AN ASSESSMENT OF BIG DATA POLICIES AND BIG DATA MATURITY IN PUBLIC ADMINISTRATION IN TURKEY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ARAS OKUYUCU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF POLITICAL SCIENCE AND PUBLIC ADMINISTRATION

SEPTEMBER 2018

Approval of the Graduate School of Social Sciences

Prof. Dr. Tülin Gençöz
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Ayşe Ayata
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Nilay Yavuz
Supervisor

Examining Committee Members

Prof. Dr. Mete Yıldız (Hacettepe Uni., SBKY)

Assoc. Prof. Dr. Nilay Yavuz (METU, ADM)

Assoc. Prof. Dr. Mustafa K. Bayırbağ (METU, ADM)

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Aras Okuyucu

Signature :

ABSTRACT

AN ASSESSMENT OF BIG DATA POLICIES AND BIG DATA MATURITY IN PUBLIC ADMINISTRATION IN TURKEY

Okuyucu, Aras
Msc. Program of Political Science and Public Administration
Supervisor: Assoc. Prof. Dr. Nilay Yavuz
September 2018, 154 pages

Governments have been the largest data collectors since the beginning of their existence. For this reason, big data applications play an important role in public administration. This thesis aims to assess big data policies and big data maturity in public administration in Turkey. In the study, a literature analysis was performed first to understand the use of big data in public administration. In the light of the analysis, big data maturity models were then examined to inform the development of big data policies. Using the country samples, a review was carried out on the big data application areas, so a framework for big data analysis was drawn. Turkey has a long tradition of data collection; it is also a developing country in terms of the use of information technologies in public administration. Therefore, this work initialy researched the historical development of big data policies in Turkey. After this discussion, a selected big data maturity assessment model was applied to Turkey's case to answer where Turkey is in its e-transformation process, and suggestions were made to improve big data applications.

Keywords: Big Data, e-Government, Public Administration, Public Policy, Turkey

ÖZ

TÜRKİYE'DE KAMU YÖNETİMİNDE BÜYÜK VERİ POLİTİKALARININ VE BÜYÜK VERİ OLGUNLUĞUNUN DEĞERLENDİRİLMESİ

Okuyucu, Aras
Yüksek Lisans, Siyaset Bilimi ve Kamu Yönetimi Bölümü
Tez Yöneticisi: Doç. Dr. Nilay Yavuz
Eylül 2018, 154 Sayfa

Devletler varoluşlarının başından beri en büyük veri toplayıcılarındandır. Bu sebeple büyük veri uygulamaları kamu yönetiminde önemli bir rol oynamaktadır. Bu tezin amacı, Türkiye'de kamu yönetiminde büyük very politikalarının ve büyük veri olgunluğunun değerlendirilmesidir. Çalışmada kamu yönetiminde büyük verinin kullanımının anlaşılabilmesi için öncelikli olarak bir literatür analizi gerçekleştirilmiştir. Analizin ışığında kamu kurumlarının büyük veri politikaları geliştirebilmeleri için büyük veri olgunluk modelleri incelenmiştir. Ülke örneklerinden yola çıkarak uygulama alanlarına ilişkin araştırma yapılmış, böylece büyük veri analizine ilişkin bir çerçeve çizilmiştir. Türkiye, veri toplama konusunda köklü bir devlet geleneğine sahiptir, aynı zamanda kamuda bilişim teknolojilerinin kullanımı açısından da gelişmekte olan bir ülkedir. Bu sebeple, bu çalışma ilk olarak Türkiye'de büyük veri politikalarının tarihsel gelişim sürecinin incelenmesini içermektedir. Bu tartışmadan sonra, büyük veri politikalarının anlaşılabilmesi için büyük veri olgunluk modeli aracılığıyla e-dönüşüm sürecinde bu politikalar açısından Türkiye'nin nerede olduğu araştırılmış, büyük veri uygulamalarının iyileştirilebilmesi için öneriler getirilmiştir.

Anahtar Kelimeler: Büyük Veri, e-Devlet, Kamu Yönetimi, Kamu Politikaları, Türkiye

To My Parents

ACKNOWLEDGMENTS

I would like to thank my supervisor Assoc. Prof. Dr. Nilay Yavuz for her guidance throughout this research. Her dedication, encouragement and positive attitude made this thesis possible. I'm sure it is hard to work with a cramming student for an advisor, for this reason I cannot express my gratitude enough for her effort, feedback and criticism throughout this process. I am also grateful to examining committee members Prof. Dr. Mete Yıldız and Assoc. Prof. Mustafa Kemal Bayırbağ for their suggestions and criticism for improving this thesis.

I would also like to thank Haruki Murakami, who has a great jazz and classical music library. Through listening to tunes from his playlist, I concentrated on to complete this thesis.

I would like to thank, Ece Başkesen, Uğur Ercan, Gültekin Yılmaz, Tuğçe Caner and Fatih Yiğitler for their friendship and support.

and thanks Merve Çubukçu for her love, motivation and never-ending "goygoyu bırak çalış"

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BDMM	Big Data Maturity Model
BRIC	Brazil, Russia, India and China
DPT	State Planning Agency (Devlet Planlama Teşkilatı)
EU GDPR	European Union General Data Protection Regulation
G2B	Government to Business
G2C	Government to Citizen
G2G	Government to Government
ICT	Information and Communication Technologies
IoT	Internet of Things
IT	Information Technology
NIST	National Institute of Standards and Technology
OGP	Open Government Partnership
TBD	Turkey Informatics Association (Türkiye Bilişim Derneği)
ROI	Return on Investment
R&D	Research and Development
SQL	Structured Query Language
TÜBİTAK BİLGEM	TÜBİTAK Informatics and Information Security Research Center (Bilişim ve Bilgi Güvenliği Araştırma Merkezi)
TÜBİTAK B3LAB	TÜBİTAK Big Data and Cloud Computing Laboratory
TDWI	Transforming Data with Intelligence

CHAPTER 1

INTRODUCTION

Digitalization of all spheres of life represented in the movies, sci-fi novels and other media becomes a reality in our era. Although this does not consist of stereotypical façade of technology, the era of information and communication technologies (ICT), improving artificial intelligence, and introduction of smart devices shows the fact that the iconic computer in Stanley Kubrick's movie 2001:A Space Odyssey, HAL 9000 is an application in handheld devices rather than a red light in a room full of sounds and flickering lights. Whether it is named as Google Assistant, Siri or Alexa, today we can communicate with computers at the level of artificial intelligence. Similarly, the Internet of Things (IoT) that enables communication among several technological devices themselves is also on the rise. We increasingly live in smart cities that employ time, cost, and energy saving technology and real-time information.

