tasks was found. This first identification step grants an overview of the organizational strategy and infrastructure aspects of the strategic alignment model per organization and is presented in Table 13. The overview in the table shows that the organizations assessed in the research project have differing experience with data activities and different types of statutory tasks. This means that the assessed organizations will most likely come from different positions in their development towards Big Data, which will use the full extent of the scope of this aspect of the Big Data Readiness assessment.

Organization	Main statutory task type	Data Collection Activity Intensity	Data Use Activity Intensity
1	Administration, Management	High	High
2	Administration, Management	High	High
3	Coordination, project-based task, no data use	Low	Low
4	Administration, Management	High	High
5	Registration, Documentation	High	Low
6	Registration, Documentation	High	Low
7	Research, Evaluation	Low	High
8	Research, Evaluation	Low	High
9	Coordination, project-based task, no data use	Low	Low
10	Research, Evaluation	Low	High
11	Research, Evaluation	Low	High

Table 13: Main statutory Tasks and Data Activities

Secondly, the IT strategy and infrastructure aspects are derived from the interest, need, applicability and feasibility answers in the questionnaire of each representative per Big Data application type. This results in an overview of the most prominent Big Data application types planned for at every organization. The overview of planned Big Data applications at the organizations provided a somewhat scattered image, in which five out of the eleven organizations were almost equally interested in two or more of the three Big Data application types. The scattered image shows that, on average, the organizations in the Dutch public executive sector are not considering the alignment

between the current situation in their organization and the required situation for their planned Big Data applications. This is a first indication that the organizations in the Dutch public executive sector may not be ready to start with using Big Data in their organizations. Public organizations in the Dutch public sector seem to have a limited understanding of Big Data applications and what the different types of applications will entail for their organizations and are therefore not ready to start using Big Data.

Combining the results of the first two steps in the alignment assessment and adding the presence of the five Big Data use characteristics in the current data activities of the organizations delivers an overview of the alignment between Big Data application types and the organization. From this overview a complete assessment of the organizational alignment for each organization can be made. For this assessment the values Very High, High, Medium, Low and Very Low are given, based on the decision rules presented in Appendix IV on page 140 and presented in Table 14.

Organization	Main statutory task type	Alignment Assessment Value
1	Administration, Management	Very High
2	Administration, Management	Medium
3	Coordination, project-based task, no data use	Low
4	Administration, Management	High
5	Registration, Documentation	Very Low
6	Registration, Documentation	Medium
7	Research, Evaluation	High
8	Research, Evaluation	Medium
9	Coordination, project-based task, no data use	Very Low
10	Research, Evaluation	Medium
11	Research, Evaluation	Medium
Average Alignment Assessment Value		**Medium**

Table 14: Organizational Alignment Assessment

The results in the table indicate that the overall organizational alignment of the considered Big Data applications types at the Dutch public executive organizations is assessed as Medium. Very few of the assessed organizations are planning for Big Data applications that are aligned with the current

situation in the organization. This indicates that organizations in the Dutch public executive sector are not ready to start with using Big Data, as they generally plan to do that in Big Data applications that do not suit their current organizations, which will severely limit the chances of successful implementation and successful future use of the currently planned Big Data applications. Public organizations in the Netherlands are considering Big Data applications that are not suited to their current organizations, which will lead to very hard implementation processes and limited success of Big Data use in these organizations. Public organizations need to be careful in planning for Big Data applications that their current data activities do not support. Big Data applications cannot operate successfully without being supported by intensive data collection and/or data use activities within the organizations.

What stands out from the organizational alignment assessment is that the range of the assessment values is very large. In the limited number of the eleven assessed organizations values from Very Low to Very high are being scored, indicating differences in the understanding of organizational alignment of Big Data applications are very large between the organizations in the Dutch public executive sector. The organizations in the Dutch public executive sector seem to have a very differing understanding of Big Data applications and the implications for their organizations.

A possible explanation for the large difference in the understanding of the organizational alignment of planned Big Data applications is the difference in knowledge and experience with intensive data activities between the organizations. It may be the case that organizations less experienced with intensive data activities are not as capable as more experienced organizations in judging the consequences and requirements of different types of Big Data applications for their organization and are therefore not as capable in selecting the best aligned Big Data applications for their organization. To investigate this possible explanation, the average alignment assessment values are taken for each of the four identified organization types with the corresponding statutory task types and data activity intensities. The result is presented in Table 15.

Main statutory task type	Data Collection Activity Intensity	Data Use Activity Intensity	Average Alignment Assessment Value
Coordination, project-based task, no data use	Low	Low	Very Low - Low
Research, Evaluation	Low	High	Medium – High
Registration, Documentation	High	Low	Low
Administration, Management	High	High	High

Table 15: Average Alignment Assessment values per main statutory task type

Before the results presented in Table 15 are discussed, it must be noted that, due to the very limited number of organizations assessed per organization type, the results are severely influenced by the assessed values of individual organizations and should therefore not be interpreted as a general observation for the entire Dutch public executive sector without that consideration.

What becomes clear from Table 15 is that the organization types in which intensive data use is performed, which are the second and fourth type of organizations, are assessed significantly higher on organizational alignment than the other two types of organizations that do not perform data use very intensively. More experience with intensive data use indeed seems to result in a better understanding of using Big Data and its applications, which can explain the higher scores for these two types of organizations on the alignment assessment. The intensity with which data collection is performed in the organizations seems not to have as much positive effect on the understanding of the alignment of planned Big Data applications and the current situation of the organization. It therefore seems that experience with intensive use of data is more important for the Big Data Readiness of organizations than experience with intensive data collection. Although this connection is quite obvious and therefore not remarkable in itself, it could provide an important point for the Big Data Readiness of public organizations and the development of it. Organizations in the Dutch public executive sector have a better understanding of Big Data when they are more experienced in structurally using more conventional data in their organizations. With the experience from using more data a better understanding of the implications of the use of Big Data can be developed, giving these organizations a head start in their development towards the implementation of the use of Big Data.

6.1.2 Organizational Maturity

In section 5.2 on page 63 the second part of the Big Data Readiness Framework is presented. In that section the assessment of the organizational maturity of public organizations for the introduction of Big Data use is discussed. The organizations are assessed on their readiness to progress and change towards a more digital and demand driven organization enabled by the introduction of Big Data use. This assessment is done by identifying the current dominant e-government growth stages of the organizations in the Dutch public executive sector and establishing how large the changes for the organizations would be towards Big Data use enabled reform. The larger the changes for organizations are, the more difficult it will be for them to successfully implement and use Big Data technologies in their organization and to take real advantage from them.

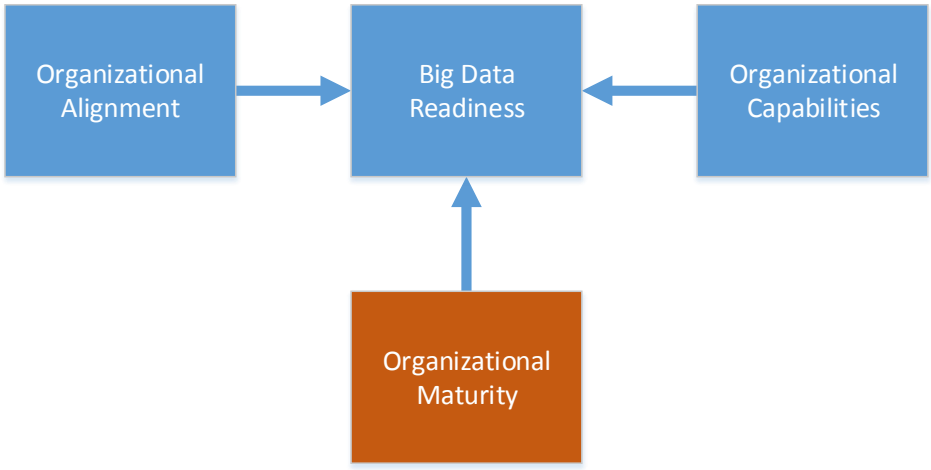


Figure 23: Organizational Maturity

In Table 16 the identified Organizational Maturity Levels are presented for each of the eleven organizations of the representatives of the Dutch public executive sector. The maturity levels per organization are determined following the assessment scorecard rules presented in Appendix IV on page 140. The scores per maturity stage are generated from the three aspects that are discussed in section 5.2; Activities and Information sharing, IT facilities and Data systems used.

Organization	Main statutory task type	Maturity Level	Maturity Assessment Value
1	Administration, Management	2	Low
2	Administration, Management	3	Medium
3	Coordination, project-based task, no data use	1	Very Low
4	Administration, Management	2	Low
5	Registration, Documentation	5	Very High
6	Registration, Documentation	2	Low
7	Research, Evaluation	2	Low
8	Research, Evaluation	1	Very Low
9	Coordination, project-based task, no data use	1	Very Low
10	Research, Evaluation	1	Very Low
11	Research, Evaluation	2	Low
Average Maturity Level and Assessment		**2**	**Low**

Table 16: Organizational Maturity Assessment

The results presented in Table 16 show that the current organizational e-government growth stages of the eleven organizations is not at a level close to the fifth and final, transformational stage for which Big Data use can be a key enabler. The Dutch public executive sector seems to have to change and develop very significantly before actual digital, flexible and demand driven government can be realized and Big Data can be used optimally for the final transition towards it. For most of the organizations the development needed towards an e-government growth stage that will enable them to take advantage of the full potential of Big Data use is very large and will take significant time and effort to execute successfully. Organizations in the Dutch public sector do not seem to currently perform activities that support extensive sharing and exchanging of data and information between organizations. This makes it harder for the organizations to successfully use Big Data, as the available data is limited in its amount and especially variety, which limits the knowledge they can create from it.

However, two remarks have to be made here. Firstly, as stated earlier, growth stages do not necessarily need to be progressed through consecutively. Stages can be skipped in the development of organizations, so organizations could potentially jump towards more advanced growth stages in a short period of time, if the right knowledge is developed and the right changes are made. However, the extent of changes needed within the organizations is properly shown with the help of their current dominant e-government growth stage. Secondly, the original framework of e-government growth stage of Klievink & Janssen (2009) from which the organizational maturity framework part is derived, is primarily aimed at service provision functions within public organizations and its corresponding organizational structures and requirements. As not all organizations assessed for this research project are public organizations with an extensive and dominant service provision function, the five used growth stages may not fully apply to all of them. As these organizations are not aimed at service provision, they have had no need to develop the organizational structure that would support it. And that could cause that these organizations are assessed with a lower maturity level than would be given based on a maturity assessment not designed with service provision in mind. Also, the representatives participating in this research project are experts on the data activities of the organizations and will have a different perspective on the service provision functions of their organizations. However, it remains that the three aspects of the e-government growth stages used in the assessment, Activities and Information sharing, IT facilities and Data systems, are important aspects of the organization in light of Big Data use. Therefore, the organizational maturity assessment does provide useful information of how ready organizations in the Dutch public executive sector are to start using Big Data.

So, as becomes clear from Table 16 is that organizations in the Dutch public executive sector have a Low organizational maturity, which indicates that the development and change needed before the potential value of Big Data use can be fully grasped is quite significant. The average maturity level in the Dutch public executive sector is currently level two. Level two indicates that organizations have currently mostly integrated activities, information sharing, IT facilities and data systems across departments within the organization, but are not structurally cooperating with other organizations on these aspects. Again, this limits the data available as input of Big Data applications, limiting the value they can add to the organization. The fact that most organizations have not integrated these aspects across organizational borders can be explained by that the field and corresponding statutory tasks of the organizations are highly specialized and often no comparable organizations with similar goals are present to cooperate and integrate with. Secondly, integration of the three aspects used in the assessment creates little value for organizations if it is done with organizations in a different field and with very different tasks and goals. Up until now that is. With the introduction of Big Data use, wildly differing data from other fields and organizations is becoming more valuable to organizations in the Dutch public executive sector. Development of information and especially data sharing and cooperation will add significant value and quality to Big Data initiatives and will become vital if the full potential of Big Data applications is to be reached by public organizations.

This means that the interviewed organizations all have extensive development and organizational changes ahead in the area of cooperation and integration of activities and information sharing with other organizations, before they are ready to implement Big Data use and really use it as an enabling technology to develop further towards a transformational form of government. Before organizations in the Dutch public executive sector can take full advantage of Big Data applications, they need to secure more and especially more varied data to use as input. This should be done by creating more structural data and information sharing facilities and activities with other (public) organizations. Now, Dutch public organizations cannot take full advantage of the potential of Big Data applications

and create the precise and granular insights needed for the organizations to develop towards demand driven, joined-up government forms.

The image drawn from the assessment of the organizational alignment that Data Use intensive organizations score higher on the Big Data readiness assessments is not continued in the organizational maturity assessment, as no clear connection between organization types and the results can be established.

6.1.3 Organizational Capabilities

The third part of the Big Data Readiness assessment of the organizations in the Dutch public executive sector is the assessment of the organizational capabilities for the use of Big Data. In section 5.3 on page 69 the framework for this assessment is presented. The findings from the assessment of the organizational capabilities is presented in this section.

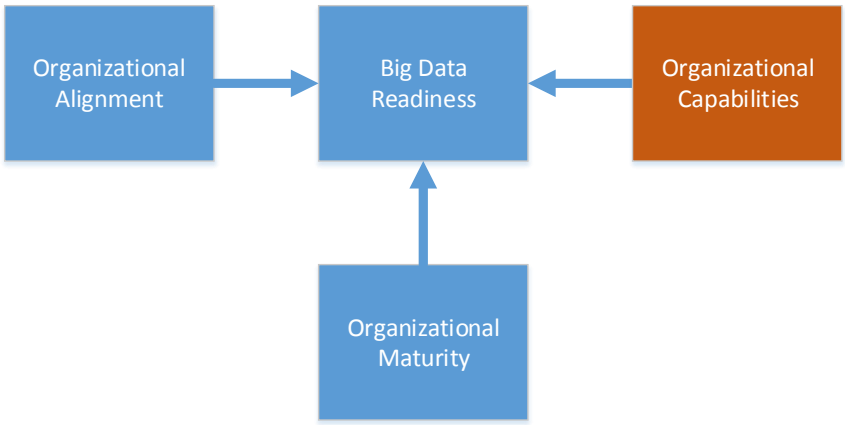


Figure 24: Organizational Capabilities

To assess the organizational capabilities for Big Data use of the participating organizations from the Dutch public executive sector the seven identified capabilities are scored by the representatives in the questionnaire. The capabilities are IT Governance, IT Resources, Internal attitude, External attitude, Legal Compliance, Data Governance and Data Science Expertise. The seven capabilities are scored on a seven point scale on three dimensions: its importance in the organization, whether it could be developed in the organization and its current presence in the organization. By summing these scores, an overall score per capability per organization is established, which indicates how well established that particular capability is within the organization. By averaging the scores on all capabilities per organization, an assessment can be made on how well the organization would be capable of handling Big Data and its implementation in the organization. And that is an important indicator for the Big Data Readiness of the organization. In Table 17 on page 86, the average scores over all capabilities are presented in percentages of the maximum possible average score.

At first glance, the organizations in the Dutch public executive sector are reasonably capable of handling Big Data in their organization. Only a few organizations are assessed with a Low value on organizational capabilities, and the rest is assessed with Medium or High values on their average scores over all capabilities. Overall, it seems that the organizations in the Dutch public sector on average have reasonably well-developed capabilities in respect to using and implementing Big Data in their organizations. However, the assessment made here is based on the average score per

organization and lacks information on individual organizational capabilities. The average scores per capability over all assessed organizations is presented in Figure 25 on page 87. From the figure the same image as from the table earlier becomes clear. All capabilities receive an average score between 70% and 81% and can therefore be assessed with a Medium or High value in the Big Data Readiness Framework. So, also in this view the Dutch public executive sector seems to be reasonably well-developed in its capabilities to use and implement Big Data technologies. The table and diagram show that Dutch public executive organizations are close to being capable of implementing and using Big Data in their organizations. However, some weak points remain and they need to be addressed before Big Data use should be widely implemented in the Dutch public executive sector.