Among all these technological advancements, one of the most important issues relates to the significance of data production and usage (Maciejewski, 2016; Jansen & Kuk, 2016). The ongoing process of civilization, hoarding vast amounts of data since the first clicking sounds of a hammer on a nail which writes the first letters on a stone tablet has come to a point that human civilization is producing enormous amount of data in daily life. Since then there has been a logarithmic increase in data production. Humanity stored so much of data that all the books that have been written in the history of civilization would fill 175 terabytes of hard disks. For a scale 2,5 quintillion bytes (2,5 million terabytes) of data produced every day. Since the emergence of many other means to collect data, such as streaming data coming from censors, financial, audiovisual, social media, resulted in huge numbers in data production, the meaning and the structure of data have also shifted to what is called "big data".

Big data is basically defined as the datasets that are too awkward to work with using traditional, database management tools (Yiu, 2012). This includes examples like a nation's health records (O'Reilly, 2015), education policies (Lavertu, 2015), Google search data, social media usage, yearly agricultural results etc. Since it is too huge to analyze and to give a better meaning, new database analyzing algorithms, ICT's, are just beginning to use these datasets in a more efficient way.

Big data applications exist in diverse sectors, including public bureaucracies. Anyone who has been to a public service building is accustomed to see thousands of folders laid down on the shelves and many more are staying in the archives of the buildings. The governments are collecting data like never before, but the raw historical data which is described as offline data is inert since categorizing, digitizing and making it machine readable makes it a vast burden on governments. Governments until this era tried to handle the challenge of recording and disposing of different types of data, and unfortunately experienced difficulties in managing both the public and private sector while performing essential tasks (Liu & Yuan, 2015)

On the other hand, the rise of ICT services in public sector and the development of the e-government practices lead governments to collect vast amounts of digital data and enable them to process it. This in turn links to the point of using the big data to increase efficiency (Tomar, 2016), increase the quality of services (Bertot, 2013), crime detection (Ho & McCall, 2016), reduce red tape (Giest, 2017) and others; that is what the users usually cannot achieve with traditional datasets. Big data not only improves the bureaucratic procedures and public policy making but it also offers many other advantages ranging from governmental transparency, to improving well-being of the citizens (Kim, et al., 2014) even to National Olympic Teams success (Marr, 2015). Governments work to find the best possible ways to increase service qualities with these datasets. As mentioned before, there are possible applications to use, they are also using an ongoing policy to open these datasets to public so that not only these vast sets restricted to the usage of data analysis or policymakers, but also to the researchers,

scientists, public sets. This would lead to finding new ways to use these data and improve the services, and at worst the privacy risks.

Many countries around the world, including Turkey, have developed and implemented policies to adopt big data applications. With the increasing use of big data and related applications in government, there has been a growing interest in academia in relation to big data-related questions. For example, studies explore whether governments are aware of the importance of ongoing big data revolution or are they just trying to cope with the stream to get on hype train (Giest, 2017). Yiu (2012) states that just because a government can do something with big data, that does not mean that it should do it, as some policies would turn out to be a failed policy in short run (Heeks, 2003). The advantages and disadvantages of using big data are described in many scholarly articles (El-Darwcihe, 2014; Gietz, 2017; Hammer, Kostroch, & Quirós, 2017). There are also evidence-based examples in the usage of big data analytics in business models such as Procter & Gambler's marketing models. However, business analytics cannot back up the public administration applications from all aspects. Since these applications are new, more theoretical and policy discussions are needed. Particularly, there are only a few studies that evaluate big data policies in public administration. In addition, measuring the maturity of big data in public administration context is not well-explored in the literature. Although there are some models that are developed to assess big data maturity levels in private organizations, application of these models to public bureaucracies are limited. Considering the fact that big data use in public administration in Turkey has been developing recently, it is essential to evaluate existing policies and big data maturity levels to inform future policies.

Within this framework, the purposes of this thesis study are twofold: 1) to assess big data policies in public administration in Turkey, and 2) to understand the overall level of big data maturity in the public sector in Turkey using a selected big data maturity assessment model. Aiming to answer these questions, a qualitative analysis of big data policies in Turkey is done by researching data and information from governmental

reports, strategic planning documents, legal documents, and the secondary survey data collected by TÜBİTAK B3LAB and Türkiye Bilişim Derneği (TBD). In addition, the big data maturity model developed by Kuraeva (2016) is applied to assess the big data maturity level of Turkish public administration.

Since, the research tries to understand where Turkey stands in big data policies, besides reporting on the legal documents issued after 2003, some other documents listed on http://www.bilgitoplumu.gov.tr/yayinlar/ webpage are also scanned: Turkey Informatics and Economic Modernization Report (1993) , KAMUNET Studies on Transition to E-Government (2002), TBD First Information Technologies Forum Report (2002), e-Turkey Initiative Action Plan(2002), 58th Government Urgent Action Plan (2003), e-Transformation Turkey Project Short Term Action Plan for 2003-2004, Result Report for 2003-2004 e-Transformation Project Action Plan, e-Transformation Project 2005 Action Plan, 2006-2010 Information Society Strategy and Action Plan, Result Report of 2006-2010 Information Society Action Plan(2012), 10th 5 Year Development Plan.

Besides, 2013-2014 National Cyber Security Strategic Plan, 2015-2018 Information Society Strategy and Action Plan, 2016-2019 National E-Government Strategy and Action Plan, National Broadband Strategy and Action Plan 2017 are reviewed and analyzed to explain a general point of view data centric approach in e-government.

Also as secondary survey data, two sources are used. TUBITAK created a Cloud Computing and Big Data lab B3LAB, the lab regularly organizes workshops since it's foundation in 2014. In 2015 Big Data Workshop, they conducted a survey study asking 70 personnel from various organizations, nine questions about their big data awareness and models in public organizations (the list of participants is provided in APPENDIX 1). These data are used to discuss the general outlook and awareness of big data in public organizations in Turkey. In May 2016, TBD created a Big Data Applications Workgroup conducted a survey study including 29 personnel from 18 public

organizations; their answers that are provided in APPENDIX 2 are also used as a secondary data to assess awareness about Big Data Applications in Turkey. The answers are then linked with the maturity model developed by Kuraeva (2016).

There are various maturity and transformation models (Andersen & Henriksen, 2012). To understand a rational place of organizations, maturity models are used for developing policy strategies for organizations. Since 1970's first introduced by Nolan, organizational maturity levels are used for policy development. The maturity models suggest a general outlook of a transformation process. A research is conducted through using databases Google Scholar, Ebsco, IEEE Xplore and CiteSeerX to find specific big data maturity model for governments come up with two results Klievink et al.s Big Data Maturity Model and and Kuraeva's Big Data Maturity for Governments and Public Agencies. Klievink's model is used institution-by-institution basis, whereas Kuraeva 's (2016) model provided a broader framework of analysis. For this reason Kuraeva's model is used to assess Turkish Public Administration's general outlook via using secondary research data and policy applications.

There are five main sections in this thesis. Chapter 2 gives a detailed literature review on big data, Chapter 3 presents countries from around the world to show examples of big data applications aiming to show well implied applications and comparable countries with Turkey. Chapter 4 describes the development of data-centric policies in Turkey via official documents, and gives examples on big data policies in Turkey, and assesses big data maturity levels. Chapter 5 summarizes the findings and discusses the results.

CHAPTER 2

LITERATURE REVIEW ON BIG DATA

2.1 Definition of Data

The daily use of the word data, is the plural version of latin word "datum". The word etymologically rooted in the word "dare" which is "-to give". Since Latin has regular conjugations, neuter past version, datum is the singular word for the root of word. Inferring the roots, data is historically incorporated with something given.

A Google N-Gram[1] search for the word "data" in English language dates the usage of the word as back to 17th Century. This usage represents the older scientific attribution to the word until the 20th Century. Although commonly used in our era, according to the Etymology Dictionary the first usage of the word "data" as a computer term goes back to 1946 when it began to be described as "transmittable and storable computer information"[2]. This usage represents an overlapping process, since the modern computer science solidified in the events following the Second World War, which saw the dawn of modern data encryption such as the success of Turing Machine algorithms against the Enigma Code of German Military and leading Allies to obtain data of German Military's strategic plans. From then on, the importance of data analytics folded tremendously since it plays as one of the important milestones for power of information in 20th century.

The benefits and outcomes of data use can be explained through incorporating

[1] N-Gram is a computational linguistic methodology for searching a word's n times frequency among text or speech. Google N-Gram scanned 15% of all printed press in the history of mankind, and used by social scientists as a "computational lexicology that tries to understand human behavior and cultural trends through the quantitative analysis of texts."(Viktor Mayer-Schonberger, 2013, p.14)

[2] (n.d) Definition of Data retrieved: 25.06.2017
http://www.etymonline.com/index.php?term=data&allowed_in_frame=0

Ackoff's classification of the human mind (Ackoff, 1989) According to Ackoff, the human mind uses five processes to understand and give meaning to any given subject. These processes include data, information, knowledge, understanding and wisdom. The raw facts or symbols and empirical observations represents the data collected. Information is the second part where these data are processed according to giving meaningful answers about basic questions when and where, so that the layers of raw knowledge begins to reveal itself. The third layer, knowledge, explains the "how" question, by doing so the multilayered structured layer of data signifies more intrinsic value. Understanding explains the "why" question. The last part, wisdom is essential to human beings, which consists of processing and giving meaning to the level of cognition materialized at this point. Ackoff states that "Information, knowledge, and understanding enable us to increase efficiency, not effectiveness. The difference between efficiency and effectiveness—that which differentiates wisdom from understanding, knowledge, information, and data—is reflected in the difference between development and growth. Growth does not require an increase in value; development does. Therefore, development requires an increase in wisdom as well as understanding, knowledge, and information" (Ackoff, 1996: 1). Although Ackoff argues that computers would not have wisdoms of human beings, the current developments in data sciences indicate that by using big data analytics and machine learning properties, computers would achieve a similar responsiveness in the future (Boyd & Crawford, 2012; Maritz, 2013).[3]

One of the earliest examples that illustrate the use of data to solve social problems was John Snow's mapping of the incidence of cholera deaths in London in 1854. The low-level cholera districts were taking pumped water from the upstream, but the high-level cholera deaths were using water pumped from downstream location of the Thames. By mapping the disease data, Snow suggested that the high level cholera might be

[3] Burke, S (2014) Maritz: Pivotal Platform Will Sidestep Amazon "Tax" for Big Data Apps retrieved: 13.07.2018 https://www.crn.com/news/applications-os/240152130/maritz-pivotal-platform-will-sidestep-amazon-tax-for-big-data-apps.htm?pgno=2

related to sewer contamination of the district's water. The officials removed the handle of the pump from the specific area and epidemic was contained.

Nowadays with the help of censors and big data spatial analysis, city police departments are reducing the crime rates, using the public transportation maps in a better way. People have been producing so much data, but the governments are recently beginning to understand how to utilize it. (Crossier, 2001).

2.2 The Meaning of Big Data

Big data is a new and fuzzy term that appears to hold different meanings for different groups and stakeholders (Chen & Mao, 2014; Stough & McBride, 2014; Zheng, 2017). Although people are aware of its importance, there are different opinions on the definition of big data.

Merriam-Webster Dictionary defines big data as "an accumulation of data that is too large and complex for processing by traditional database management tools". According to Oxford Dictionary, big data refers to "extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions". In another definition, big data refers to the use of a massive amount of data to conduct analyses so that the data patterns and relationships can be used for classification, clustering, anomaly detection, prediction, and other needs in decision making (Mills, 2012). McKinsey Global Institute (2011) defines big data as, "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (p. 1). UN's Big Data for Development report defines big data as explosion in the quantity and diversity of high frequency digital data. It is generated by the people and collected by the government and commercial firms (Mills et al., 2012). The reason behind this is that the concept which data scientists and other users approaches is so vast and stratified that a common explanation would not be enough to unfold the idea behind what big data is.

Three questions were raised at first about big data: “how to store, how to move, how to search?”. Fisher, Drucker & Konig (2012) explains the complexity of analyzing big data by underlining that big data is too complex to store, capture and analyze with traditional data-management tools.

Over the years the value and interest in big data have increased (Chen & Mao, 2014). IBM, Microsoft, Oracle and, Facebook have been investing in big data since 2005. Tech companies’ increasing interest affected academia, and Nature published a special issue in 2008; Science in 2011.