Organization	Capabilities score	Capabilities Assessment Value
1	76%	Medium
2	72%	Medium
3	69%	Low
4	86%	High
5	70%	Low
6	72%	Medium
7	82%	High
8	81%	High
9	65%	Low
10	86%	High
11	75%	Medium
Average score	**76%**	**Medium**

Table 17: Organizational Capability Assessment

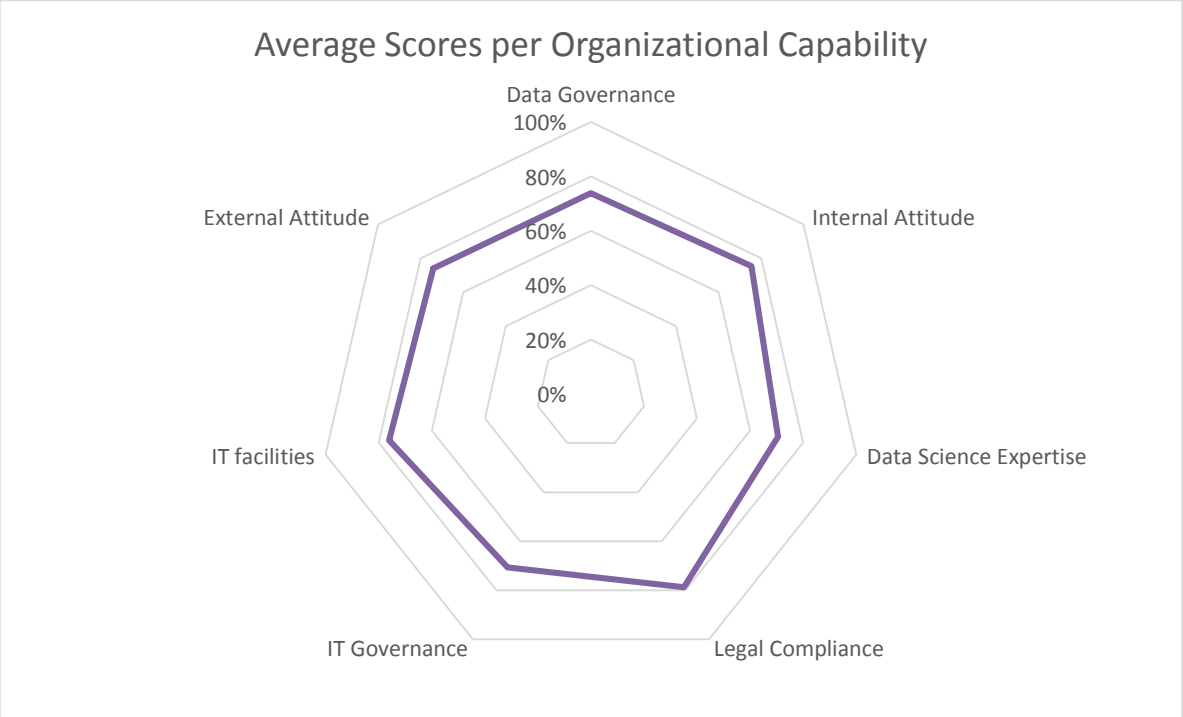


Figure 25: Average Scores per Organizational Capability

The assessment of Big Data Readiness on the organizational capabilities is based on the average scores over all assessed capability per organization. The danger of using average scores is that information on the scores of the individual capabilities per organization is lost. Although the average scores do not necessarily give a false image of the actual Big Data Readiness of the organizations capabilities, it is valuable to look deeper into the individual scores per capability. What is especially interesting is the amount of capabilities that have been scored very low in the questionnaire, as these scores indicate situations in organizations in the Dutch public executive sector that severely reduce the Big Data Readiness of those organizations. By investigating the frequency of occurrence of the individual scores per capability in score classes, the amount of very low scoring organizational capabilities can be found. The results are presented in Figure 26.

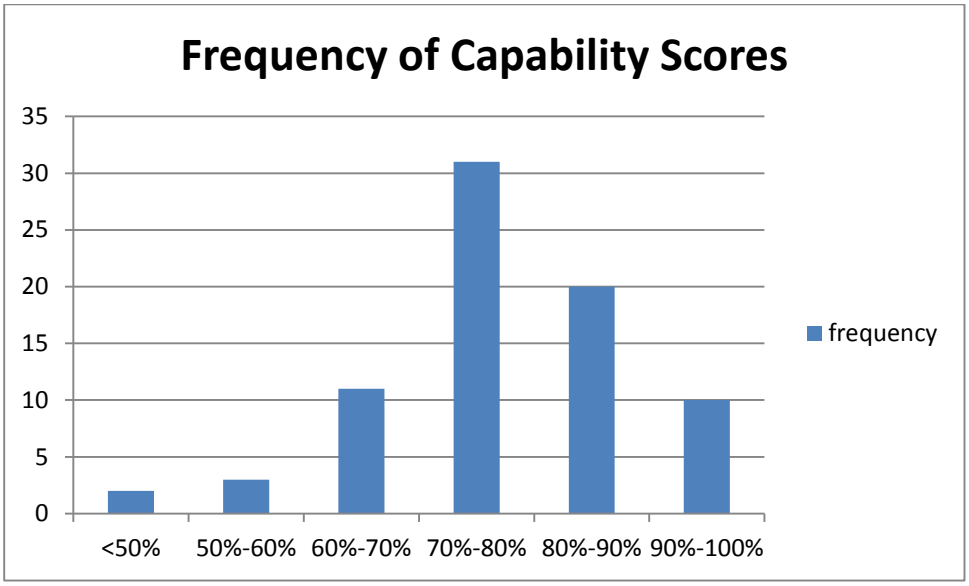


Figure 26: Frequency of Capability Scores

It is clear the amount of capabilities that score very low is quite limited. This indicates that the finding that organizations in the Dutch public executive sector are indeed reasonable capable to start implementing and using Big Data is defensible. However, one further test should be executed, before this finding can be put forward with certainty.

Low scores on only one organizational capability for using Big Data in an organization can indicate that that particular organization is not ready to start using Big Data. So, although a limited amount of very low scoring capabilities have been found in Figure 26, they could still severely hamper the introduction of Big Data use in the Dutch public executive sector if they are distributed in such a way that virtually all organizations have one or two very low scoring organizational capabilities, indicating a situation in which none of the organizations is ready for the introduction of Big Data use. To investigate whether this is the case of not, the frequency of very low capability scores for each separate organization has to be made visible. This is done in Figure 27 on page 89.

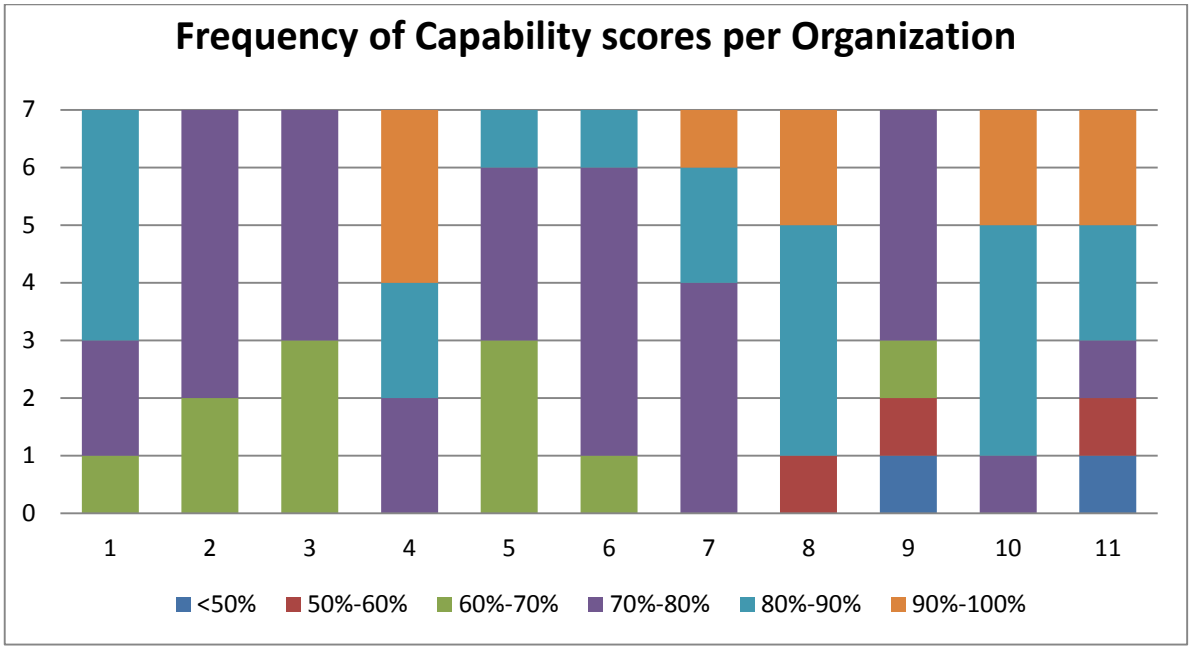


Figure 27: Frequency of Capability scores per Organization

From Figure 27 it becomes apparent that only three out of eleven organizations currently have very low scoring organizational capabilities, indicating that problematic capabilities hindering the introduction of Big Data are not spread over all organizations, but are more concentrated in a couple of organizations. Also important to note from this figure is that organization 8, which has one low scoring organizational capability scores the Organizational Capabilities Assessment Value High, as all of its other capabilities individually score high and very high. This shows that although the average score over all capabilities per organization gives a representable view of the overall Big Data Readiness of organizations in Dutch public executive sector, attention to more detailed scores is also important for accurate assessment of Big Data Readiness per organization. Very few organizations in the Dutch public sector have weak points in their capabilities to implement and use Big Data in their organizations. However, public officials should be careful with the implementation of Big Data use in their organizations, as early mistakes could swing public opinion on Big Data use and block future development before the added value of Big Data can be proven.

Lastly, the connection between the organizational capabilities assessment values and the organization types is investigated. The average assessment values per organization type are therefore presented in Table 18. Just as with the organizational alignment, organizations with intensive data use activities score higher on their Big Data Readiness. These organizations seem to have better developed capabilities for Big Data use and its introduction through their extensive experience from their current intensive data use activities. Again, the connection seems fairly obvious, but could have important implications in the future development of Big Data use in the Dutch public executive sector.

Main statutory task type	Average Capabilities Assessment Value
Coordination, project-based task, no data use	Low
Research, Evaluation	High
Registration, Documentation	Low- Medium
Administration, Management	Medium - High

Table 18: Organizational Capabilities Assessment per main statutory task type

6.1.4 Big Data Readiness

Combining the scores of all three aspects of the Big Data Readiness framework allows us to assess the complete Big Data Readiness for each of the participating organizations from the Dutch public executive sector. The values on all three aspects are combined in Table 19. By coding all scores to a range of 1 to 5 points for Very Low to Very High values, a summed Big Data Readiness score can be established, which can range from 3 to 15 points. Each of the possible Big Data Readiness scores has been given a Big Data Readiness assessment value. The details of the decision rules of the assessment method are presented in Appendix IV on page 140.

#	Organizational Alignment	Organizational Maturity	Organizational Capabilities	BDR score	Big Data Readiness Assessment
1	Very High	Low	Medium	10	Ready for further development of Big Data Use.
2	Medium	Medium	Medium	9	Ready for planning and early design of Big Data Use
3	Low	Very Low	Low	5	Ready for orientation on Big Data Use
4	High	Low	High	10	Ready for further development of Big Data Use.
5	Very Low	Very High	Low	8	Ready for planning and early design of Big Data Use
6	Medium	Low	Medium	8	Ready for planning and early design of Big Data Use
7	High	Low	High	10	Ready for further development of Big Data Use.

8	Medium	Very Low	High	8	Ready for planning and early design of Big Data Use
9	Very Low	Very Low	Low	4	Ready for orientation on Big Data Use
10	Medium	Very Low	High	8	Ready for planning and early design of Big Data Use
11	Medium	Low	Medium	8	Ready for planning and early design of Big Data Use

Table 19: Big Data Readiness Assessment

The Big Data Readiness assessment of the eleven organizations presented in Table 19 sketch an image that organizations in the Dutch public executive sector are not yet ready to start using Big Data in their organization. Although some organizations score quite high, the overall picture is that even these organization have substantial further orientation, planning and development to do, before Big Data can be successfully used in their organizations. None of the assessed organizations is given consistently high assessment values over all three of the Big Data Readiness Framework aspects and is coming close to the level of Big Data readiness that is required for using Big Data successfully.

When the Big Data Readiness assessment values are averaged on the four organization types identified in section 5.1, it becomes clear that current intensive execution of both data collection and data use activities seem to have a positive influence on the Big Data Readiness Assessment of organizations in the Dutch public sector. The BDR score presented in Table 20 furthermore indicates that the positive effects of experience in both data activities enhance each other and results in an even higher Big Data Readiness Score for organizations that are experienced in both data activities.

Org. Type	Organizational Alignment	Organizational Maturity	Organizational Capabilities	BDR score	Big Data Readiness Assessment
1	Very Low - Low	Very Low	Low	4,5	Ready for orientation on Big Data Use
2	Medium – High	Very Low - Low	High	8,5	Ready for planning and early design of Big Data Use
3	Low	Medium - High	Low- Medium	8	Ready for planning and early design of Big Data Use
4	High	Low- Medium	Medium - High	9,67	Ready for further development of Big Data Use.

Table 20: Big Data Readiness per main statutory task type

From the overall Big Data Readiness assessment another important observation can be made. The average organizational capabilities of Dutch public executive organizations are quite well developed, which could lead to organizations believing they are ready to start using Big Data. On itself this is true, as they have developed the capabilities needed to use and implement Big Data and the required technologies. However, these organizations should not forget that the scope of a radical new concept such as Big Data is much wider than just the organizational capabilities required for it. Big Data applications are so comprehensive and potentially invasive for organizations that the organizational alignment of these applications is equally important for the success of Big Data use. Organizations may be capable of using Big Data in their organizations, if they have to undertake data activities that do not fit their current organizations and main statutory tasks. They will not significantly gain from these Big Data applications, as the applications will struggle to provide added value in line with the main activities of the organization. Organizational alignment of Big Data applications is vital for its success and should not be neglected, even if organizational capabilities are very high. Furthermore, organizational maturity is of organizations in the Dutch public sector is poor, resulting in limited exchange and sharing of information and data between organization, which in turn limits the amount and especially variety of data available as input for the Big Data applications. Not only limits this the potential added value of these applications, it also severely limits the ability of Dutch public organizations to develop the granular, individual information and insights needed for them to develop towards a more demand driven, joined-up form of government.

Organizations in the public sector contemplating the implementation of Big Data use in their organizations may focus on their organizational capabilities for it, but should also strongly consider the other aspects of Big Data Readiness presented in this research project. They should develop their understanding of the connection between Big Data applications and their organization and their organizational maturity by structurally sharing and exchanging information and data with other organizations.

6.2 Areas of Improvement

After the Big Data Readiness assessment of the organizations in the Dutch public executive sector is completed, the identification of possible areas of improvement of that Big Data Readiness is presented in this chapter. The following sections will describe a number of the most promising possible areas of improvement of the Big Data Readiness of Dutch public executive organizations. The areas of improvement will be presented in regard to their corresponding part of the Big Data Readiness Framework. First the possible improvements in organizational alignment are presented, followed by the areas of improvement for the organizational maturity and lastly the possible areas of improvement for the organizational capabilities.

6.2.1 Areas of Improvement for Organizational Alignment

Improvements in Big Data Readiness on organizational alignment can be made in two separate areas. Both are part of the IT side of the Strategic Alignment Model described in section 5.1 on page 58. These two areas are expressed in the Big Data Readiness Framework as the planned Big Data application type and the big data use characteristics present in the current data activities of the organization. Both can be improved to make the alignment between the organization and the

planned Big Data application better, which will result in an improved Big Data Readiness of the organization. First the selection of the optimal Big Data Application is discussed, followed by the development of the required Big Data Use characteristics.

Big Data Application types

As was presented in section 6.1.1 on page 77, the organizational alignment of organizations in the Dutch public sector is assessed as Medium and can be improved significantly. The most important improvement organizations can make is to increase their understanding of the different types of Big Data applications and their unique requirements and characteristics. They should also recognize that certain application types fit best with certain main statutory tasks and corresponding current data activities and with certain Big Data use characteristics present in the current data activities of organizations. Dutch public organizations should better understand that not all Big Data applications fit their organization and current activities. Big Data applications of the wrong type will require organizations to employ intensive data activities that they would normally not employ for their main activities and this could severely decrease the overall benefit of Big Data applications.

In practice this means that public organizations should first plan to develop Big Data applications of a type that best fits their current organization's statutory task and current data activities. From the questionnaire it became apparent that multiple organizations in the Dutch public executive sector are planning Big Data applications of multiple application types, of which some do not optimally align with their organization. Some organizations indicated advanced interest in Big Data applications, while their current organization is totally unaligned with it. The Big Data Readiness of organizations with plans that are simply too ambitious would improve if they realized they have a much higher chance of success with the use of Big Data in their organization if they start with it in an application best suited to their current organization.

Big Data Use Characteristics

The second area of improvement for the organizational alignment part of Big Data Readiness are the present Big Data Use characteristics in the current data activities of the organizations. By further developing the possibilities and expertise to perform data activities with these characteristics within their current data activities, the organizations will improve their alignment with future planned Big Data applications as their current infrastructure will already be better geared towards the introduction of the planned Big Data applications. By developing their current data use incrementally towards the future data activities once Big Data use is implemented, organizations can enhance their experience and organizational infrastructure towards the state needed for Big Data use and therefore improve their organizational readiness for it. Practitioners should let their organizations experiment with for example real-time or unstructured data, to prepare their infrastructure and employees for the transition towards Big Data use in the future.

In Figure 28 the frequency of possible improvements on the five Big Data use characteristics over all assessed organizations is presented. From the graph it becomes clear that the most possibilities for improvement in this area can be found in the Big Data use characteristics using a combination of structured and unstructured data and in the innovative use of data in the current data activities of organizations in the Dutch public executive sector. However, these areas of improvement are not

very specific, so the possible improvements are also plotted per identified organization type in Figure 29 on page 95.