Changes in the storage technology, increasing connectivity with the Internet and developments in computer science have led to an increase in the importance of data generation, collection and process. The history of “big data” as we know today goes few years back with these developing storage, data speed and processing models. The big data is the word of 2010’s (Gillis & Stephany, 2014). Interest in academic research on big data in public administration has also been rising since 2011 gradually (Fredriksson et al., 2017)

2.3 Properties of Big Data

The research firm Gartner defines big data as a “high-volume, high-velocity, and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making and process automation”. The properties of big data mentioned in this definition -volume, velocity and variety- are called as 3V’s of big data (McAfee, et al., 2012) and are depicted in Figure 1 below. There are also other views where the number of properties are increased by adding “veracity” (Gantz & Reinsel, 2012; Jarmin & O,Hara, 2015), “value” (Hitzler & Yanowicz, 2013; Buccholts et al., 2014;); and “variability” and “visualization” (McNulty & Eilenn, 2014). Transforming Data with Intelligence (TDWI) adds another three and reaches 10V’s by adding “validity”, “vulnerability”,

and “volatility”. The practices of adding varies, because each sector approaches from different perspective to the big data concept.

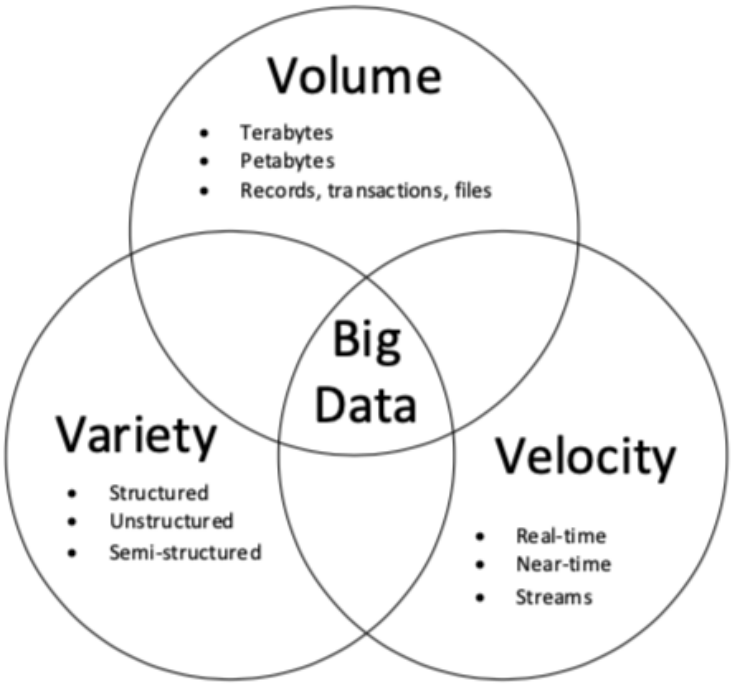


Figure 1: 3V’s of Big Data

Source: Russom, P. (2011). Big data analytics. TDWI best practices report, fourth quarter, 19(4), p. 11.

Volume is the amount of data generated, collected and processed (Buchholz, et al., 2014). The volume property is what first comes to mind actually when a person who thinks about what big data is. The relativity of volume of the data should be taken into account while describing the data, because amount of data generated everyday increases, a huge volume of today’s data might be a smaller daily part of data in the future. Bill Gates said in 1981, “649 KB ought to be enough for anybody.” and 10 MB’s of hard disks (costing $3398) were described as the “hard disk you’ve been waiting for”. The volume is not the predominant concern for big data (Ills, et al.., 2012: 12). For example, until 15 years ago a 3½ floppy disk was a data standard which can store 1,44 megabytes of data. The dawn of DVD’s with 4.8 gigabytes of storage and

more modular and higher storage USB sticks erased them from common usage. Today there are memory cards that are smaller than pinky finger and have 512 GB of storage, which indicates capacity of a half a million floppy disk can be crammed into it. With increasing data storage, another advantage for volume is introducing compression algorithms that make this component of data management a less concerned part of it. Since the dawn of bureaucracy the state is one of the biggest collectors of data. The common representation of state in audiovisual media is given with stacks of documents on the walls. Therefore the volume property is an intact and omnipresent part of the public administration process.

In big data context, *velocity* represents the increase in the speed of data. For this reason, Japan invested $87.5 millions to develop a 400 gbps high speed network infrastructure (Gamage, 2016). Laney, who defined the 3V's of Big Data in 2001, explains velocity as the point-of interaction. Speed is increasingly perceived as a competitive differentiator, so is an organization's ability to manage data velocity. When the organizational structure of public agencies is taken into account, like private business the velocity of data is more important than the volume because it leads to the responsiveness of the communicating ends. Under the press of daily demands for action, often constructed as ''crises,'' decision makers feel the need to act without delay.

Variety consists of structured, semi-structured and unstructured data types. The different sources of data, ranging from strings to audiovisual data creates the variety of data. The NIST report emphasizes the importance of variety by mentioning the rise of information coming from new sources both inside and outside the walls of the organization that creates integration, management, governance, and architectural pressures on IT. Connecting datasets with various structures would end up with new results that could not have been possible to analyze with traditional database management models. Desouza and Jacob (2017) emphasize that as datasets becomes increasingly complex, structured to unstructured variety of data might lead better

understanding of the defining characteristics of an organization's real, or potential, data repositories; and it would lead insights into the level and types of investments needed for organizations to achieve better big data analytics capacity.

Veracity is the quality of the data analysis outcome, which makes it trustworthy. By avoiding systematic errors and controlling the trustworthiness of the sources through new data analysis technique, the veracity of the data occurs (Gandomi & Haider, 2015).

By exploiting better sources and getting three main V's out of them, *Value* is created (Boyd & Crawford, 2012). It is the economical outcome of the processed data. According to Buccholtz et al. (2016), exploiting the insights gathered from vast amounts of data can not only increase companies' revenues, lower operating costs or create new services, but also improve national security, decrease health risks and help policy makers to target their policies in a better way.

Variability is constantly changing data, influx of various data. Since the datasets are exploitable not for a specific end, the datasets would give differing outcomes through different analysis (Cavoukian & Jonas, 2012).

Visualization refers to the representation of the data. A common source for data analysis is geospatial data. As one of the earliest examples of big data mentioned in academia, Cox & Ellsworth (1997: 1) stated that: "Visualization provides an interesting challenge for computer systems: Data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of big data."

Although TDWI adds *Validity* as the sources' accuracy, it represents more or less what *Veracity* signifies. *Vulnerability* is about security concerns in big datasets and one of the most commonly mentioned risks about using them. Even anonymized data can be

de-anonymized through analysis. *Volatility* is also relatable with value and velocity, storing inept data through time would only cause increasing expenditures for keeping them. If needed a dataset can be reachable quickly but on the other hand it should not be expensive to maintain.

2.4 Unstructured, Semi-Structured, and Structured Data

Big Data is located in various internal and external sources, and can consist of structured data, unstructured data such streaming data, social media data, and geospatial data (Braun, 2015).

Due to different structures of data, an e-book file, or a streaming video file has different analytical methods for analysis. The structured data makes up 15% of the whole data we are processing throughout the day and easiest to work on conventional models. On the other hand, what actually makes up big data comes with a fuzzier concept. A streaming video service, collected data from censor data from a specific district, or even a street bump application for drivers creates different layers on the subject. Within this context, this part is going to explain different data structures.

With the technology and computing power of the last decades, it is mostly possible to process structured data. One of the defining properties of big data is the unstructured property of it. According to the Big Data Public Working group, “Fifteen percent of the information today is structured information, or information that is easily stored in relational databases of spreadsheets, with their ordinary columns and rows. Unstructured information, such as email, video, blogs, call center conversations, and social media, makes up about 85% of data generated today and presents challenges in deriving meaning with conventional business intelligence tools” (Mills et al., 2012, p. 10). Stéphane Hamel a digital analytics consultant explains this by saying that the Simplest definition of big data is that “it doesn’t fit in Excel.” Although in excel unstructured data can be stored

Structured data has defining characteristics such as the name, number, addresses, date of data. Traditional "Relational Database Management Systems" RDBMS are used to process structured data. Since this data has structured property, easy to analyze and sort, a common example of RDMBS is Structured Query Language (SQL). This data is used by search engines to come up with meaningful results. An example is given by Khutarniuk (2015) as, even though a human can give meaning the "giraffelivesinsavannah", the machine does not make up any meaning with it.[4] So the data should be structured into meaningful words for a machine to come up with results. With the development of the machine reading technologies, Google Search results can understand and divide these words and "did you mean?" questions comes less and less and directly "giraffes lives in savannah" query result is given to the user.

Semi-structured data are structured data but are hard to analyze with structured data models like, SQL. If a document is sorted through metadata tags, the unstructured properties of it become analyzable through this metadata even though the file itself is unstructured data.[5] For example, when a public agency wants to employ an officer, by collecting resumes over time, each candidate would give some hints and metadata tagging about them, such as previous work experiences, what was their responsibility, for how long they worked there etc. It is suggested that, while staffing, not only skill-based search is enough but it would be better to set up a network of intelligence that allows better candidates by for example comparing who worked for the same project or company over the time, so by sorting out this network of candidates, the human resources agencies can staff comparatively (Loshin, 2005).

[4] Khutarniuk, Y (2015) A Beginners Guide to Structured Data, Retrieved 12.06.2017 from https://www.link-assistant.com/news/structured-data.html?fb_comment_id=1469902979742360_1470992599633398#f23820b65dddad4

[5] Rouse, M (n.d) Retrieved 03.04.2018 from: https://whatis.techtarget.com/definition/semi-structured-data

Unstructured data is all data without a predefined or organized data model. Health records of 70 million people is an example of unstructured data. Uncoincidentally, most of the big data analytics are possible through unstructured data. Audio files, videos, pdf files are some examples. Holzinger et al., (2013) estimates that 70%-80% of data is unstructured in organizations. While imposing open data policies, one of the expected things from public agencies is to make their data machine-readable so that an analysis is possible within a structured framework. Otherwise they would fall into unstructured data property.

2.5 Data Analytics for Big Data

Data analytics refers to "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions" (Davenport & Harris, 2017, p. 7). Traditional structured data analytics tools for data are CRM (Customer Relationship Management) and ERP (Enterprise Resource Planning) can only work on structured data. Since big data differs from traditional data, big data analytics needs more developed technical architectures to work. New methods such as Hadoop and MapReduce are used on more diverse and high volume data (Braun, 2015). An organization's big data program needs to meet the requirements of collecting, managing and analyzing potentially huge volumes of disparate data, at the right speed, and within the right time frame.

The new technologies such as NoSql and Apache Hadoop are used for processing big data sets via database management systems. Cassandra is giving way to actualize the possibilities, and both of these systems are not older than a decade. These database management tools can handle both the unstructured and semi-structured data.
NoSQL is a broad class of database management tools. Unlike its predecessor SQL, NoSQL tools allows high-performance, agile processing of information at massive scale (Lo, 2018). NoSQL databases are unstructured in nature, does not only uses tables, uses distributed databases so that an unstructured data can be stored on many

servers. NoSQL can work better with getting enhanced hardware, high volume data can easily be processed with more computing power. Some of the database solutions using NoSQL are HBase, Cassandra, and MongoDB.

Hadoop is an open source software framework for storing data and running applications on clusters of commodity hardware. Hadoop is a distributed file system, which allows data to be stored on many devices. (Marr, 2015) Map Reduce is a prominent part of Hadoop, which basically maps the unstructured data and reduces it to operable scale. For example, to retire all +67 years of male academicians from a database of CV files in Council of Higher Education Catalogue, MapReduce can be used. Hadoop Common property is used to read from the database of the user and YARN manages the resources of the system, which data are stored in.

The concept **BOLD,** Big, Open and Linked Data actually gathers the whole idea together and it represents the way and which a government can share, collect, and reimagine a dataset. The discussion on the grand narrative about BOLD is that it can be part of both government transparency, and on the other hand it can also be a matter of privacy discussion (Jansen & Van den Hoven, 2015). The "linked" part of the data is "connecting the structured and machine readable data that can be semantically queried" (Bizer et al., 2009) and is a step further on the big and open data discussion.

2.6 Big Data and Cloud Computing

One of the main concerns for big data applications in public administration is that, rather than having integrated data storage systems, there are offline or rather less connected data silos used by the governments. This leads to a problem in big data management, since one of the most useful implementations of big data in public sector is to have integrated and connected systems. Big data is therefore directly linked with cloud computing.

Cloud computing, which represents the interconnectedness of the devices, huge influx of data stored in shared hosts and connected servers around the world, rather than common physical IT architectures, was the response to an increase in velocity and volume of data. Cloud technologies' most important advantage is its real-time data fluctuation capacity. Like the importance of data compression formats cloud computing is an infrastructure for developing big data capacities. United States Federal Chief Information Officer (CIO) develops policy on cloud-computing technologies for development, test, integration, pilots and production. From an organizational point of view, cloud computing helps to succeed and fail quickly so that they can learn a lesson and act accordingly. Public administration organizations also benefit from this infrastructure through gaining spatial advantage, since there will be no need of physical data centers for specific centers if the organizational structure is appropriate enough to use cloud computing via distant access.