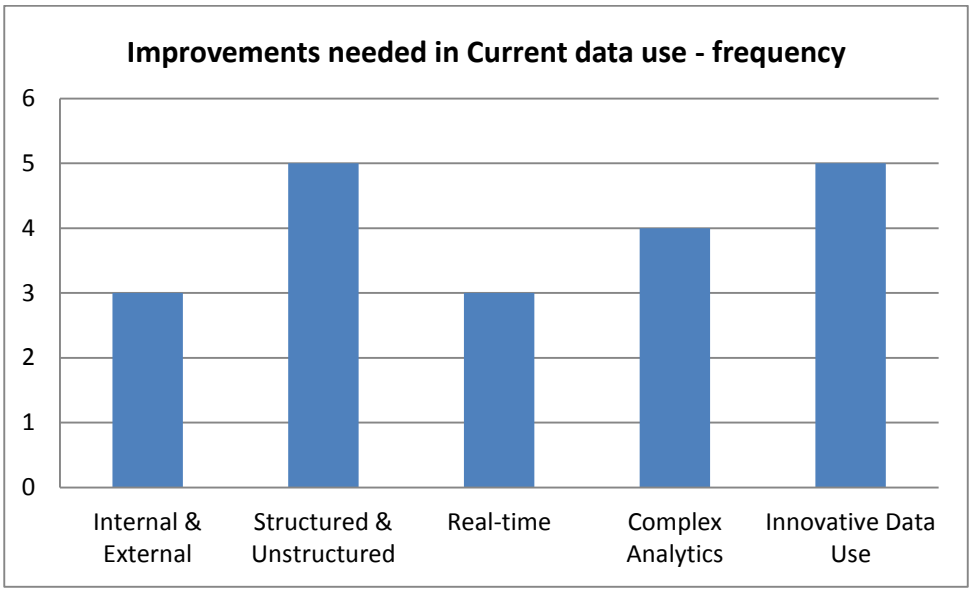


Figure 28: Improvements in Big Data Use Characteristics

Figure 29 gives a much more specified image of the possible areas of improvement in organizational alignment for organizations in the Dutch public executive sector. Two clear opportunities for improvements stand out. The first important specific area of improvement that can be identified from the figure is that organizations with main statutory tasks related to research and evaluation can best focus on improving their possibilities to use a combination of structured and unstructured data in their data activities when they strive to improve their organizational alignment for research type Big Data applications. These organizations have a limited amount of internal data, but are advanced in data use and analysis. By extending their possibilities in analysis by improving their use of unstructured data, these organizations can gain significant advancement in their readiness to use Big Data in their organizations and the potential value it can bring them.

Secondly, organizations with main statutory tasks related to registration and documentation can best improve their organizational alignment with object/subject evaluation type Big Data applications by developing their current data activities by making more innovative use of their vast stores of data. These organizations have large collections of internal data and can significantly improve their Big Data Readiness and its potential value if they learn how to use that large amount of internal data more innovatively. This means they should develop new and creative ways to use their internal data, to add more value to their organizations from it.

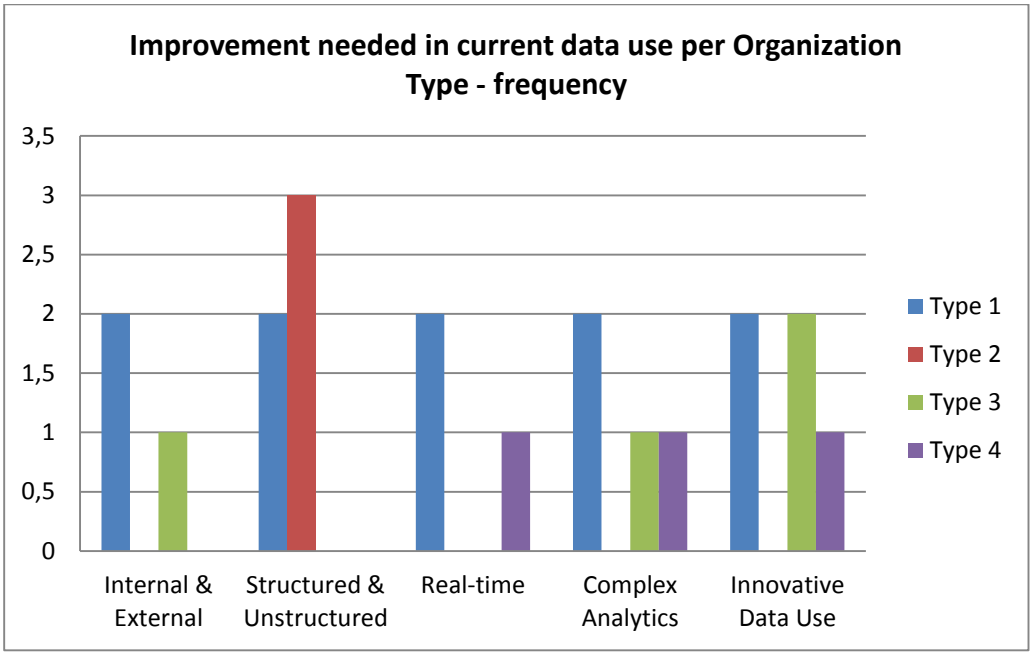


Figure 29:Improvements in Big Data Use Characteristics per Organization Type

6.2.2 Areas of Improvement for Organizational Maturity

The possible areas of improvement for the organizational maturity of organizations in the Dutch public executive sector are presented in this section. The most promising areas of improvement are again identified from the frequency of possible improvements of all used factors over all organizations assessed in the research project. The possible improvements are plotted in Figure 30 and divided in blue for the Maturity Level of the improvement and red for the three aspects used for the assessment of the growth stages of the organizations for Organizational Maturity assessment.

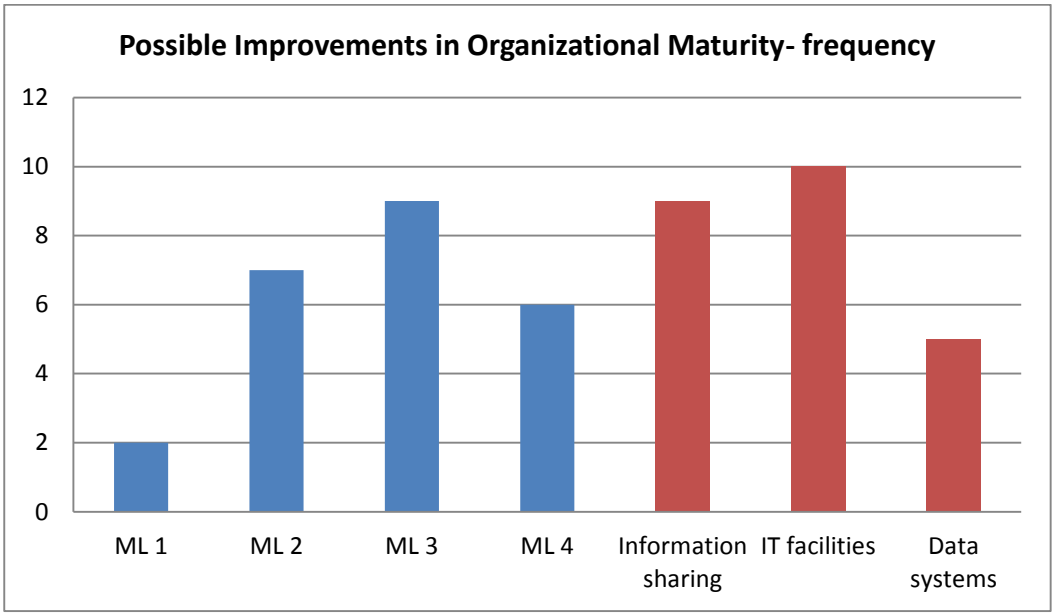


Figure 30: Possible Improvements in Organizational Maturity

Again, the results from the general figure are not very specific and no clear promising area of improvement can easily be identified from the figure. The only real information that can be derived from the figure is that no emphasis is needed on the first maturity level and that currently used data systems also do not need extensive attention, but that is not very helpful. To create more insight into the most promising areas of improvement, the possible improvements are again plotted per organization type. The result is presented in Figure 31.

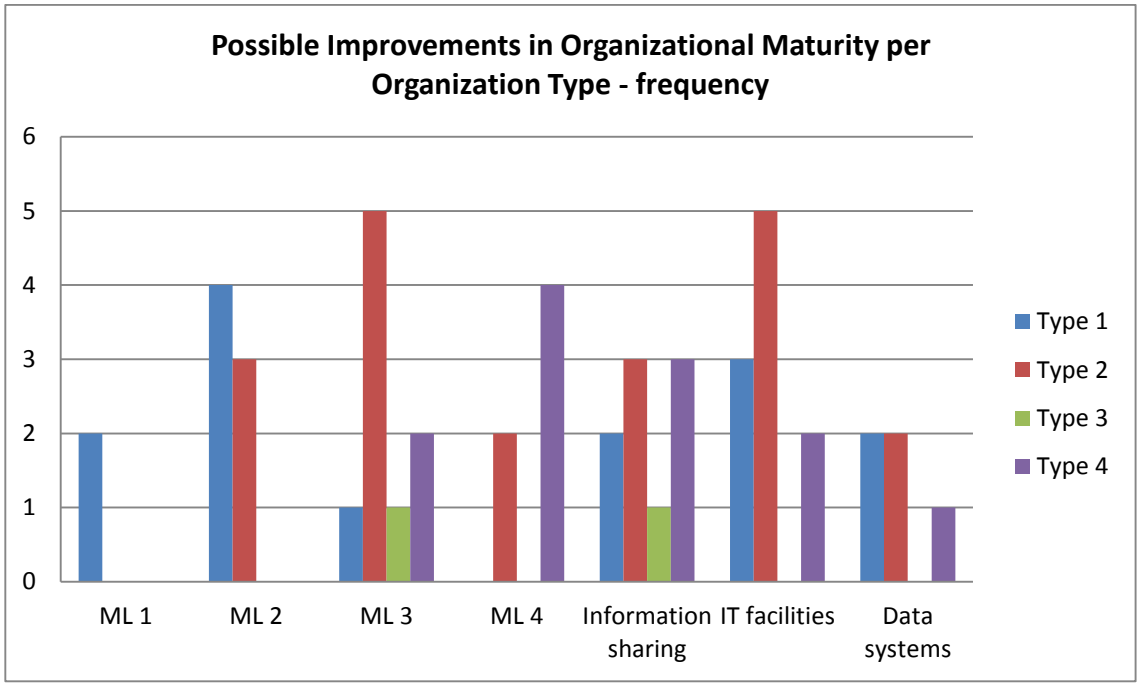


Figure 31:Possible Improvements in Organizational Maturity per Organization Type

From Figure 31 the following four promising areas of improvement of organizational maturity for the use of Big Data of organizations in the Dutch public executive sector can be derived. Organizations with coordinating or project based statutory tasks or without any intensive data activities can improve their organizational maturity for Big Data best by improving the availability and sharing of data and IT facilities within their organizations. Organizations with research type statutory tasks can best improve their organizational maturity for Big Data by enabling their IT facilities and data to be more accessible for outside stakeholders. Organizations with registration and documentation type of statutory tasks can best improve by redesigning their current activities to be more accessible and available for outside stakeholders and share more information with them. Administration and management type of tasked organizations can improve their organizational maturity for Big Data use best by increasing their efforts to actively share and cooperate on activities and information creation and provision with other organizations.

In general organizations in the Dutch public executive sector should increase their sharing and exchanging of data and information with other organizations to improve the amount and especially variety of data to use as input for their future Big Data activities. Only in this way will their future Big

Data applications be able to provide the individual and granular information that is needed for more demand driven, joined-up forms of government.

6.2.3 Areas of Improvement for Organizational Capabilities

The last areas of improvement to be presented in this chapter are the areas which are focused on the organizational capabilities important for Big Data use in the organizations in the Dutch public executive sector. The organizational capabilities for Big Data use are scored in percentages and the three lowest scoring capabilities of each of the assessed organizations is used to compile the possibilities of improvement in Figure 32.

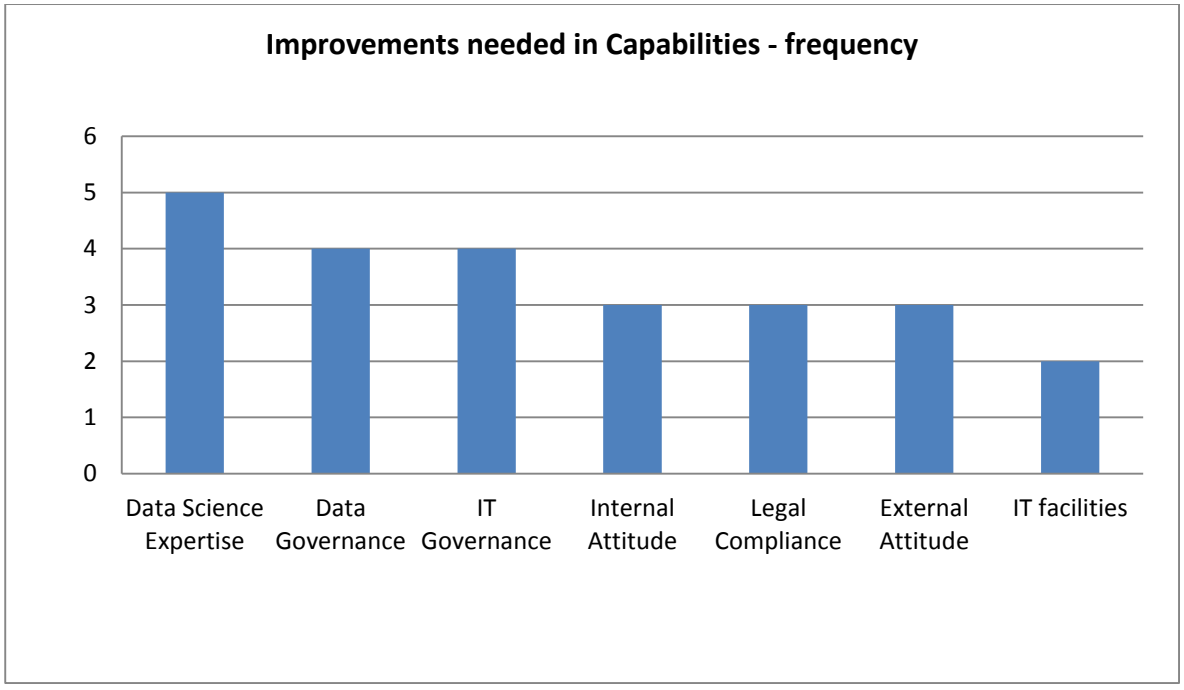


Figure 32:Possible Improvements in Organizational Capabilities

From the ranking in Figure 32 it becomes clear that the organizations in the Dutch public executive sector can best improve their Big Data Readiness by improving their organizational capabilities to introduce and handle Big Data in their organizations on the capability to develop and/or bundle Data Science Expertise, the capability to design and execute a solid Data Governance program and the capability to design and execute a suited IT Governance program.

The ranking of the frequency of possible improvements of the organizational capabilities in Figure 32 is useful and does create some insight into possible promising areas of improvement for the Big Data Readiness of organizations in the Dutch public executive sector. However, the identification of the possible areas of improvement needs to be more specific to be useful, so again the possible improvements are plotted per organization type. The result is presented in Figure 33 on page 98.

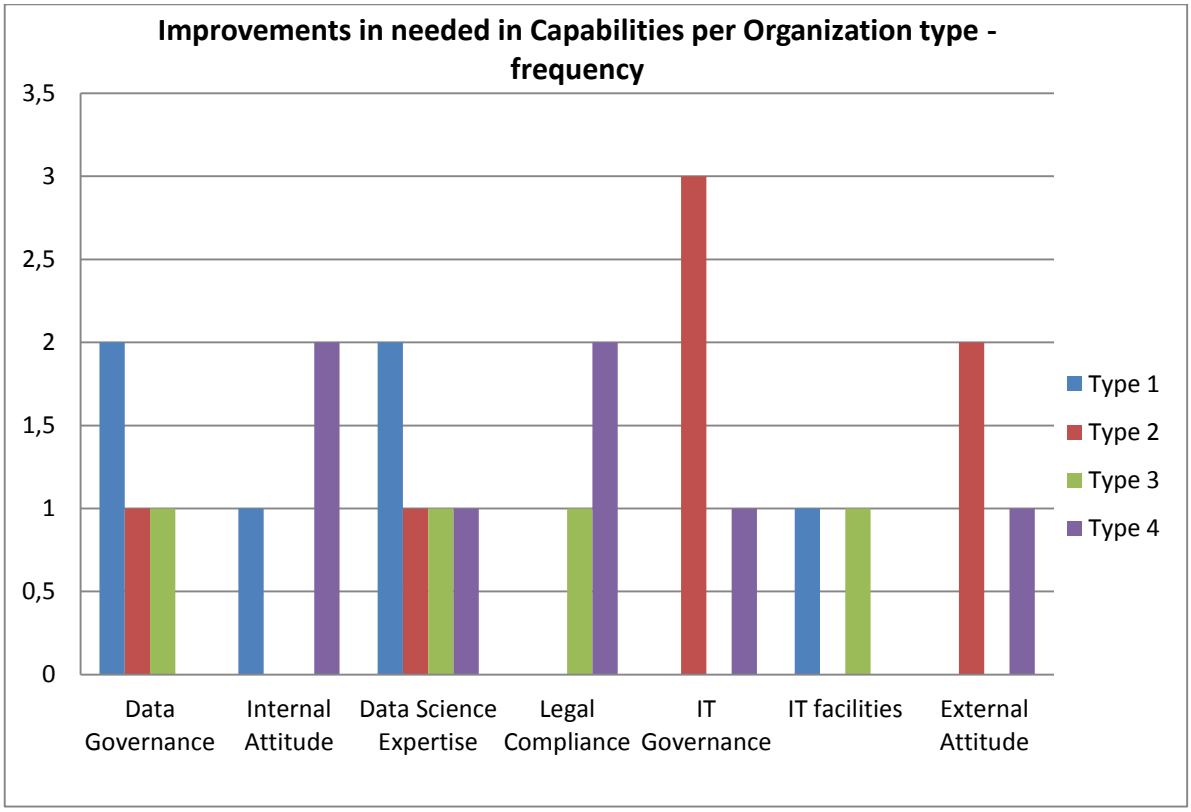


Figure 33: Possible Improvements in Organizational Capabilities per Organization Type

Similar to the areas of improvement for the organizational maturity for Big Data use presented in the previous section, the most promising areas of improvement are identified and presented per organization type.