The cloud computing creates the hot data and cold data concepts. Hot data is frequently accessed stored data and cold data is rarely accessed data type that requires slower storage read and write speed. For this reason, hot data is closely related with cloud storage which presents solid state disks instead of hard disks, which helps increased data access and storage speed. For example disaster reporting streaming data by crowdsource is an example of hot data, a fast analysis and storage is needed to benefit from data. However, inactive data that needs rare access and usage such as storing national archives of a country is an example of cold data. A cold data also can be stored in cloud but, with lower maintenance and cheaper storage prices are needed for storing cold data.

2.7 Big Data and the Internet of Things

Gartner estimates that there will be 26 billion interconnected devices by 2020. Internet of Things is a developing concept as McKinsey's IoT report from 2018 expects a $11.1 trillion value in 2025. It is basically defined as connecting the devices to the internet

which are until this age generally accepted as offline, such as domestic appliances, heavy-duty machines, vehicles, health gadgets and etc. For example autonomous trucks, by using censors, which is the main hardware of many IoT devices, is expected to change logistics businesses, both as employment and through improving efficiency of transportation. Nowadays, investments are being made on the IoT, such as public health, public security, resource management, disaster management and traffic control. The relationship of the IoT and big data is that, the Internet of things are the hardware and their constant real-time data output is what makes up of the unstructured big-data.

The earliest implementations of the IoT began to emerge in public administration with the implementation of the "smart cities" concept, which primarily envisages the improvement of public services by employing various technologies. For example, the use of smart parking areas, i.e. the use of open parking spaces in certain areas of the city by means of sensors to transfer data from the interface to the drivers, or the sensors are applied as the improvement of waste management by transferring the occupancy rates of solar energy garbage cans to the center. In addition to this, in terms of security policy, noise rates of the high sensitivity sound sensors, detects suspicious increases and transmits the data to the center. The crime rate is decreased in some pilot implementations in the USA due to the security forces being transferred to the scene immediately.

High volume, high velocity data transferred in real-time to track public transportation, tourism, environment (water supply accordingly), security etc. are other examples. By analyzing them, high value information is achieved. For example, newest smart watches are capable of measuring body temperature and heartbeat through censors, if opt-in, in the U.S. These data flow to the family doctor for tracking of an emergency. Here is when technology and government clashes. The EU privacy laws are defined, child-tracking smart watches are trending around the world. By using sensors, a child's smartwatch can be snooped by the parents both for geospatial data and audio data. According to EU privacy laws, this affects the privacy law and the German

Government demands from parents to destroy these watches. Although works hand in hand with big data technologies, except the local government policies, the IoT usage in the strategic plans is not as much as big data hype. This might be due to IoT being a technology but big data is an overreaching method that is applicable in diverse areas.

2.8 Big Data and Open Data

As Yu and Robinson (2012) pointed out, the earliest appearance of the term "open data" in a policy context comes from science policy in the 1970s: When international partners helped NASA to operate the ground control stations for American satellites, the operative international agreements required those partners to adopt an open data policy to work together.

The open data concept is hand in hand with big data when it comes to the big data analytics, since not only data scientists can use these data but also it can be used for transparency purposes and public researchers can use the data to discuss and understand these raw, machine readable sets of collected data to use it in their studies. The open data initiatives around the world for the last decade began to be a part of governments primary digital policies. Data.gov opened in 2009 and United States open data initiative is in process since the 2013. UK uses the open data platform since September 2009. India Open Data Platform is in use since 2012, these platforms are encouraged by the World Bank with Open Data Initiative.

The emerging tendency for opening huge volumes of data resulted by the discussion of whether it is possible to result in useful and practical outcomes for governments. For example opening government data to the public, to help finding new results about the prevalence of the certain diseases, or using open datasets to detect the microtransactions for avoiding frauds from the users are gaining importance. On such condition, the datasets do not only represent countless documents on the shelves, but also begin to represent the possible outcomes of discussing and understanding more

open and diversified consequences. This leads to the point of leading data analysis. The economists expect that in the next 10 years the American Job Market would be in need of vast amount of new data analysts. UN E-Government Survey 2018 does not differentiate the open government data policies and big data analytics for the sustainable development, plans for institutional performance and effectiveness.

2.9 An Overview of Big Data Related Studies in the Literature

An academic literature review is essential for understanding the role of big data in public administration because it is crucial to understand both practitioners and academics point of view on the subject and to get a general outlook of the positive and negative outcomes and main issues. Fredriksson Mubarak, Tuohimaa, & Zhan (2017) made a systematic review of big data usage in public administration and identified 156 articles in different databases. The research suggests that since 2011 there is an exponential growth of big data research in public administration context (Fredriksson et al. 2017). This literature review below will try to understand and create a framework for big data use in public administration.

While the e-government scientific output is broadening day by day, Mayer-Schönberger (2013) suggests that e-government is transforming into information government by surpassing the technical abilities and bureaucratic reduction and such, but also adds a crucial part of using data to develop into a more proactive and open state by developing e-democracy strategies. For this reason, a subset of e-government research, big data usage in public administration is rooted in information technologies, a counterpart to the subject as open data research, and public policy research together. For example Yıldız (2007: 648) states that the Urban Information Systems (URBIS) project, which was conducted from 1973 to 1978 at the University of California, Irvine, by a multidisciplinary team, became one of the gamechangers in e-government because this was the “first large, systematic, empirical study to focus specifically on policy and outcomes related to computer use in complex service organizations”. Since

the popularization of the word big data and increasing usage of ICT in government, big data gradually began to enter into the public administration realm. When taken as a whole, big data is not the technology itself, since clusters of varying data does not signify a technology but to analyzing, getting meaningful results from those datasets what makes big data a hype word. Processing big data is a highly complex task, which includes many different formats.

Business, academy and government sectors are enjoying the benefits of big data revolution (Kim et al., 2014). However, business and government differentiate on the area of profit making (Kim et al., 2014). Although the New Public Management principles lead public administration closer to business area, achieving sustainable development, achieving peace, promoting welfare and growth are still the prominent policies of the governments rather the private sectors profit model. Kim (2014) differentiates the different values of data for the business and government sectors. The value of data in government sector still stays in data silos, which are not connected and does not have interoperability. For this reason value creation is harder than compared to the business model, along with traditional data production machine generated data also added for volume. Governments have to break away from data silos to actualize the value of their data. Tight regulations prevent the governments to take advantage of data that they possess. Compared to businesses profit driven analytics, Kim (2014) suggests government will actualize value of data via providing sustainable solutions such as enhancing government, transparency and balancing communities.

The early 2010's saw an increase in the importance of big data usage in many areas starting first from the business models. Although what would become many data analytics tools infrastructure, the MapReduce programming model dates back to 2004. Hadoop one of the most common analytical devices to analyze big data, first initial release was on 2011. The general interest in using such datasets increased gradually over time.

While researching for the review, there were common themes of definition of big data, mostly about its properties in almost all the papers. Another common subjects were gradual policy recommendation for digital transformation (Janowski, 2015) possible advantages and disadvantages in public administration (Chen & Zhang, 2014), case studies, future prospects. Although case studies are on a central government, there were many articles about ICT & big data usage in smart cities. (Batty, 2013; Townsend, 2013) which has an increasing frequency over time.

As Mergel (2017, p. 928) pointed out, "Big data in the public sector is context specific and needs to be meaningfully combined with administratively collected data to have value in improving public programs". There are many different areas and policies that big data can be used in public administration but as any computer scientist would say, there is no one size fits all policy for any government agency or in any specific context. There are some early examples of studies mentioning big data in public administration before academic interest was raised; some of the research institutions were the first comers on the area. In McKinsey Global Institute's report "Big Data: The next frontier for Innovation and Productivity" (2012), although public administration was not the main concern of the research paper, importance of big data for public sector was underlined as follows: "In advanced economies facing unprecedented pressures to improve productivity, our analysis of public administration in Europe demonstrates that big data levers can make a significant contribution to this effort around the world" (MGI, 2012: 63). Consecutively, in 2014 Digital Agenda for Europe, the importance of data driven policies was highlighted (Digital agenda for Europe, 2014: 5). For adapting better big data policies, there are big data maturity assessment programs, which can be used to define where a public agency in adapting the big data application is (Braun, 2015).

The 2012 report entitled "Demystifying Big Data: Practical Guide to Transforming Business of Government" prepared by TechAmerica Foundation is one of the earliest examples on the subject, which takes big data in consideration extensively from a

public administration point of view. The report, similar to many academic papers, first defines what big data is, V's of big data, mission, value and gives results about it. The first important point about the article is that, it diversifies where government can use big data such as, healthcare, transportation, education, fraud detection in healthcare and tax collection, cyber security, weather and employment policies (Mills et al., 2012). The report's suggestion about Medicare fraud detection began to be used by the US Government. Similarly, the report provides specific case studies about the subject from NASA Space Center usage and National Archive and Records Administration (NARA), and NASA Space Center suggestion about using historical data for educational purposes show a very early stage of big data maturity adoption (Mills et al., 2012). The report also gives an early analysis of big data technologies such as defining streaming data and map reduce framework, and data warehousing.

Fernandes et al. (2012) discuss the usage of big data to improve efficiency and get better outcomes for the governments, particularly in health services. The authors claim that "Government organizations may be the biggest beneficiary of Big Data solutions." If as a huge stacker of the data the government can implement big data policies, they can detect of fraud and abuse patterns, identify best practices for safer and more efficient care delivery, and obtain better resulting surveillance. While there are advancements related to big data in health sector, one of the most common concerns about big data comes into the scene as the privacy of the citizens and patients. Furthermore, when highly developed data analysis is done from anonymized data, a patient can be reached through certain measures.

For using big data in the public sector, Maciejewski (2017) defines three general advantages of big data: increasing accuracy in decision making, improving performance through analytics, and reducing costs. A thoroughly analysis of big data would get better results in data driven decision-making such as an analysis of certain big geospatial data to analyze city crime rates (Pries & Dunnigan, 2014; El-Darwiche, 2014). Maciejewski (2017) points out that there are three possible areas to apply big

data: first public supervision to identify irregularities and being proactive, second for public regulation, for social conduct and shaping social relations of permits and prohibitions, and third for public service delivery to provide services or products such as infrastructure.

Throughout literature there are varying analysis and abstractions from different perspectives of scholars, but Maciejewski (2017, p.123) underlines a different one: "Big data methods can uncover knowledge that was previously impossible to reveal. In turn, this new knowledge allows new tasks (previously impossible or even unimaginable) to be successfully carried out."

Some applications of big data expect to get results from the particular documents hoarded by the governments. For example, by using historical data of child abuse, IBM reached 90% accuracy in re-occuring of child abuse or neglect cases in certain areas.[6] A historical data analysis caused predictive results. When real-time data considered, there are streams of data coming from censors and devices, which is a property of the unstructured data. For example, constant streaming data from the CCTV's are used by Chinese government to catch suspicious people. [7]

Unstructured data proves opportunities, however various government policies about big data stays on the level of structured data (Kim, 2016) which means that governments have a high variety of data but they do not know how to work with them or how to benefit from them. This also overlaps with McKinsey report (2011) the report suggests that, although governments have vast amount of data, it is the hardest one to use because of coordination problems, different data centers, offline/inept data and lack of machine readability. Big data sets do not necessarily create meaning as

[6]Reilly, N.S (2016) How Big Data Can Help Save Endangered Kids retrieved: 12.07.2018 from https://nypost.com/2016/10/16/how-big-data-can-help-save-endangered-kids/

[7] Joyce Liu, In Your Face China's All Seeing State, retrieved: 17.07.2018 https://www.bbc.com/news/av/world-asia-china-42248056/in-your-face-china-s-all-seeing-state

Mergel (2017) pointed out, computer scientists finds correlation between some datasets but this does not necessarily find causation with the existing policies. Correlation does not mean causation is a common quotation among statisticians, big data helps to find this causation mostly without human interference. However, public agencies have to have awareness insights on data analytics to benefit from them.