Organizations with coordinating or project based statutory tasks or without any intensive data activities can best focus on improving their Data Governance and Data Science Expertise capabilities. This area of improvement is a logical consequence of the currently missing experience of working with data intensively in these organizations. However, they do need to develop their capabilities in this area significantly before they will be able to introduce the use of Big Data in their organizations.

Organizations with research type statutory tasks should primarily focus on improving their IT Governance capabilities and their capability to improve the external attitude towards the use of Big Data at these organizations. Improving IT Governance capabilities is necessary as especially these organization have to significantly alter the nature of their current IT systems towards the use of Big Data, as their current IT systems are more geared towards supporting project based research applications and not data intensive, more continuous Big Data applications. Improving the capability to improve the external attitude will be vital for organizations of this type, as these organizations do not intensively collect data themselves, but are dependent on other organizations to share or sell the required datasets.

Organizations with registration and documentation type of statutory tasks do not seem to have very specific capabilities to develop in their organization type, but can focus on capabilities for Data

Governance, Data Science Expertise, Legal Compliance and IT facilities. These capabilities should especially be developed in the light of the intensifying of the actual use of all of the data this type organization has stored. In these organizations most of the organizational capabilities important for Big Data are developed quite well. However, they are developed primarily for the collection and storage of data and not for the use of data within the organization. It is therefore important for these organizations to develop and adapt their organizational capabilities more towards data use, as this will be vital for using Big Data in these organizations in the future.

Organizations with administration and management type of statutory tasks will primarily have to focus on developing their capability to improve the internal attitude towards Big Data and the capability to design and execute a successful legal compliance program for their data activities. The capability to improve their internal attitude towards Big Data is important for a successful introduction of Big Data use in their organizations as that introduction will represent a significant change throughout these organizations. Furthermore, changing towards a more data oriented culture will have the most effect in these organizations, as both data collection and data use in intensively executed in them. Developing their capability for the design and execution of solid legal compliance programs will be important for the use of Big Data in these organizations, as they control and execute the whole Big Data value chain within their organization. These organizations should therefore have very solid and comprehensive legal compliance programs.

Overall, organizations in the Dutch public sector have quite well developed organizational capabilities for the implementation and use of Big Data is their organizations. There are areas for improvement, and they are different for each organization and organization type. In general, Dutch public executive organizations can improve their organizational capability for Big Data use by improving their expertise on data science, improve their handling of data and their IT strategy. Big Data applications need all of these capabilities to be on a very high level to be successful.

6.3 Evaluation of the Big Data Readiness Framework

Now that the framework has been used to create an overview of the Big Data Readiness of organizations in the Dutch public executive sector, the performance of the proposed framework can be examined and evaluated. Extensive evaluation of the framework is very hard to execute, for several reasons. Firstly, the proposed Big Data Readiness Framework is designed to solve a research and practical problem that organizations in the public sector are not able to assess their organizational readiness to start using Big Data. A problem for which no previous solution was available. The Big Data Readiness Framework is unique in this sense and can therefore not be fully evaluated by comparing its performance to other specifically designed solutions for this problem. Secondly, the amount of data used in this research project is not sufficient to thoroughly test the framework and all of its aspects in great detail. Because of the limited research objective to establish an overview of the Big Data Readiness of organizations in the Dutch public executive sector, only eleven representatives of organizations participated in the research project and that number is insufficient to evaluate the performance of the framework in great detail.

However, some brief evaluation of the framework with the use of validation and verification is possible and will be presented in this section.

Validation

Firstly, the full framework is validated on whether it performs all of its desired functions and if it fulfills its set goals. The main objective of the framework is to create an overview of organizations Big Data Readiness for the organizations of the Dutch public executive sector. For this overview the framework has to address the three main uncertainties that public organizations in the Netherlands experience in their own assessment on their organizational readiness for Big Data. The main uncertainties are alignment between organization and Big Data applications, maturity of the organizations to change for Big Data and the organizational capabilities to introduce and handle Big Data in the organization.

It can be concluded that the framework is validated in regard to its main objective. The framework provides a clear overview of organizational readiness for Big Data from the questionnaire and addresses all three of the main uncertainties of public organizations. It can be concluded that the framework functions appropriately in regard to its main objective and can therefore be successfully validated.

The second objective of the framework is to demystify the concept of Big Data and to connect the ambiguous term to established and more grounded scientific knowledge. If the framework can also accomplish this objective, the term Big Data will become more understandable and useable for both practitioners and academia. By first explaining the concept of Big Data in light of its actual use in public organizations by describing differentiating characteristics and the process and activities of using Big Data, the framework is able to provide the concept with better handles and makes it more understandable and specific. This allows for the Big Data Readiness framework to be able to connect the concept of Big Data to established scientific knowledge, such as theory on organizational alignment and change for IT implementation, e-government development and organizational core and dynamic capabilities.

As the Big Data Readiness framework was indeed able to connect frameworks based on these theories to the concept of Big Data and in the same process create more understandable handles on the concept of Big Data that were well understood by the practitioners that participated in the questionnaire, it can be concluded that the framework has also delivered on its second objective. The Big Data Readiness framework can therefore be validated for the purposes it has for this research project.

Verification

The verification of the Big Data Readiness framework is unfortunately much harder to do. As stated earlier in this section, no alternative way to establish the readiness for Big Data use introduction of public organizations, are similar concepts is available. This means that the results of the analysis that is made using the Big Data Readiness Framework cannot be compared with results using other methods and thus cannot be verified in this way. Verifying each step of analysis taken in the framework would be an alternative way of verification of the Big Data Readiness framework, but this would entail a massive amount of research, that if possible at all, unfortunately falls outside the scope of this research project. It also remains questionable if this verification process would produce meaningful results with the limited number of organizations assessed in this research project and the limited timeframe used.

However, one method of verification can be applied to the results of the analysis using the Big Data Readiness framework, albeit it is not a very accurate method of verification. Each of the organizations was assessed on the basis of answers to a questionnaire by a public official from that organization. Prior to that, these representatives were interviewed in research step 1, to establish the current situation regarding Big Data in the Dutch public executive sector and to gain knowledge from practice on Big Data to help describe the use of Big Data. From these interviews a general estimation of the organizational readiness for Big Data of each of the assessed organizations can be formulated. This has been done in Table 21, where all eleven organizations are given a classification based on the information given in the interviews of being well-developed, somewhat developed or not very developed towards Big Data. By comparing these estimations with the Big Data Readiness scores produced by the analysis based on the Big Data Readiness framework, the frameworks results can be examined on their accurateness in comparison to the image gathered from the interviews.

Organizations	Estimations after interviews	BDR Score	Comparability?
1	Well-developing towards Big Data	10	Yes
2	Somewhat developed towards Big Data	9	Yes
3	Not very developed towards Big Data	5	Yes
4	Well-developing towards Big Data	10	Yes
5	Not very developed towards Big Data	8	No
6	Somewhat developed towards Big Data	8	Yes
7	Well-developing towards Big Data	10	Yes
8	Somewhat developed towards Big Data	8	Yes
9	Not very developed towards Big Data	4	Yes
10	Well-developing towards Big Data	8	No
11	Somewhat developed towards Big Data	8	Yes

Table 21: Verification of Big Data Readiness Assessments

From Table 21 it can be concluded that the results from the Big Data Readiness framework are very comparable to the estimations made from the interviews. Of course, this is not very surprising, as the framework was designed on the basis of the information gathered in these interviews. However, is does reflect the fact that the framework is able to produce results that are very comparable to the intuitive opinion of the interviewer. The framework however shows the image of Big Data Readiness with much greater detail, consistency and transparency and that is its main added value.

The fact that two organizations do not have a comparable estimation and Big Data Readiness assessment can be explained by deviations in their assessments on organizational maturity. Organization 5 scores much higher than estimated and organization 10 much lower. It is possible that these differences come from incorrect estimations during the interviews, or from different

interpretations of the questions on organizational maturity in the questionnaire. However, it could also be the case that the organizational maturity aspect of the Big Data Readiness framework is not performing optimally and that it should be improved to make the results of the framework more accurately. The limited number of assessed organizations makes it impossible to state which of the above explanations is most accurate, but further research on this aspect may be helpful to improve the accuracy of the assessments from the framework further.

For the purpose of this brief evaluation of the Big Data Readiness framework, it can be concluded that the relatively high comparability between the estimations from the interviews and the assessments form the framework shows that the framework can be tentatively verified to accurately assess the Big Data Readiness of public organizations.

7. Discussion

In this section the findings from the research project, their meaning and their implications are discussed. In the discussion the findings from the analysis using the Big Data Readiness framework are translated into the conclusions and recommendations of the research project.

7.1 The Assessments

The Big Data Readiness framework is created for two purposes. One, it is an analysis tool that can provide more structured and detailed insight into the current readiness for Big Data of organizations in the public sector and the possible areas of improvement. Two, the framework demystifies the concept of Big Data and its implications for public sector organizations and improves its understandability for both practitioners and academia by connecting the concept to more established scientific concepts. As can be concluded from the evaluation of the framework in the previous section, the framework seems to deliver on both purposes. But the real value of the Big Data Readiness framework comes forward when the demystification of Big Data and the results on Big Data Readiness of the Dutch public executive sector are combined to give meaning to the Big Data Readiness scores that are presented in chapter six.

The first aspect of the Big Data Readiness framework is the organizational alignment between the organizations current statutory task, intensity of performed data activities, Big Data characteristics and its planned Big Data application. A low organizational alignment means that organizations plan to introduce Big Data applications that are not compatible with their current activities. The incompatibility causes the introduction of those Big Data applications to be much harder and much more radical changes are needed for it to be successfully implemented. A low organizational alignment score often indicates that an organization has insufficient understanding of the implications certain types of Big Data applications have on the organization. Of course, a high organizational alignment assessment indicates the opposite, it shows that organizations have o good understanding of what type of Big Data application would best fit their organization and current activities. In the Dutch public sector only a few organizations score low on the organizational alignment assessment, indication that most organizations in the Dutch public sector have a reasonable understanding of what Big Data use in different application types would mean for their organizations. Those organizations that do score low in this respect are often organizations who are currently not performing data use activities very intensively, have less experienced in that regard and therefore have a lower understanding of the implications of the introduction of Big Data applications in their organizations. This results in that these organizations are definitely not ready to start with using Big Data in their organizations. Overall, organizations in the Dutch public sector do not understand well enough that Big Data applications will only add value to their organizations if they are supported by the main activities of the organizations. Organizations more experienced in using data better understand the implication Big Data application will have for organizations are closer to its implementation than organizations who do not intensively use a lot of data.

The second aspect of Big Data Readiness is organizational maturity, which indicates the current stage of e-government development the organization is at. The e-government stage of an organization is an indication to what extent that organization is advanced in working with IT systems to support its current activities, communication and service provision to citizens and other organizations and its cooperation with other public organizations in those areas. A low organizational maturity indicates that organizations are not very far in their development towards intensive cooperation and sharing

of information in the organization and with other organizations. In regard to Big Data this means that organizations with a low organizational maturity are not very advanced in sharing data, information and knowledge with other organizations. In terms of e-government development, a low organizational maturity entails that organizations still have a long way to go, before they are able to become more demand-driven and flexible in their service provision and activities. Big Data technologies could be an important enabler of such developments, but also has to be supported by shared data, information and knowledge, before it can reach its full potential value. Implementing Big Data technologies in organizations with a low organizational maturity will mean that the technologies added value is severely limited, as they will rather form very advanced analytics applications than actual Big Data applications. Without the input of huge amounts of varied data from various sources, Big Data applications will never be able to provide public organizations with the value and benefits they are promised to provide. The positively enforcing interaction between Big Data technology and organizational maturity is one of the main reasons of the high potential value of these technologies for the public sector. The interaction however does need a starting point sufficiently developed to lend Big Data applications with enough data and possibilities to add enough value to enable public organizations to develop further. In this regard the current situation in the Netherlands is not good. Across the Dutch public executive sector organizations are assessed with a low organizational maturity as only very few organizations are structurally and intensively cooperating with other organizations in regard to continuous data and information sharing. The few organizations that do receive a high assessment of organizational maturity do so, as they are intensively sharing their data with other organizations, both in the public and in the private sector. Dutch public organizations do not share and exchange enough data to have a large and varied amount of it as input for their future Big Data applications. Without that input, producing the granular and individual information needed for demand driven government forms will be much harder to do.

The third and last aspect of Big Data Readiness is organizational capabilities. The seven capabilities taken from both literature and the interviews with practitioners provide an overview of the attention, development and presence of different aspects of the ability of public organizations to implement, use and control Big Data technologies in their own organization. Of course, this comprehensive ability contains much more than seven capabilities for organizations to develop, but the seven used in the organizational capabilities part of the framework are seven of the most important capabilities. A low score on organizational capabilities indicates that organizations do not have the required full package of sufficiently developed capabilities for Big Data use in their organization. This means that organizations with a low score have one of three problems, or a combination of problems. Either organizations are not able to successfully implement Big Data technologies. This will most likely be the case with low scores for the capabilities Internal and External Attitude and IT Governance. Low scores for Legal Compliance and Data Governance could indicate problems with controlling Big Data use in the organization. And problems with the effective use of Big Data will most likely occur when the capabilities Data Science Expertise and IT facilities score low.

However, the biggest problem with organizational capabilities is not an overall low score, but the single capabilities that are not sufficiently developed in organizations. In the Dutch public executive sector only a few organizations score low overall on their organizational capabilities, but the majority of assessed organizations scores higher. So, at first glance, it seems that the Dutch public executive sector is quite well-developed in regard to organizational capabilities for Big Data use. However, only a very few of the assessed organizations scored high on all capabilities, indicating that the vast

majority of organizations has at least one capability important for Big Data that is not at a sufficient or near sufficient level to start with using Big Data. Data Governance and Data Science Expertise seem to be the two capabilities causing these kind of problems, but the connection is not very clear. The danger however with organizations that are only lacking in one organizational capability is that they may assess themselves, incorrectly, to be ready for the introduction of Big Data. And this increases the risk of failing to take advantage of Big Data.

Another risk for Dutch public executive organizations is that they overestimate their readiness for Big Data use because of their well-developed capabilities to implement and handle it. Big Data is not just another IT resource you can tap into, it has a much larger scope and much larger implications. Not only the right capabilities are enough to use and handle it effectively. Public organizations should be aware that a good match between the present main activities and future Big Data applications is just as important as being able to handle the data itself. Without this match Big Data applications will not be able to deliver information that is really valuable to the organization. Secondly, the unmatched Big Data applications may require additional activities to be performed outside of the scope of the main activities of the organization, increasing cost and risk, and thereby decreasing its overall benefit significantly. Also, Dutch public executive organizations should not overlook the fact that currently they lack a level of structural information and data sharing between organizations that is needed to provide their future Big Data application with the amount and variety of data it needs. Without it Big Data applications will never be able to produce the individual and granular information needed for real demand driven government forms.

Big Data is an unique opportunity for the public sector to structurally change and improve government organizations and it should not be wasted. The decision support information, the new insights and the richer images of reality offered by the three Big Data application type described in this thesis report grant public organizations the opportunity to significantly improve their effectiveness, efficiency and transparency and progress on the e-government maturity ladder. However, as the findings on Big Data Readiness in the previous chapter show, much work still has to be done before that potential can be unlocked.

Public organizations should not start using Big Data before they are ready, as this will most likely lead to severe problems with the darker sides of Big Data. Breaches in privacy and security of personal data, unfair treatment of citizens through over-extensive and unethical datafication of decision making processes, wrong or sub-optimal decisions because of incorrect handling, analyses and interpretation of data, severely incompetent and faulty IT facilities and extreme investments in IT that never pay off are just some of the examples of the dangers of Big Data. These dangers are real and pose a substantial threat to the promise of Big Data and its potential value for society. And they can only be avoided with careful planning and development of adequate organizational alignment, maturity and capabilities to start using Big Data. Only if organizations are able to fulfil all of these prerequisites should they consider to start using Big Data in their organizations.

7.2 Possible Improvements

So, it is clear the Dutch public executive sector is not ready for the introduction of Big Data use. Organizations do not always understand which Big Data applications align best with their organizations, are not mature enough to take full advantage of the potential of Big Data and do not always have the required capabilities to implement, use and control Big Data in their organizations. Organizations in the Dutch public sector should therefore refrain from premature introduction of

Big Data use in their organizations. They should wait for the right moment for that introduction. And in the time in between they should develop their organizations towards the requirements of Big Data use as much as they can. However the questions remains where these developments should take place and what organizations in the Dutch public sector can do to improve their organizational readiness for Big Data. With the results from the Big Data Readiness assessment in section 6.2 on page 92 an answer to that question will be provided in this section.

Currently, organizations in the Dutch public executive sector should continue with the development activities for Big Data. They should continue to learn and orientate themselves on what Big Data use will entail for their organization and what specific added value Big Data can bring to them. Public organizations should establish specific plans or roadmaps that will guide their development. By making sure the identified weak points of the Big Data Readiness of their organizations are addressed in these plans, the organizations in the Dutch public sector can swiftly develop their understanding of Big Data alignment, organizational maturity and organizational capabilities. This will allow them to be able to introduce Big Data use and its added value in their organizations in the near future and start their further development on the e-government maturity ladder and provide more value for society.

Besides the development of the individual weak points in organizations, more can be done to bring the introduction of Big Data use in the public sector closer. Public sector organizations should have more attention for the recruitment and training of so-called Big Data talent or data scientists. By employing more experts in the various profession concerning data, the in-house expertise for data intensive activities can be significantly improved, which will both benefit current data activities, Big Data Readiness of the organizations and the successful utilization of future Big Data applications. The acquisition of these employees has to start as soon as possible, as the demand for them from organizations across the board is increasing quickly.

Furthermore it should be stressed that public organizations developing Big Data should not do so individually and solely in the confines of their own organizations. From the findings on possible areas of improvement it is apparent that the organizations in the Dutch public executive sector have different strong and weak points in their organizational readiness for Big Data. Therefore, organizations can potentially learn from each other and develop faster and more efficient that way. Organizations should look for interesting partners in their development of Big Data. Not only in their own field of expertise, organization type or even organizations under the same Ministry, but specifically organizations that are strong and well-developed on points important for Big Data use that they are not. Also, organizations in the public sector should not solely focus on other public organizations for their Big Data development partners. A large number of both small and large private organizations are currently also working on developing for Big Data use. Cooperation with them could also prove to be very fruitful. Private firms are eager to gain access to government data, while the innovative and different perspectives of private organizations could greatly benefit the public organizations in developing valuable Big Data application they did not think of by themselves.

Finally, organizations in the Dutch public sector should realize they are embarking on a new territory here. Intensively working with data, being it Big Data or more conventional forms, requires a very different approach to governing, knowledge creation and decision making. Organizations should consider this radical change by acting carefully during the transition. Rapid and extensive development of the required expertise for Big Data is good and very helpful, but actual deployment of Big Data applications should be handled more carefully. Public sector organizations should start small with single function Big Data applications that suit their current organizations and that are

completely known and understandable to them. They should make sure these applications are successful and only then start to scale up towards larger and different applications. This will allow the organizations to learn from the use of Big Data and create a positive attitude among stakeholders while minimizing the risk of problems with one or more of the previously mentioned dangers of Big Data.

Individual public organizations are not the only actors that have to take action to develop the Big Data Readiness of the Dutch public executive sector. National government bodies, such as Ministries and supporting government agencies should support the public organizations in the executive sector that are developing for Big Data use. They can help the development of Big Data Readiness in the public sector by establishing a national legal framework specifically designed for the legal implications of Big Data use. A legal framework to deals with the challenges of the collection, storage, use and re-use of data will be an important asset in the development of high quality legal compliance capabilities in public organizations. Secondly, it is not only valuable for public organizations, but also for private sector organizations struggling with similar issues in their development towards Big Data use. Furthermore, the framework should deal with the trade-off between privacy protection and value for society from using personal data in Big Data applications. By finding the right balance between the protection of privacy of citizens and the possibilities for both public and private organizations to create value from Big Data applications handling personal data, national government can set a clear benchmark of what can and cannot be done with Big Data.

Secondly, national government can play an important role in promoting more, easier and faster data sharing between public organizations in all fields and regions of government. A standard solution and format for data sharing can greatly benefit organizations in the public sector who have a demand for more data. Current Open Data policies are a good start and are making progress, but more movement towards fully sharing all data between government organizations should be accelerated and extended.

Lastly, national government could set up a knowledge or expert centre where experts from all public organizations can share experiences and knowledge. In this way public organizations can better learn from each other and easier identify potential interesting development partners. The centre could also include a knowledge base on all Big Data vendors, designers and consultant and their achievements at other public organizations, so that selection of these important partners for Big Data development and implementation projects will be much easier for public officials.

8. Conclusion & Recommendations

In chapter eight the conclusion and recommendations of the research project are presented. Firstly the answers to the research questions are briefly presented in section 8.1. In section 8.2 the conclusion is given, followed by the recommendations in section 8.3. Finally, suggestions for future research are made in section 8.4.

8.1 Answers to Research Questions

First the answers to the sub-research are presented, followed by the answer to the main research question.

Sub-Research Questions

1. What is the current situation in the Dutch public sector regarding Big Data?
    a. What are organizations in the Dutch public sector currently doing regarding Big Data?

Organizations in the Dutch public executive sector are very interested in applications of Big Data use for their organizations in the near future. Interviewed public officials from the Dutch public executive sector indicate they expect that Big Data can add significant value to their organizations in the future. They expect that organizations in the sector will want to start using Big Data in the near future.

Currently, organizations in the Dutch public executive sector are not using Big Data in their operational processes. However, some public officials indicate that their organizations are performing intensive forms of conventional data use, which is by some tentatively named Big Data. They do however acknowledge that currently no full-fledged Big Data applications are being used.

Almost all public officials from the Dutch public executive sector indicate that organizations in the sector are currently working on some form of R&D activities for Big Data. Some organizations are performing orientation research into the possibilities of Big Data, while others are performing more advanced activities in development of Big Data expertise and applications.

More briefly, organizations in the Dutch public executive sector are currently not using Big Data in their organizations, but are very interested in its use and are therefore working on its development in a varied range of R&D activities.

    a. Which uncertainties are experienced by organizations in the Dutch public executive sector in assessing their organizational readiness for using Big Data?

Organizations in the Dutch public executive sector have difficulty to assess whether they are ready to introduce Big Data use. This difficulty impedes the further development of Big Data use in the sector. Three uncertainties are identified by public officials from the Dutch public executive sector, that are impeding the assessment of organizational readiness for the introduction of the use of Big Data. These are:

- Uncertainty on the form of Big Data that is suited for the organization
- Uncertainty on the maturity of the organization to change for Big Data implementation
- Uncertainty on the capabilities of the organization to successfully use Big Data

2. What does using Big Data entail for organizations in the Dutch public executive sector?
    a. What makes using Big Data different from using more conventional data in Dutch public executive organizations?

From literature and interviews with practitioners from the Dutch public executive sector five differentiating characteristics of using Big Data in comparison to conventional data are formulated.

- The use and combination of multiple large datasets, from both external and internal sources.
- The use and combination of structured and unstructured data in analysis activities.
- Real-time or near-real-time streams of incoming data are structurally handled and analyzed.
- The use of advanced analytics and algorithms, distributed computing and/or advanced technology to handle very large and complex computing tasks.
- Innovative use of (existing) datasets and/or data sources for new and radically different applications.

The use of Big Data does not necessarily have to show all five characteristics together for it to be qualified as Big Data use. A combination of two or more characteristics will usually entail that Big Data is used. The characteristics show the implications the use of Big Data will have on the IT infrastructure and data activities of public organizations.

    b. How does the process of using Big Data in organizations in the Dutch public executive sector look like?

The process of using Big Data in organizations in the Dutch public sector is experienced by practitioners as a cyclical process with four distinct main data activities. The four activities are Data Collection, Data Combination, Data Analytics and Data Use. The connection between Data Collection and Data Combination is usually facilitated by Data Management systems. The connection between Data Analytics and Data Use are often facilitated by Business Intelligence systems. The connection between Data Combination and Data analytics represent the IT Infrastructure that supports Big Data technologies. The connection between Data Use and Data Collection represents the strategic choices organizations make in their data activities.

    c. For which applications can Big Data be used by organizations in the Dutch public executive sector and what is their specific added value?

Big Data applications in the Dutch public executive sector can be as diverse as the full package of functions that public organizations in the sector have. Possible applications of Big Data use are only limited by the data available and the creativity of its designers. However, these possible applications can be classified into three distinct types of Big Data applications with each a specific added value for their organizations. The proposed typology of the Big Data applications is based on the different types of knowledge that are created by the applications of the three types.

The first type is object/subject evaluation Big Data applications. This type of applications evaluates objects or subjects on their individual attributes and pre-set decision rules. The added value of

object/subject evaluation Big Data applications is higher quality and more granular decision support information.

The second type of Big Data applications is research applications. These applications combine huge amounts of diverse data with advanced analytical methods to find patterns and connections in the data. Research Big Data applications provide organizations with new insights they would not have been able to find in conventional data research.

The third type is continuous monitoring Big Data applications. These applications have a signalling or dash boarding function for which vast amounts of (near) real-time streaming data are used from very large sensor networks. Because of its continuous stream of data it is possible to use these applications for real world experimentation. This type of application provides organizations with a richer image of reality.

3. How can the organizational readiness for using Big Data of organizations in the Dutch public executive sector and its potential areas of improvement be assessed?

The organizational readiness of organizations in the Dutch public sector for using Big Data can be assessed by the Big Data Readiness Framework. The framework has three separate aspects that correspond with the three uncertainties stopping organizations in the Dutch public executive sector experience from assessing their readiness for Big Data. The three aspects of the framework are organizational alignment, organizational maturity and organizational capabilities. By combining the results of the three separate aspects, the total Big Data Readiness of organizations can be assessed. The produced overview of the Big Data Readiness of the organizations in the Dutch public executive sector also gives an indication of the possible areas of improvement of the Big Data Readiness of the organization by identifying the current weak spots in the alignment, maturity and capabilities of organizations in regard to the use of Big Data.

    a. How can the uncertainties in assessing organizational readiness for using Big Data experienced in practice be addressed is the assessment method?

The three uncertainties experienced by practitioners in assessing the organizational readiness for Big Data in their organizations are translated in three distinct aspects that are used as the starting points for the formulation of the Big Data Readiness Framework. The uncertainty on if and in what form Big Data is suited for the organization is translated into the alignment between the planned Big Data application and the current organization. The uncertainty on the maturity of the organization to change for Big Data implementation is translated in the organizational maturity based on e-government theory. The uncertainty if Big Data can be successfully used by the organization is translated into organizational capabilities in regard to Big Data use and its implementation. All three aspects are used to create the framework.

The first aspect of the framework is Organizational Alignment and the assessment method is adapted from the Strategic Alignment Model (Henderson & Venkatraman, 1993). The Strategic Alignment Model allows the evaluation of the alignment between a new IT system and the organization. It has four parts, which are all translated towards specific parts of the connection between public organizations and Big Data applications. The four parts of the organizational alignment aspect are the main statutory task (its organizational strategy), the strategic data activities (its organizational

infrastructure), the planned Big Data application type(s) (IT strategy) and the currently present Big Data Use characteristics (IT infrastructure). An optimal alignment is presented in the framework, which is used as a reference to check whether the organizations in the Dutch public executive sector are planning for Big Data applications that align with their main statutory tasks, strategic data activities and currently present Big Data use characteristics.

The second aspect of the framework is Organizational Maturity and the assessment method is based on the growth stage models from e-government research and more specifically the e-government growth stages model (Klievink & Janssen, 2009). The organizational maturity of organizations in the Dutch public executive sector is based on scores on activity and information sharing, IT facilities and used data systems characteristics per growth stage in the model. The dominant e-government growth stage for organizations is established and that determines the extent in which organizations need to develop to grow towards the fifth and final growth stage of demand driven, joined up government that provides the best environment for successful Big Data use. The assessment of the organizational maturity of organizations in the Dutch public sector for Big Data use is based therefore based on the distance between the current maturity level of the organization and the fifth and most advanced maturity level.

The third aspect of the framework is Organizational Capabilities. The assessment in this aspect is based on the e-government maturity model (Valdés, et al., 2011), which uses dynamic capabilities to assess how advanced public organizations are in their e-government development. By adapting the capabilities used in the model towards Big Data related capabilities, the model is made applicable for the Big Data Readiness assessment of public organizations. The adapted capabilities are identified from different fields of IT research literature and insights from practice discussed in the interviews with public officials. The seven capabilities are IT Governance, IT Resources, Internal attitude towards Big Data, External attitude towards Big Data, Legal Compliance, Data Governance and Data Science Expertise. Each of the capabilities is scored on attention, development and presence in the assessed organizations and expressed as a percentage of the maximum possible score. This provides a clear and understandable overview of the organizational capabilities for Big Data Readiness in organizations in the Dutch public executive sector.

b. Is the assessment method suited to answer the main question of the research project?

The proposed assessment method for the organizational readiness for Big Data use in public executive organizations is able to provide extensive information on Big Data Readiness in the assessed Dutch public executive organizations. The assessment method was able to provide information on the three chosen aspects, that were translated from the main uncertainties in Big Data Readiness assessment in practice. The assessment produced gave a comprehensive and clear overview of Big Data Readiness and its potential areas of improvement of the Dutch public executive sector. The assessment method could therefore be successfully validated for its main functions and goals. Verification of the assessment method is harder to do, as the framework developed in this research project is the first assessment method for Big Data readiness. Results produced by the assessment method can therefore not be comprehensive compared and tested with alternative assessment methods. However, some work on the verification of the assessment method can be done by comparing the Big Data Readiness results produced by the assessment method with the images of organizational readiness for Big Data generated from the interviews. Although these images are only based on the interviewers opinion, they do give a first indication if the proposed

assessment method concurs with them. As a successful comparison is the case, the assessment can be tentatively concluded to be suited to answer the main question of this research project.

Main Research Question:

> What is the organizational readiness for Big Data use in the Dutch public executive sector and where can it be improved?

Organizations in the Dutch public sector are not ready to start to use Big Data in their organizations. Although organizations, especially the ones experienced in using data in their organizations, have well developed capabilities to implement and use Big Data, they are not ready for it. Because of the well-developed capabilities for Big Data use, the organizations could believe they are close to be ready to use Big Data, but they should understand that the use of Big Data requires more from an organization. Dutch public executive organizations lack a full understanding of Big Data applications and their implications for their organizations. They do not understand fully that certain Big Data application need support of organizational activities and expertise that may not be present in their organizations, as they are not part of their main tasks and activities. Trying to implement these Big Data applications will result is unsuccessful use of Big Data and limited added value for the organization. And as organizations in the Dutch public sector still contemplate the implementation of not aligned Big Data applications, they are not ready for Big Data use. Furthermore, Dutch public organizations are not ready for Big Data as their organizations have not developed enough in their maturity to take full advantage of the potential value of Big Data use. Dutch public executive organizations insufficiently share information and data with each other to provide for adequate data as input for Big Data applications to produce the granular and individual information on objects and subjects public organizations need to become more demand driven. In short, Dutch public executive organizations are not ready for the implementation of Big Data use in their organizations and should develop their readiness further, before they consider starting to use it.

The readiness for Big Data use in the Dutch public executive sector can be improved in three areas. Firstly, possible improvements in the alignment between the organizations and their planned Big Data applications are developing a better understanding of the implications and prerequisites of Big Data applications amongst responsible public officials. In that way they are better able to determine which applications are best suited for their organizations. More specific areas of improvement are the current execution of Big Data use characteristics at organizations in the sector. Organizations that use and analyse data should develop and experiment with using more forms of data, for example real-time and unstructured data to develop their organizations more towards Big Data use. Organizations that collect and store data should improve their current use of that data by developing new and innovative ways to use that data and create more value from it.

Secondly, organizations in the Dutch public executive sector can improve and extend their efforts in structurally sharing activities and information with other organizations by making their data more accessible for outside stakeholders and acquiring more data from other organizations. This will improve the amount and variety of data available for future Big Data applications at all organizations, improving knowledge creation and with that the added value of Big Data use in the public sector.

Thirdly, organizations in the Dutch public sector can improve their capabilities to use and implement Big Data by increasing their expertise in data science. They can do this by attracting so-called Big Data talent, or by training their current data experts in the use of Big Data. Dutch public executive organizations should furthermore focus on improving their Data Governance programs and create data strategies to take full advantage of the data that is available to them. Lastly, organizations should improve their in-house IT expertise, strategy and decision making to ensure the implementation and integration of new IT systems for Big Data use can be done successfully. More specific, organizations that primarily use data, but do not have a lot of internal data should focus on the development the attitude of external stakeholders towards their Big Data use. This will ensure that a sufficient amount of Big Data is made available to them and that they have permission to use it in their organization. Organizations that are both collecting and using data actively can improve their readiness for Big Data use by improving the internal attitude towards Big Data use by convincing employees and management of the potential value of Big Data applications. This could for example be done by building successful business cases and small scale experiments. Additionally these organizations can improve their readiness for the implementation of Big Data use by improving their legal compliance activities. By making sure the organizations has a good strategy to ensure that the collection, storage and use of Big Data is conducted within the relevant legal frameworks in the organization, the organization can ensure the use of Big Data will be safe and secure and therefore more successful.

8.2 Conclusion

Currently, public sector organizations are not using Big Data. They are interested in using Big Data and see its potential, but are unable to establish whether they are ready for its implementation or not. And this could be a reason why they are currently not using Big Data. They also have trouble to identify where they can improve their organizations in order to develop towards a situation in which they are confident they are able to effectively start to use Big Data in their organizations. To contribute to this situation, an assessment method for Big Data Readiness in public executive organizations is developed in this research project. By assessing the Big Data Readiness of the Dutch public executive sector, it can be established if the readiness can indeed be an impeding factor in the use of Big Data in the public sector and where that readiness can be improved.

The dynamic and ambiguous definition of Big Data from literature provides little help for practitioners in understanding the concept of Big Data, so an alternative approach was proposed that explains the actual use of Big Data in organizations. By providing handles on the concept in the form of differentiating characteristics of Big Data use, the description of the process of using Big Data and by formulating three types of Big Data applications, the concept of Big Data is made more understandable and more applicable to reality and established academic knowledge.

Using three more detailed uncertainties from practice on Big Data Readiness as starting points, a Big Data Readiness Assessment framework was formulated combining concepts from literature from other fields of IT research with experience and observations from practice. In the framework the concept of Big Data, with the additional handles proposed in this thesis, is connected with established theoretical frameworks on IT alignment and e-government. Because of this the framework is able to assess the Big Data Readiness of organizations in the Dutch public executive sector.

From the assessment it becomes clear that the Dutch public executive sector is not ready for the introduction of Big Data use in their organizations. Although the capabilities for the implementation and use of Big Data of organizations in the sector are well developed, they lack a proper understanding of the implications Big Data application will have on their organizations. They do not understand sufficiently that Big Data applications need to match their current activities and that non-matching Big Data applications will not provide them with the valuable information they expect. Furthermore, Dutch public executive organizations are not structurally sharing data and information with each other on a level that will allow for them to provide their future Big Data applications with a sufficient amount and especially variety of data. This will significantly limit the value these applications can produce for the public executive organizations and will not allow them to produce the granular and individual information needed for the organizations to progress towards a more demand driven form of government. By not being ready for the use of Big Data in their organizations, Dutch public executive organizations severely limit the potential value Big Data applications can have for them. Further development of the organizations is needed, before the use of Big Data can be introduced in the Dutch public executive sector, as only then the full potential of Big Data can be exploited.

Dutch public executive organizations can improve their readiness for Big Data use by improving their understanding of what type of Big Data applications best fits with their current activities and goals and will therefore provide the most valuable knowledge to them. Secondly, organizations should improve their structural sharing and exchanging of information and data with other organizations, in order to increase the amount and variety of data that they can use as input for their Big Data applications. Additionally, organizations in the Dutch public executive sector should improve their expertise in data science and data governance, by hiring and training employees specifically for future Big Data activities. Attention to IT governance and legal compliance strategies for the use of data will also improve the chance of successful implementation and use of Big Data in the organizations.

The concept of Big data is hard to decipher, as the definition of the term is very dynamic and ambiguous. Technological progress and large differences between organizations on what constitutes Big Data to them, make the definition unusable in practice and very hard to connect to more established concepts in literature. Because of this the second objective of this thesis was to demystify the concept of Big Data and show the implications of it on public organizations. By explaining the use of Big Data in organizations in this thesis, the concept is given handles that make it more understandable and recognizable. Furthermore, the handles allow the concept to be connected with established academic concepts as e-government maturity, organizational change and IT implementation and core/dynamic capabilities of organizations. This allows for a better connection between the concept of Big Data and current established literature which positions the concept of Big Data in regard to other fields of research. In particular the field of e-government and t-government research gain from the more detailed insights into the implications of Big Data use on public organizations that are proposed in this thesis. Hopefully this will ultimately result in more successful implementation efforts of Big Data in the public sector in the future, that can add value to public sector organizations and society as a whole.

8.3 Recommendations

The first set of recommendations from this research project are of course the identified possible areas of improvement that have been described in detail in section 6.2 and again briefly in section 8.1. These areas of improvement provide practitioners with vital information on in which areas their organizations are lacking in regard to readiness for Big Data use and which they should therefore focus on in the further development of the organizational readiness for Big Data use. Apart from these very specific recommendations, a number of other points are presented here that could help practitioners with improving the Big Data Readiness of the Dutch public executive sector.

- The Dutch public executive sector should not yet start with the implementation of Big Data use for operational purposes. The organizations in the sector are not ready for it yet and significant development efforts should be considered first. Development efforts should be focused at the weak points in the organizational Big Data Readiness first, as this will raise the organizational readiness in the most effective way.
    - Improvement efforts should focus on creating a better understanding of the implications of Big Data applications for organizations and the alignment between them. Organizations should start by selecting to develop Big Data applications that fit best with their current activities and provide the type of knowledge that they need and use in their daily operations.
    - Furthermore, public organizations should make sure they improve their structural sharing and exchange of information and data with other organizations, to be able to have a sufficient amount and variety of data to use as input for their Big Data applications.
- Practitioners in the Dutch public executive sector should work towards the introduction of Big Data use, but should be careful with the introduction of it. They should make sure implementation of Big Data applications is done right, as incorrect action in this could evaporate the potential added value of Big Data use in public organizations. Incorrect or incomplete implementation could mean that the dangers of Big Data cannot be controlled, which will do more harm than good for society and will significantly impede future support for Big Data use.
- The development of the Big Data Readiness of organizations in the Dutch public executive sector should be conducted in structural cooperation with other organizations, from both the public and private sector. The widespread distribution of weak points and strong points in the Big Data Readiness of organizations in the sector allows for cooperation with other organizations for mutual learning and further development. By working together on the development of Big Data use, all parties will develop faster and have more expertise and data at their disposal.
- The Big Data Readiness Framework is an assessment tool that can help practitioners assess how ready their organizations are for the introduction of Big Data use. It should be used regularly in the development process to check progress and to identify new weak points in the organizational readiness. Furthermore, the different aspects of the framework provide for a comprehensive list of requirements for Big Data use in public organizations that can be used for the development of development plans and Big Data development roadmaps.

8.4 Future Research

As the Big Data Readiness is proposed and used for the first time during this research project, a number of suggestions for future research on the framework alone are obvious. The framework is currently only used to assess the Big Data Readiness of eleven public organizations from the Dutch public executive sector. As is described in section 6.3, this makes extensive evaluation of the framework currently not possible. The first suggestions for future research are therefore specifically formulated to improve this situation, so that the framework can be properly and extensively evaluated.

- The Big Data Readiness Framework should be applied to more organizations from the public executive sector, to establish if it can also create an overview of Big Data Readiness for more and more varied organizations.
- Future research should also focus on the detailed connection between Big Data and the three separate parts of the framework. Now these connections are made based on interviews with eleven practitioners from the Dutch public sector and they should be tested more extensively.
- Future research could aim to establish improved methods and decision rules for the analysis of the answers to the questionnaire to produce the assessment scores for the framework. The analysis methods are developed in this research project after several options have been tested on the gathered data. The methods and decision rules could however be less applicable on other, more extensive datasets and may be improved for them. Specifically the averaging of the scores for the seven capabilities to one average score over all capabilities may be improved to better deal with weak points in the capabilities of an organization. An example of a more extensive and perhaps better performing assessment method can be found in the e-government maturity model (Valdès, et al., 2011). The method was deemed to extensive for the scope of this research project, but could prove to perform very well.
- The form of the Big Data Readiness framework is currently based on the three identified uncertainties on organizational readiness assessments in organizations in the Dutch public sector. This makes the framework well suited for assessments in the Dutch public executive sector, but perhaps not very well suited for assessment in other countries as uncertainties may be different there. Future research into uncertainties on Big Data Readiness assessment in other countries could provide further insights into the applicability of the framework to other countries.
- The aspect of organizational maturity on the basis of e-government growth stage models seemed to be less applicable than anticipated. Future research should focus on finding an alternative established concept in literature that could potentially be more applicable to assess whether organizations are mature enough to make the changes required for Big Data use implementation.

Further suggestions for future research besides the framework itself are also interesting and could potentially yield important insight on the concept of Big Data.

- Future research on the applicability of the different parts of the explanation of Big Data use in more organizations can provide further insights into what the use of Big Data precisely entails for organizations. In this line of research organizations from science and the private

sector should also be included, to see if the concepts proposed in this thesis report are recognizable for practitioners from different field working with Big Data.

- Research into the implications of Big Data in the practice of the organizational capabilities linked with Big Data in this thesis report could provide further insights for practitioners on how to develop their organizations for the introduction of Big Data use better. Especially interesting would be to investigate the detailed implications of Big Data applications on IT Governance and Data Governance in organizations. Also, legal compliance strategies will be severely pressurized by the use of Big Data in organizations.

9. Reflection

After the completion of the research project and the largest part of the thesis report, it is time for reflection on the perspectives, assumptions and choices used in this research project. Firstly, the perspective from literature used as a basis for this thesis is discussed.

As is described in the introduction of this report the perspective is taken that Big Data has the potential to be an enabling technology for further development in e-government maturity and possibly even for a structural transformation towards true digital-era, demand driven and joined up government. Several authors state that Big Data technologies can and will have substantial influence on the form of government in the future. I tend to agree with them that in the most promising cases Big Data technologies can have substantial influence on government organizations and their activities in the future. I believe Big Data technologies can potentially enable the transformation of parts of functions and activities of public organizations towards more flexible and demand driven forms. Further cooperation and joint functions over multiple public organizations are also quite likely in my view. Therefore I have chosen to use this perspective from literature for my thesis. However, I also have reservations about the almost unlimited potential that Big Data is presumed to have according to a large number of authors. The unlimited potential that has created a global hype for everything Big Data in the last two years. Yes, using Big Data has potential, and yes, it has the potential to greatly improve and transform certain parts of our society, but it is by no means a panacea that can be applied in any situation and for every problem in our world.

Some authors have us believe using Big Data is as easy as dumping a large pile of data on a server, crunch it with HADOOP and 10 minutes later you have the most wonderful insight you would never have found without Big Data. Of course, reality is a lot less comfortable. Even collecting, cleaning, classifying, selecting, extracting, transforming and loading that pile of data is a daunting task in itself, let alone selecting and properly using the right analysis methods and finally interpreting the result with the very large probability of not finding anything of use. Big Data is not a miracle cure, it is a lot of hard work. Therefore, it may be more probable that Big Data is not a key enabler for the public sector to develop towards new forms of government. Big Data may just be a very advanced IT tool. Valuable for the specialists that work with it and important for the extremely complex activities it is used for, but not the accessible IT revolution that can be used by every data savvy public official to create extraordinary results for the reform of his of hers organization.

And even if Big Data technologies do develop into an all-round and valuable IT solution that can be of use for virtually any data related activity in our society, than it remains questionable if Big Data technology itself is the enabler of e-government reform. A lot of literature on the development of e-government, which is also used as a theoretical basis in this thesis, is based on the idea that new technologies drive that development. And this assumption of technology push in this development may not necessarily be true. Although most of the e-government activities performed by public organizations today cannot exist without present day technology, it remains questionable if the introduction of these technologies pushed the public organizations to develop their e-government activities (Lips, 2012). Yes, these technologies enabled them to develop their e-government activities further, but the new technologies did not necessarily push them forward. The development of e-government was not pushed by technology, it was pushed by institutional policies and cultural demands from within public organizations and higher government bodies. Technology is only the enabler that allows organizations to develop their digital activities as they see fit. Big Data in this regard is no different. In the most positive perspective Big Data has the potential to enable government to transform to more advanced organizational forms and activities, but Big Data will not

push public organizations towards it. Organizational change will always be the product of the interplay between institutional factors, culture and technology and not just one of the three. In most literature on Big Data this important point is often overlooked and Big Data technologies are credited with a transformational power that is does not have. Big Data technologies should be seen as an advanced tool that can enable an organization with a data orientated culture and non-restrictive institutional factors to develop more towards a data driven, flexible and cooperating version of itself. Big Data should not be seen as anything more, nor as anything less.

Another point of critique on the huge potential Big Data technologies are credited with is already partially mentioned in the previous paragraphs. It is the point that using Big Data is much more complex than they are often portrayed to be. There is a distinct possibility that using Big Data may prove to be just too complex for public organizations to use on a large scale. And this can happen for a number of reasons. One can be that the predicted shortage of Big Data experts in the near future will cause public organizations to not be able to acquire the required expertise to develop the use Big Data. Private sector organizations will always be able to offer higher fees and will attract the bulk of the talents available. Secondly, the vast amounts of personal data on civilians handled by public organizations may prove to be too difficult to use legally in Big Data applications, evaporating the potential value Big Data technologies have to offer to the public sector. Thirdly, Big Data applications need diverse data from multiple sources to really create added value in regard to more conventional data use. If the sharing of data and information between public organizations is not improved, the potential value of Big Data technologies will never be realized in the public sector. Big Data may be too difficult for the public sector to perform on a large scale and with that, whether or not it has the potential to change government for the better will not even matter.

Currently Big Data is mostly seen in the public sector as a new form of IT system that has to be implemented into the organization, in the same fashion that systems like EDI, BI and ERP systems. This paradigm of radical or disruptive IT implementation in the form of whole systems has been used for the last few decades and is likely to continue, especially in the more traditional public sector. However, some authors insist, Big Data should not be implemented in this fashion, but much more in a more natural, incremental and experimental way that allows for continuous learning, innovation and development (Franks, 2012; Hopkins & Evelson, 2012). Supporters of this line of thought indicate that using Big Data and finding or creating valuable and useful applications for it is a very creative process, which works most effective in very flexible and small environments. Incremental trial and error processes and experimentation are important facets of Big Data development according to them and it is questionable whether these facets can be found in the larger and more traditional organizations in the public sector. That this paradigm works has been proven by many successful, small startups that have developed single applications for Big Data and exploited them, before moving on to the next problem they could solve with Big Data. If this paradigms is the more suited and successful one for the development of Big Data applications, then public organizations will have to change their views of the implementation of Big Data in their organizations substantially. In that case, two possible solutions can be followed. One, development of Big Data applications could be outsourced to small, flexible and creative private organizations that design the Big Data applications and run them as a service for the larger public organizations to buy. Two, larger public organizations create in-house expert centers for Big Data applications that have the same characteristics as smaller, more flexible and creative organizations with separate environments in which they can safely and independently test their Big Data applications before putting them in practice in the larger public organization.

Finally, in this thesis it is assumed that the main role of public organizations is that they are the users of Big Data. It is assumed that all public organizations in the Dutch public executive sector have the desire to use Big Data in their own organizations, build their own Big Data applications and collect or acquire the required data for their own organization. And while this currently seems to be the dominant paradigm in the Dutch public executive sector in this early stage of Big Data development, it is not the only way to approach the development of Big Data. Some of the representatives of public organizations that were spoken to for this research project mentioned that their organizations are contemplating the role they should take regarding Big Data. Public organizations that do not have a pressing need for the value that Big Data can potentially offer can also choose not to jump on the Big Data bandwagon. Big Data may not be needed at all for every public organizations out there and not every public organization should participate in Big Data use for it to develop to a success. Public organizations can take up a role as provider of data. They can focus on collecting data and improving its quality, so that other organizations can use it in their Big Data applications. The other way around is also possible. Public organizations with substantial Big Data expertise and facilities could provide these for other organizations with less structural need for them as a service. Private organizations can also be involved in this shift towards specialization within the data value chain.

In what form the development of Big Data in the public sector will take place remains uncertain for the coming years. The scenarios discussed here are all possible and a favorite has yet to be identified. What is clear is that the current Big Data hype is set too positive and that both practitioners and academics should focus more on how to mitigate the dangers Big Data brings. I believe that Big Data has a large potential to add value to our society. In what form and led by who will be the question. The perspective taken on Big Data in this thesis reflects both the hopes and expectations of practitioners and academia that Big Data can help public organizations perform better. However, I hope that the sketched overview of the readiness of organizations in the Dutch public executive sector will also send a clear message that starting with using Big Data now is much too early for the public sector and that more development is needed first. If this thesis is able to deliver that message and simultaneously help with a better understanding of the concept of Big Data and its implications and requirements, I believe the main goal for it is achieved.

References

Adelman, S., Moss, L., & Abai, M. (2006, Winter). Data Strategy. *Information Builders Magazine*.

Adrian, M. (2011). *Information Management Goes 'Extreme': The Biggest Challenges for 21st Century CIOs.* Stamford, Connecticut, United States: Gartner Inc.

Algemene Rekenkamer. (2007). *Lessen uit ICT-projecten bij de overheid.* Den Haag: Algemene Rekenkamer.

Algemene Rekenkamer. (2013). *Bezuinigingen op uitvoeringsorganisaties.* Den Haag: Algemene Rekenkamer.

Almeida, F., & Calistru, C. (2013). The main challenges and issues of big data management. *International Journal of Research Studies in Computing*, 11-20.

Andersen, K. V., & Henriksen, H. Z. (2006). E-government maturity models: Extension of the Layne and Lee model. *Government Information Quarterly*, 236–248.

Andersen, K., Henriksen, H. Z., & Medaglia, R. (2010). Fads and Facts of E-Government: Fads and Facts of E-Government:. *International Journal of Public Administration*, 564–579.

ANP. (2014, april 25). 'Nederland miljarden kwijt door falend ict'. *Volkskrant*.

Bertot, J. C., & Choi, H. (2013). Big Data and e-Government: Issues, Policies, and Recommendations. *The Proceedings of the 14th Annual International Conference on Digital Government Research.* Quebec City, QC, Canada: ACM.

Bertot, J. C., Jaeger, P., & McClure, C. R. (2008). Citizen-centered E-Government Services: Benefits, Costs, and Research Needs. *'08 Proceedings of the 2008 international conference on Digital government research* (pp. 137-142). ACM.

Bollier, D. (2010). *The Promise and Peril of Big Data.* Washington D.C., United States: The Aspen Institute.

Bresnahan, T. F., & Trajtenberg, M. (1995). General Purpose Technologies "Engines of Growth?". *Journal of Econometrics, Annals of Econometrics, vol. 65, no. 1*, 83-108.

Brown, B., Chui, M., & Manyika, J. (2011, October). Are you ready for the era of 'big data'? *McKinsey Quarterly*.

Brown, B., Court, D., & McGuire, T. (2014, March). Views from the front lines of the data-analytics revolution. *McKinsey Quarterly*.

Bryant, R., Katz, R., & Lazowska. (2008). Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society. In *Computing Research Initiatives of the 21st Century.* Washington DC, United States: Computing Research Association.

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond Computation: Information Technology and Organizational Transformation. *Journal of Economic Perspectives*, 23-48.

Bughin, J., Livingston, J., & Marwaha, S. (2011). Seizing the Potential of Big Data. *McKinsey Quarterly*.

Chen, H., Chang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 1165-1188.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Network Applications*, 171-209.

Cordella, A., & Bonina, C. M. (2012). A public value perspective for ICT enabled public sector reforms: A theoretical reflection. *Government Information Quarterly*, 512–520.

Courtney, M. (2012, September). The Larging-up of Big Data. *Engineering & Technology Magazine*, pp. 72-75.

Cox, M., & Ellsworth, D. (1997). Application-Controlled Demand Paging for Out-of-Core Visualization. *Proceedings of the 8th conference on Visualization '97* (pp. 235-244). Los Alamitos, CA, USA: ACM.

Cumbley, R., & Church, P. (2013). Is "Big Data" Creepy? *Computer Law & Security Review*, 601-609.

DAMA International. (2009). *The DAMA Guide to the Data Management Body of Knowledge.* Bradley Beach, NJ: Technics Publications.

Daniel, E. M., & Wilson, H. N. (2003). The role of dynamic capabilities in e-business transformation. *European Journal of Information Systems*, 282-296.

Davenport, T. H., Barth, P., & Bean, R. (2012). How Big Data is different. *MIT Sloan Management Review*, 22-24.

Davenport, T., & Dyché, J. (2013). *Big Data in Big Companies.* Portland, OR, USA: International Institute of Analytics.

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, 377-383.

Department of Finance and Deregulation. (2013). *Big Data Strategy — Issues Paper.* Parkes, Australia: Commonwealth of Australia.

Dhar, V. (2013, December). Data Science and Prediction. *Communications of the ACM*, pp. 64-73.

Dunleavy, P., & Margetts, H. (2010). The second wave of digital era governance. *American Political Science Association Conference.* Washington DC, United States.

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New Public Management Is Dead—Long Live Digital-Era Governance. *Journal of Public Administration Research and Theory*, 467–494.

Ebrahim, Z., & Irani, Z. (2005). E-government adoption: architecture and barriers. *Business Process Management Journal*, 589-611.

Executive Office of the President. (2014). *BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES.* Washington DC, USA: The White House.

Federal Big Data Commission. (2012). *Demystifying Big Data.* Washington D.C., United States: TechAmerica Foundation.

Finney, S., & Corbett, M. (2007). ERP implementation: a compilation and analysis of critical success factors. *Business Process Management Journal*, 329-347.

Franks, B. (2012, October 3). To Succeed with Big Data, Start Small. *Harvard Business Review*.

Gartner. (2012). *Information Economics, Big Data and the Art of the Possible with Analytics.* Stamford, CT, USA: Gartner Inc.

Ghoneim, A., Irani, Z., & Sahraoui, S. (2011). Guest Editorial. *European Journal of Information Systems* , 303-307.

Gilbert, D., Balestrini, P., & Littleboy, D. (2004). Barriers and benefits in the adoption of e-government. *The International Journal of Public Sector Management*, 286-301.

González-Bailón, S. (2013). Social Science in the Era of Big Data. *Policy & Internet*, 147-160.

Gottschalk, P. (2009). Maturity levels for interoperability in digital government. *Government Information Quarterly 26*, 75-81.

Groenfeldt, T. (2012, March 30). *Morgan Stanley Takes On Big Data With Hadoop*. Retrieved from Forbes: http://www.forbes.com/sites/tomgroenfeldt/2012/05/30/morgan-stanley-takes-on-big-data-with-hadoop/

Grover, V., Teng, J., Segars, A., & Fiedler, K. (1998). The influence of information technology diffusion and business process change on perceived productivity: The IS executive's persppective. *Information & Management 34*, 141-159.

Gurbaxani, V., & Whang, S. (1991). The impact of Information Systems on Organizations and Markets. *Coomunications of the ACM 34-1*, 59-73.

Heeks, R. (1999). *Reinventing Government in the Information Age: International practice in IT-enabled public sector reform.* London: Routledge.

Henderson, J., & Venkatraman, N. (1993). Strategic Alignment: Leveraging Information Technology for transforming organizations. *IBM Systems Journal*, 472-484.

Henschen, D. (2012, October 24). *Why Sears Is Going All-In On Hadoop*. Retrieved from Information Week: http://www.informationweek.com/it-leadership/why-sears-is-going-all-in-on-hadoop/d/d-id/1107038?

Hopkins, B., & Evelson, B. (2012, February 3). *Forrester: Big data – start small, but scale quickly*. Retrieved from ComputerWeekly.com: http://www.computerweekly.com/opinion/Forrester-Big-Data-Start-Small-But-Scale-Quickly

ICT Roadmap team. (2012). *Roadmap ICT for the Top Sectors.* Den Haag: Ministerie van Economische Zaken.

IDC. (2011). *Extracting Value from Chaos.* Framingham, MA, United States: IDC.

IDC. (2012). *The Digital Universe in 2020: Big Data, Bigger Shadows and Biggest Growth in the Far East.* Framingham, MA, United States: IDC.

Jeyaraj, A., Rottman, J. W., & Lacity, M. C. (2006). A review of the predictors, linkages, and biases in IT innovation adoption research. *Journal of Information technology 21*, 1-23.

Joseph, R. C., & Johnson, N. A. (2013, November). Big Data and Transformational Government. *Computing Now*, pp. 43-48.

Junqué de Fortuny, E., Martens, D., & Provost, F. (2014). Predictive Modeling With Big Data: Is Bigger Really Better? *Big Data*, 215-226.

Kamal, M. M. (2006). IT innovation adoption in the government sector: identifying the critical success factors. *Journal of Enterprise Information Management*, 192-222.

Kelly, J. (2013, April 27). *Amazon Web Services: 1 Million Hadoop Clusters And Counting*. Retrieved from Services Angle: http://servicesangle.com/blog/2012/04/27/amazon-web-services-1-million-hadoop-clusters-and-counting/

Kirkpatrick, M. (2010, Augustus 4). Google CEO Schmidt: "People Aren't Ready for the Technology Revolution". *Huffington Post*.

Kiron, D. (2013). Organizational Alignment is Key to Big Data Success. *Sloan Management Review*, 1-6.

Kiron, D., & Schockley, R. (2011). Creating Business Value with Analytics. *MIT Sloan Management Review*, 57-63.

Klievink, B., & Janssen, M. (2009). Realizing joined-up government — Dynamic capabilities and stage models for transformation. *Government Information Quarterly 26*, 275–284.

KPMG. (2012). *Big Data. The Risks and Rewards Locked in Vast Oceans of Data.* United States: KPMG LLP.

Laney, D. (2001). 3D Management: Controlling Data Volume, Velocity and Variety. *META Group Research Note 6*.

Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. *Government Information Quarterly*, 122–136.

Lee, G., & Kwak, Y. H. (2012). An Open Government Maturity Model for social media-based public engagement. *Government Information Quarterly 29*, 492-503.

Lin, L. M., & Hsia, T. L. (2011). Core capabilities for practitioners in achieving e-business innovation. *Computers in Human Behavior*, 1884-1891.

Lips, M. (2012). E-Government is dead: Long live Public Administration 2.0. *Information Polity*, 239-250.

Lohr, S. (2012, February 11). The Age of Big Data. *The New York Times*.

Malik, P. (2013). Governing Big Data: Principles and Practices. *IBM Journal of Research and Development*.

Margetts, H., & Sutcliffe, D. (2013). Addressing the Policy Challenges and Opportunities of "Big Data". *Policy & Internet*, 139-146.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work and Think.* London: John Murray Publishers.

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*.

McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition and productivity.* New York, NY, United States: McKinsey & Co.

Melville, N., Kraemer, K., & Vijay, G. (2004). Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS Quarterly*, 283-322.

Milakovich, M. E. (2012). Anticipatory Government: Integrating Big Data for Smaller Government. *Internet, Politics, Policy 2012: Big Data, Big Challenges?* Oxford: University of Oxford.

Miller, H. G., & Mork, P. (2013, February). From Data to Decisions: A Value Chain for Big Data. *Computing Now*, pp. 57-59.

Ministerie van Binnenlandse Zaken en Koningsrijkrelaties. (2012). *Big Data. Wat moet de overheid ermee?* Den Haag: Ministerie van Binnenlandse Zaken en Koningsrijkrelaties.

Ministerie van Infrastructuur en Milieu. (2014). *Voorbereiding thematische kenniskamer Omgaan met open data en Big Data.* Den Haag: Ministerie van Infrastructuur en Milieu.

Ministerie van Infrastructuur en Milieu. (2014). *Voorbereiding thematische kenniskamer Omgaan met open data en Big Data + Samenvatting conclusies kenniskamer.* Den Haag: Ministerie van Infrastructuur en Milieu.

Mullich, J. (2013). *Closing the Big Data Gap in Public Sector.* New York City, United States: Bloomberg Businessweek Research Services.

Negash, S., & Gray, P. (2003). Business Intelligence. *Ninth Americas Conference on Information Systems*, (pp. 3190-3199).

Ngai, E. W., Law, C. C., & Wat, F. K. (2008). Examining the critical success factors in the adoption of enterprise resource planning. *Computers in Industry*, 584-564.

OpenTracker. (2013, september 23). *Definitions of Big Data*. Retrieved June 6, 2014, from OpenTracker: http://www.opentracker.net/article/definitions-big-data

Orlikowski, W. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, 398-427.

Policy Exchange. (2012). *The Big Data Opportunity. Making government faster,smarter and more personal.* London: Policy Exchange.

Premkumar, G. (2003). A Meta-Analysis of Research on Information Technology Implementation in Small Business. *Journal of Organizational Computing and Electronic Commerce*, 91-121.

Proffitt, B. (2012, April 27). *Toronto Hospital Detects Infection With Analytics*. Retrieved from IT World Canada: http://www.itworldcanada.com/article/toronto-hospital-detects-infection-with-analytics/45411

Provost, F., & Fawcett, T. (2013). Data Science and its relationship to Big Data and data-driven. *Big Data*, 51-59.

Prybutok, V. R., Zhang, X., & Ryan, S. (2008). Evaluating leadership, IT quality, and net benefits in an e-government environment. *Information & Management 45*, 143–152.

Puron-Cid, G., Ramon Gil-Garcia, J., & Luna-Reyes, L. (2012). IT-Enabled Policy Analysis: New Technologies, Sophisticated Analysis and Open Data for Better Decision Making. *The*

Proceedings of the 13th Annual International Conference on Digital Government Research (pp. 97-106). College Park, MD, USA: ACM.

Rajagopalan, M., & Solaimurugan, V. (2013). Big Data Framework for National e-Governance Plan. *Eleventh International Conference on ICT and Knowledge Engineering.* IEEE.

Ramakrishnan, R., & Gehrke, J. (2003). *Database management systems (Vol. 3).* New York: McGraw-Hill.

Ramón Gil-García, J., & Pardo, T. A. (2005). E-government succes factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, 187–216.

Rijksoverheid. (2013). *Hervormingsagenda Rijksdienst.* Den Haag: Rijksoverheid.

Robey, D., Im, G., & Wareham, J. D. (2008). Theoretical foundations of empirical research on interorganizational systems: assessing past contributions and guiding future directions. *Journal of the Association for Information Systems*, 4.

Ross, J., Beath, C., & Quaadgras, A. (2013). You May Not Even Need Big Data After All. *Harvard Business Review*, 90-98.

Russom, P. (2011). *Big Data Analytics.* Renton, WA, United States: TDWI Research.

Sagiroglu, S., & Sinanc, D. (2013). Big Data; a Review. *IEEE*, 42-47.

Scholl, H. J., & Scholl, M. C. (2014). Smart Governance: A Roadmap for Research and Practice. *iConference 2014 Proceedings*, (pp. 163–176).

Scholl, H., & Klischewski, R. (2007). E-Government Integration and Interoperability: Framing the Research Agenda. *Intl Journal of Public Administration*, 889–920.

Seifert, J. W. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. *Government Information Quarterly*, 461–480.

Shiri, A. (2014). Linked Data Meets Big Data: A Knowledge Organization Systems Perspective. *Advances In Classification Research Online*, 16-20.

Simon, P. (2013). *Too Big to Ignore: The Business Case for Big Data.* Hoboken, New Jersey, USA: John Wiley & Sons Inc.

T.W. Powell Co. (2008). *The Knowledge Value Chain:How to Fix It When It Breaks.* New York, New York, USA: T.W. Powell Co.

Tambe, P. (2014). Big Data Investment, Skills, and Firm Value. *Management Science*.

The Economist. (2010). *Data, Data everywhere. A special report on managing information.* London: The Economist Newspaper Limited.

Toorn, v. d. (2014, July 10). Big data kent restricties. *Public Mission*.

United Nations Department of Economic and Social Affairs. (2014). *UNITED NATIONS E-GOVERNMENT SURVEY 2014. E-Government for the Future We Want.* New York, USA: United Nations.

Valdés, G., Solar, M., Astudillo, H., Iribarren, M., Concha, G., & Visconti, M. (2011). Conception, development and implementation of an e-Government maturity model in public agencies. *Government Information Quarterly 28*, 176–187.

Valdès, G., Solar, M., Astudillo, H., Iribarren, M., Concha, G., & Visconti, M. (2011). Conception, development and implementation of an e-Government maturity model in public agencies. *Government Information Quarterly 28*, 176–187.

Veenstra, A. F. (2013). *Big Data in Small Steps: Assessing the Value of Data.* Delft: TNO.

Waarts, E., Everdingen, Y. M., & Hilligersberg, J. v. (2002). The dynamics of factors affecting the adoption of innovations. *The Journal of Product Innovation Management*, 412–423.

Weerakkody, V., & Dhillon, G. (2008). Moving from E-Government to T-Government: A Study of Process Reengineering Challenges in a UK Local Authority Context. *International Journal of Electronic Government Research*, 1-16.

Weerakkody, V., Janssen, M., & Dwivedi, Y. K. (2011). Transformational change and business process reengineering (BPR): Lessons from the British and Dutch public sector. *Government Information Quarterly 28*, 320–328.

Wikipedia. (2014, May 21). *Big data*. Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Big_data

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS quarterly*, 17-41.

Wu, J. H., & Hisa, T. L. (2008). Developing e-business dynamic capabilities: an analysis of e-commerce innovation from I-, M-, to U-commerce. *Journal of Organizational Computing and Electronic Commerce*, 95-111.

Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of computer information systems*, 23-32.

Yildiz, M. (2007). E-government research: Reviewing the literature, limitations, and ways forward. *Government Information Quarterly* , 646–665.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research 21(4)*, 724-735.

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information Technology and the Changing Fabric of Organization. *Organization Science 18(5)*, 749-762.

Appendices

I - Organizations and Interviewed Representatives

Organization	Representative(s)	Main Tasks/Activities	Organization type
Zorginstituut Nederland	Hans (JJGM) van den Hoek Advisor strategy	"-adviseren en duiden verzekerde aanspraken Zvw en AWBZ en het systematischj doorlichten van de zorg in aandoeningencluster (ICD-10) -bevorderen van de kwaliteit van de gezondheidszorg (in brede zin) - Uitvoren van de zogenaamde burgerregelingen (wanbetalers, oncerzekerden, buitenland, gemoedsbezwaarden, onverzekerbare vreemdelingen). - uitvoering risicovereving -Fondsbeheer Zorgverzekeringsfonds en Algemeen fonds bijzondere ziektekosten"	ZBO/ non-departmental public body
Gemeente Utrecht	Pieter in 't Hout Information and process manager	Local government of the city of Utrecht. Governmental services for citizens for	Municipality/Local public administration
DORDRECHT	Jan Wind Area manager East Gemeente Dordrecht MO/WGW	Local government of the city of Dordrecht. Governmental services for citizens for	Municipality/Local public administration
RDW	Martine van Heijnsbergen Advisor Information Provision	RDW is de registratie- en toelatingsautoriteit voor voertuigen.	ZBO/ non-departmental public body

Organization	Representative(s)	Main Tasks/Activities	Organization type
BRANDWEER Amsterdam-Amstelland	Barry van 't Padje Program manager Brandweer Intelligence	"- Risicobeheersing van brand en ongelukken met gevaarlijke stoffen (van proactie tot nazorg). - Hulpverlening in relatie tot ongelukken die de fysieke veiligheid bedreigen."	Emergency services
Koninklijk Nederlands Meteorologisch Instituut *Ministerie van Infrastructuur en Milieu*	Raymond Sluiter Researcher Geo-ICT, with focus on data	Het KNMI garandeert als onafhankelijke autoriteit Nederland de best beschikbare informatie op het gebied van weer, klimaat en aardbevingen. Dit draagt bij aan de veiligheid, bereikbaarheid, leefbaarheid en welvaart van Nederland	Agency of Ministerie van Infrastructuur en Milieu
kadaster	Dick Eertink Advisor strategy and policy	Het Kadaster waarborgt de rechtszekerheid in de vastgoedmarkt en zorgt voor een optimale informatievoorziening van vastgoed- en geo-informatie.	ZBO/ non-departmental public body
Rijkswaterstaat *Ministerie van Infrastructuur en Milieu*	John Steenbruggen Innovation manager on Geo data, ICT and Big Data	Rijkswaterstaat is de uitvoeringsorganisatie van het ministerie van Infrastructuur en Milieu die in opdracht van de Minister en Staatssecretaris de nationale weg- en waternetwerken beheert en ontwikkelt en werkt aan een duurzame leefomgeving. Rijkswaterstaat werkt samen met anderen aan droge voeten, voldoende en schoon water, een duurzame leefomgeving, vlot en veilig verkeer over weg en water en betrouwbare en bruikbare informatie in relatie tot de genoemde taken.	Agency of Ministerie van Infrastructuur en Milieu

Organization	Representative(s)	Main Tasks/Activities	Organization type
Planbureau voor de Leefomgeving	Hiddo Huitzing Researcher, project manager Johan van der Schuit project manager Geo information	De kernactiviteiten zijn dus strategische beleidsanalyses over ruimtelijk/milieu en natuur beleid, ex ante, ex durante en ex post (voor, tijdens en na afloop van het beleid). Daarnaast agenda vormende onderzoeken naar onderwerpen die voor het nationale beleid op het gebied van milieu, natuur en ruimte belangrijk zijn.	Independent planning agency of Ministerie van Infrastructuur en Milieu
Centraal Bureau voor de Statistiek	Piet Daas Methodologist and research coordinator Big Data	Produceren van relevante en kwalitatief betrouwbare cijfers over Nederland	ZBO/ non-departmental public body
Landelijk Informatiecentrum Voertuigcriminaliteit RDW / Politie / VbV Belastingdienst	Jeroen de Graaf Data Scientist	Het LIV is een informatie- en kenniscentrum dat een bijdrage levert aan de opsporing en bestrijding van criminaliteit. Door de nauwe samenverwerking tussen bovenstaande partijen ontstaat een betere informatie-uitwisseling. Hierdoor wordt voertuigcriminaliteit in een vroeg stadium gesignaleerd. Dit is onder andere mogelijk doordat de partijen - waar dat wettelijk mag - gebruik maken van elkaars informatiesystemen.	public-private partnership organization.

II - Interviews

Conversation topic list for the unstructured interviews

- Introduction of the research project, its goals and scope

- Big Data definition, development and possible hype.
    - Big Data term, definition.
    - Giant leap forward, emergency, forced or possibilities?
    - Successful examples in the private sector
    - Influence of Big Data on the organization, way of working, decision making and strategy.

- Current situation regarding (Big) Data at the organization
    - What is the organization doing with Big Data? Are you currently using Big Data, is it under development, is it being monitored?
    - Data collection/acquisition/sensors
    - Data storage/combination/accessibility
    - Data analysis/mathematics/computing/statistics/modeling
    - Using the data/knowledge creation from data.
    - Currently used data systems/concepts and their development.

- Using Big Data in the organization
    - Difference using Big Data and conventional data
    - Applications

- Initiative for using Big Data, execution and evaluation in the organization

- Big Data in the future. Changes in organization and changes in the field

- Conditions for using Big Data
    - Data (quality, open/linked data, Structured/unstructured, formats)
    - Analytics
    - Compliance/Personal data (collection/sharing/using data)
    - Privacy and security
    - Big Data expertise, data scientists
    - Technology, hardware/software, in-house/outsourced/cloud?

- Big Data Challenges/Issues/Obstacles/Uncertainties
    - Why currently not used?/Why is it not currently developed?
    - What is holding the organization back?

- Other organizations cooperating with the organization involved in Big Data?

Summary of interview answers

Organizations	1	2	3	4	5	6	7	8	9	10	11
Big Data activities											
Collection - automatic	No	No	No	Yes	No	No	Yes	No	No	No	Yes
Collection - manual	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No
Storage/Stream	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Analytics	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Used for operations/DM	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Collection focus	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Usage focus	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Big Data Characteristics											
Combination - internal	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Combination - external	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No
Structured data	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unstructured data	Yes	No	No	Yes	Yes	No	Yes	No	No	Yes	No
(Near) Real-Time	No	No	No	Yes	No	No	No	No	No	No	Yes
Advanced Analytics	Yes	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes
Advanced computing	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Innovative use of existing data	Yes	No	Yes	No	No	No	No	Yes	Yes	Yes	No
Big Data Applications											
Object evaluation	No	Yes	Yes	No	No	No	No	Yes	No	Yes	No
Continues monitoring	No	No	No	Yes	No	No	No	No	No	No	Yes
Research	Yes	Yes	No	No	Yes	No	No	Yes	Yes	Yes	Yes
Data Characteristics											
Personal data	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No
Open/linked data	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes		Yes
Volume	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Velocity	No	No	No	Yes	No	No	No	No	No	No	Yes
Variety	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No
Big Data conditions											
Data Quality	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Analytics Quality	Yes	Yes	No	Yes	No	No	No	No	Yes	Yes	Yes
Data Driven DM	No	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes
IT Resources	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Data Science Expertise	Yes	Yes	Yes	No	No	No	No	No	Yes	Yes	Yes
Legal Compliance	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	No
IT Governance	No	Yes	Yes	No	No	No	Yes	No	Yes	Yes	Yes
Expectations											
Big Data in own words?	Yes	No	No	Yes	Yes	No	No	Yes	No	Yes	No
Big Data in the future?	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Opportunity?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Necessity?	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	No
Current activities											
none	No	No	No	No	No	Yes	Yes	No	No	No	No
Research on possibility	No	No	No	Yes	Yes	No	No	No	No	No	Yes
Research on usability	Yes	No	No	No	No	No	No	No	Yes	No	No
development of data use	No	Yes	Yes	No	No	No	No	Yes	No	Yes	No
development of big data	No	No	No	No	No	No	No	No	No	No	No

III - Questionnaire

Legend

Standard open questions

Scale questions (7 point scale)

Multiple choice questions

Your organization

1. Could your briefly describe the main activities performed by your organization?
2. Could you briefly describe the main statutory tasks for your organization?
3. Which type of organizational tasks best describes your organization?
    a. Management and Administration
    b. Evaluation and Research
    c. Registration and Documentation
    d. Indirect coordination and/or diverse project based tasks
    e. None of the above
4. What role does digital data and information play in performing the main tasks of your organization?
5. What role does IT and IT expertise play in performing the main tasks of your organization?
6. How intensive is your organization structurally performing the activity “Data Collection”? “Data Collection” is the purposeful creation of datasets from the processes of the organization. Acquisition of data from other organizations is NOT included.
7. How intensive is your organization structurally performing the activity “Data Use”? “Data Use” is the structural use of data and derived information for the operational processes of the organization.

Current activities with data

8. Can you specify to what extent the current data activities match the following statements?
    a. My organization uses and combines multiple datasets, from both internal and external information sources.
    b. My organization uses and combines both structured and unstructured data.
    c. My organization analyses and uses large volumes of data that are coming in real-time of near real-time.
    d. My organizations uses complex analytical methods and advanced hardware to process and analyse data.
    e. My organization uses existing data and data sources innovatively for new and very different applications than what the data is collected for.
9. Is your organization using Business Intelligence or similar systems?
    a. Yes
    b. No
10. Is your organization using Database Management or similar systems?
    a. Yes
    b. No

11. Is your organization currently sharing data with other (public) organizations?
    a. Yes
    b. No
12. Is your organization currently providing Open Data?
    a. Yes
    b. No
13. Is your organization currently sharing almost all of its data with other organizations or providing it almost all as Open Data?
    a. Yes
    b. No
14. Can you specify to what extent the following statements on your current activities and available information are applicable to your organization?
    a. Current activities and available information are divided between separate departments within the organization.
    b. Current activities extend beyond the limits of the departments and information is available throughout the organization.
    c. Current activities and available information are accessible for outside the organization.
    d. Current activities and available information are extensively shared with other organizations.
    e. Current activities and available information are organized centrally, so that all involved organizations and stakeholders can access and use it on demand.
15. Can you specify to what extent the following statements on IT facilities are applicable to your organization?
    a. When possible all activities are actively supported with IT facilities and applications.
    b. A organization wide IT infrastructure makes data available throughout the organization.
    c. The IT infrastructure makes it possible that data and applications are accessed from outside of the organization.
    d. The IT infrastructure makes it possible that all data and applications are shared with other organizations.
    e. All IT facilities are organized centrally (for example nationally) and accessible for all interested organizations and stakeholders.

Definition of Big Data

16. What do you and your organizations think the term Big Data means?

Interest for Big Data

17. Are you and your organizations interested in using Big Data in your organization? And if so, why are you interested?
18. What value and/or benefits do you expect to gain from using Big Data?

Big Data Use in your organization

19. Is your organization currently working with Big Data and/or developing it?
    a. Yes
    b. No
20. What is your organization currently primarily doing with Big Data?
    a. Using Big Data for primary, operational processes in the organization.
    b. Developing Big Data through orientation, research and development projects/activities.
21. Do you expect your organization to start developing Big Data through orientation, research and development projects/activities in the near future?
    a. Yes
    b. No
22. Could you specify which of the following statements is most applicable to the current or near future research & development activities regarding Big Data in your organization?
    a. Almost no internal activities, only monitoring external developments.
    b. Research into the possibilities of the use of Big Data.
    c. Research into the feasibility of the use of Big Data.
    d. Development of knowledge/expertise for working with large amounts of data.
    e. Development of knowledge/expertise for working with Big Data.
    f. Developing specific Big Data applications
    g. Other:
23. Do you expect your organization to start using Big Data for the primary, operational processes in the near future?
    a. Yes
    b. No
24. Why do you think your organization will not start working with Big Data in the near future?

Big Data Applications

25. Could you briefly describe what kind of Big Data applications your organization is using or is planning to use in the future?

On the following three pages a type of Big Data application will be presented with a short explanation on each page, followed by three questions.

Type 1 – Object/subject evaluation

Applications in which Big Data is used to evaluate the status of objects and/or subjects. Examples are: Risk profiling, Client segmentation, Fraud detection, Preference predictions, etc. This type of Big Data application primarily offers decision support information.

26. Could you specify to what extent your organization would be interested in this type of Big Data applications?
27. Could you specify to what extent this type of Big Data applications is applicable for the main tasks/activities of your organization?
28. Could you specify to what extent this type of Big Data is feasible to perform in the near future for your organization?

Type 2 – Research

The use of Big Data to find new relations and patterns, researching causalities and evaluate large scale effects/developments. This type of application primarily provides new insights.

29. Could you specify to what extent your organization would be interested in this type of Big Data applications?
30. Could you specify to what extent this type of Big Data applications is applicable for the main tasks/activities of your organization?
31. Could you specify to what extent this type of Big Data is feasible to perform in the near future for your organization?

Type 3 – Continuous monitoring

The use of Big Data to accurately measure, understand and monitor the environment in real-time. This type of Big Data application primarily provides information that provides a richer image of current events and reality.

32. Could you specify to what extent your organization would be interested in this type of Big Data applications?
33. Could you specify to what extent this type of Big Data applications is applicable for the main tasks/activities of your organization?
34. Could you specify to what extent this type of Big Data is feasible to perform in the near future for your organization?

Need for Big Data

35. Could you specify to what extent your organization has a need or will develop a need for the following possible benefits of Big Data use ?
    a. Increased transparency of decision making process, as decision are based more on data.
    b. More insight into actual effects of policy choices and decisions, as more information on it is available.
    c. Circumstances in reality can be made more insightful, as more and richer information on it is available.
    d. Higher quality and more effective decision making, as more and higher quality information is available.
    e. Faster, more efficient decision making, possible even automated, as more information is available and available much faster.
    f. Future policies will be more effective, as effects and consequences of past policies are made more insightful.
    g. With more real-time data, real world experiments can be executed, which will improve future policy effectiveness.
    h. More and faster available information will speed up decision making and policy (re-)design, which will improve efficiency of these processes.
    i. Continuous monitoring will lead to faster discovery of suboptimal decisions and ineffective policies, so they can be adapted sooner.

Capabilities for Big Data

36. Could you specify to what extent the organizational capabilities described below are important for Big Data use in your organization?
    a. Data Governance. This includes data quality management, database management, data acquisition, data sharing, etc.
    b. Managing internal attitude towards Big Data. This includes top management support, presence of a Big Data champion, experience in using (Big) data in the organization, overall opinion on Big Data, etc.
    c. Data Science Expertise development. Data Science is the bundling of knowledge in the organization on IT/programming , Mathematics/Statistics and the organizations core business.
    d. Compliance strategy. Legal compliance concerning protection of privacy, purposeful data collection and data use, data ownership, etc.
    e. IT Governance/Strategy. Purposeful design and execution of decision making processes and responsibilities concerning IT in the organization.
    f. IT infrastructure. Design, maintenance and development of IT infrastructure , such as hardware, network, enterprise systems, etc. and its corresponding expertise.
    g. Managing external attitude towards Big Data. This includes relationships with Big Data consultants, suppliers and experts, Big Data partnerships with other organizations, managing expectations and demands from higher level public organizations and citizen and stakeholder opinions on Big Data use in the organization.
37. Could you specify to what extent the organizational capabilities described below are present in your organization?
    a. Data Governance. This includes data quality management, database management, data acquisition, data sharing, etc.
    b. Managing internal attitude towards Big Data. This includes top management support, presence of a Big Data champion, experience in using (Big) data in the organization, overall opinion on Big Data, etc.
    c. Data Science Expertise development. Data Science is the bundling of knowledge in the organization on IT/programming , Mathematics/Statistics and the organizations core business.
    d. Compliance strategy. Legal compliance concerning protection of privacy, purposeful data collection and data use, data ownership, etc.
    e. IT Governance/Strategy. Purposeful design and execution of decision making processes and responsibilities concerning IT in the organization.
    f. IT infrastructure. Design, maintenance and development of IT infrastructure , such as hardware, network, enterprise systems, etc. and its corresponding expertise.
    g. Managing external attitude towards Big Data. This includes relationships with Big Data consultants, suppliers and experts, Big Data partnerships with other organizations, managing expectations and demands from higher level public organizations and citizen and stakeholder opinions on Big Data use in the organization.
38. Could you specify to what extent your organization is capable of developing the organizational capabilities described below?
    a. Data Governance. This includes data quality management, database management, data acquisition, data sharing, etc.

b. Managing internal attitude towards Big Data. This includes top management support, presence of a Big Data champion, experience in using (Big) data in the organization, overall opinion on Big Data, etc.
c. Data Science Expertise development. Data Science is the bundling of knowledge in the organization on IT/programming , Mathematics/Statistics and the organizations core business.
d. Compliance strategy. Legal compliance concerning protection of privacy, purposeful data collection and data use, data ownership, etc.
e. IT Governance/Strategy. Purposeful design and execution of decision making processes and responsibilities concerning IT in the organization.
f. IT infrastructure. Design, maintenance and development of IT infrastructure , such as hardware, network, enterprise systems, etc. and its corresponding expertise.
g. Managing external attitude towards Big Data. This includes relationships with Big Data consultants, suppliers and experts, Big Data partnerships with other organizations, managing expectations and demands from higher level public organizations and citizen and stakeholder opinions on Big Data use in the organization.

Obstacles for Big Data Use

39. Are you and your organization currently experiencing obstacles that are preventing the use of Big Data in your organization?
40. What could contribute to mitigating these obstacles?
41. Who or which organizations could contribute to mitigating these obstacles?

IV – Assessment scorecard decision rules

Alignment Readiness assessment

A. Application fit with statutory task type and data intensity:

Highest scoring (>50%) application fits statutory task: 3 points

Multiple high scoring (>50%) applications, among which best fitting application: 2 points

Best fitting applications scores low (<50%): 1 point

B. Big Data characteristics currently present in data use fitting with needed characteristics:

All three characteristics highly present (>70%): 3 points

One characteristic not highly present (>70%):: 2 points

Two or more characteristics not highly present (>70%):: 1 point

Score (= A + B: range 2-6 points)	Alignment assessment
2	Very Low
3	Low
4	Medium
5	High
6	Very High

Maturity Readiness assessment

Maturity Level = highest level in which average score > 75% AND all three aspects are above 60% AND previous level also qualifies.

Maturity Level	Alignment assessment
1	Very Low
2	Low
3	Medium
4	High
5	Very High

Capabilities Readiness assessment

Capabilities score is average score of all separate capabilities. Separate capabilities scores is actual score divided by maximum score expressed in a percentage.

Average score (%)	Alignment assessment
0%-60%	Very Low
61%-70%	Low
71%-80%	Medium
81%-90%	High
91%-100%	Very High

Overall Big Data Readiness assessment

Overall score is the sum of assessment scores of the three aspects of the framework

Assessments	Score
Very Low	1
Low	2
Medium	3
High	4
Very High	5

Readiness Score range is therefore 3-15.

Readiness Score	Big Data Readiness assessment
3-5	Ready for orientation on Big Data Use
6-7	Research into Big Data Use possibilities and requirements
8-9	Ready for planning and early design of Big Data Use
10-11	Ready for further development of Big Data Use.
12-13	Ready for detailed development and testing of Big Data Use
14-15	Ready for Big Data Use implementation

Big Data Readiness Assessment based on the phases typically formulated in design and implementation projects and used in general IT roadmaps in practice.

V - Detailed Big Data Readiness Assessments per Organization - Example

Organization	1
Alignment	
Main Statutory task type	Administration, Management
Data Collection intensity	High
Data Use intensity	High
Best aligned application	All, but especially Continues Monitoring Applications
Corresponding Big Data characteristics	- Real-time or near real-time - Advanced Analytics & Algorithms - Innovative use of existing data
Application Types	
Object/Subject Evaluation	82%
Research	82%
Continuous monitoring	82%
Big Data characteristics	
Internal & External	71%
Structured & Unstructured	57%
Real-time	71%
Advanced Analytics	71%
Innovative Data Use	71%
Alignment assessment	Very High
Alignment area for improvement	-
Maturity	
Stove-piped organizations	*86%*
Activities and Information sharing	86%
IT facilities	71%
Data systems/concepts	100%
Integrated organizations	*86%*
Activities and Information sharing	86%

IT facilities	71%
Data systems/concepts	100%
Nationwide portal	*68%*
Activities and Information sharing	43%
IT facilities	71%
Data systems/concepts	90%
Inter-organizational integration	*71%*
Activities and Information sharing	86%
IT facilities	29%
Data systems/concepts	100%
Demand driven, joined-up government	*73%*
Activities and Information sharing	71%
IT facilities	71%
Data systems/concepts	76%
Maturity level	2
Maturity assessment	Low
Maturity area for improvement	Sharing more activities and information with other organizations and creating IT facilities that makes all data in the organization accessible by other organizations.
Capability	
Data Governance	81%
Internal Attitude	71%
Data Science Expertise	62%
Legal Compliance	86%
IT Governance	81%
IT facilities	81%
External Attitude	71%
Capability Average Score	**76%**
Capability Assessment	Medium
Capability area for improvement	Increasing Data Science Expertise, Improving Internal and External attitude towards Big Data

Alignment Assessment	Very High
Maturity Assessment	Low
Capability Assessment	Medium
Big Data Readiness score	10
Big Data Readiness Assessment	Ready for further development of Big Data Use.