Mergel (2017) underlines that until 2010's data was only collected by individuals only for administrative reasons. However, in our era, it is possible to analyze many financial transactions, flows, statements on the Internet, and credit card payment information etc, as a combined data to demonstrate the characteristics of the citizens. Using this analysis for a specific group, such as a neighborhood with higher handicapped population detected via analytics, it is possible to build better roads and present municipality social services specific to those citizens. The constructive effects of the data are not limited to government to citizen (G2C) as summarized with the previous example, but also useful for government to government (G2G) benefits, too. In both cases, vertical (G2C) & horizontal (G2G) data share is beneficial for the interests of individuals and institutions. (Layne & Lee, 2001)

An example of data sharing from the UK represents a communication failure. Health Services and the Police Department of London once wanted to share data to be able to work more efficiently but uncooperative approaches of the departments led this initiative to turn into a failed policy.

Mergel (2017), defines the differentiating opportunities and challenges about big data across management, political science, public policy, information and technology management and computational social sciences . One of the rare challenges in the Public Policy area when compared to other disciplines is "no breakthroughs in quasi-experimental research designs". Public administration discipline's applications has direct impact on the society. The IT personnel in this area do not find it specifically

meaningful to create experimental research designs but rather execute the policy to see whether it is working or not.

A good example from Turkey is the Şahinbey Municipality of Gaziantep. The Municipality wanted show its existence in e-government area as creating an open data portal. This portal let the municipality to win the first prize in 2015, e-TR(e-Türkiye) Awards. At the moment, the webpage has not been in use since 2016. In other words, after its initial release to the public in 2015 the web site functioned only for one year. This shows that without having an actual aim other than winning an award or without realizing the demands of the citizens, top down approaches in some initiatives may lead failure in the active use of the systems. The policymaking process at this part is important. Another example from Turkey is genealogy query, which went live in National E-Government portal on February 2018. With the help of this new tool, Turkish citizens could have seen their ancestors registered in the Census Bureau. For the first two weeks, whole E-Government Turkey servers 404'd, which shows both infrastructural problems and also a possibility of e-government policies to become a fad (Giest, 2017). Since, everybody already used the system with a hasty manner, the data stays inept on the system.

A highly structured data high in volume could lead an interactive solution about kinship, geospatial analysis and roots of Turkish citizens forgotten due to serving completely structured data. A proactive government strategy could be to analyze the spatial history of the citizens and compare it with certain diseases, because the Turkish government has a detailed open data portal for healthcare providers to see where their patients are coming from, their background, common illnesses on the area etc. Heeks (2003) reports 35% of e-government projects are total failures, 50% are partial failures and only 15% are successes, for this reason this can be said that what happened to genealogy query system in Turkey is not limited to the country.

In the literature, there are also discussions on big data specifically intended for Turkish public administration, such as a possible legislative framework for big and open data (Kaya, 2017: 193), public policy adoption possibilities of big data (Altun & Şahin, 2017), or analysis of strategic documents on big data (Köseoğlu, Ö, 2017).

2.10 Opportunities Related to Big Data Use

There are several opportunities expecting the practitioners, academics, and also IT personnel in the practice of big data in public administration.

2.10.1 Economical Benefits

Although governments are using big amounts of data for decades in policy analysis and to improve management (Pirog, 2015), with the transformation of ICT and the other revealed properties (V's), actual data would lead researchers to have much more comprehensive analysis to improve on these areas. Big data is seen as the next step of an ongoing process of increasing efficiency and leveraging the competitive advantage, reducing risks, costs and operational losses while improving customer (citizen) experience (Kitchin 2017, p. 118).

Benefiting from digital economy is one of the most common expectations of big data usage by governments, because having one of the greatest datasets, governments would like to benefit from mobile data, IoT, social media inflow by incorporating them to create economic efficiency and market possibilities. Since data-sharing between private and public organizations, such as health sector, is possible, it is suggested that there are many benefits for governments along with e-commerce (Munné, 2016), as well as probable optimization in the labor market (Höchtl, 2016).

There are varying degrees of opportunities for government services. Zheng (2017) summarizes the big data adoption to build a more efficient (Bekkers & Homburg 2007)

and capable government for better public service. Milakovich (2012) improves upon efficiency and adds effectiveness and transparency to big data applications. The reason for transparency of big data is to create big and open data policies that are linked to each other in developmental stages. UN e-Government Survey (2018) suggests that by using big data, efficiency and effectiveness measures increase, because big data would lead decision-makers to receive more precise, vast and predictive measures, since they are trying to overcome uncertainties. Kitchin (2014, p. 154) summarizes opportunities of big data for public administration under two titles, (i) first "improvements to public management and cost saving through enhanced operational efficiency, reduction in the cost of errors and fraud, and increase in taxes with a better-informed citizenry, and (ii) second "increasing state security and the tackling of crime" (154). John Manzoni, Chief Executive of the UK Civil Service and Permanent Secretary of the Cabinet Office summarizes the subject in a clearer manner and defines the usage of big data compared to Kitchin. The benefits of data usage for UK government have been improving the citizen experience by making government more efficient eveloping ICT affects governments through various ways, increasing efficiency also affects budgetary constraints. Since public administration becomes more effective because of the global economic condition of developing countries, governments have to improve their services through ICTs. Chen and Zhang (2014) suggest big data as a potential booster of budget and resource, as it can be a source for decreasing budget deficits and reduce national debts.[8] For example return on investment levels are high on IT investments. EU Big Data Report suggests up to €100 billion of return in short term. (Chen & Mao, 2014) Some other estimates gives potential savings up to € 300 billion.

Also, opening datasets on various sectors such as health, education and tourism may lead businesses, citizens and governments a possible usage different from their original purpose of collection. This includes creating services that are more responsive to the

[8] https://www.sciencedirect.com/science/article/pii/S0020025514000346?via%3Dihub#!

needs of citizens as well as government itself being more efficient in its response (Bekkers & Homburg, 2007)

IMF produced a report on the factors analyzing possible efficiency outputs for countries, such as using monetary policy analysis via using big data, for effective forecasting and financial stability. Furthermore, structural transformation of statistical agencies for an efficient data collection and analysis is made possible. As a whole, each new service comes with its efficiency possibilities.

2.10.2 Decision-making

Availability of big data analytics and integration brings some other benefits about decision making both in public and private sectors (Desouza & Jacob, 2017) by providing predictive analysis for mission outcomes and improved operational intelligence.[9] Processing high volumes of data helps organizations to have better informed decisions in a shorter time period (LaValle et al., 2011). Since, it helps one of the main proponents of management and planning, big data helps decision making through linking various datasets(Horvath, 2016). These new analysis tools help specialists to have better control on decision-making process (Gänßlen & Losbichler, 2014, p. 2).

Desouza (2017) suggests that before the influx of the real-time data, the datasets were slow and stationary being updating annually, at best monthly, but the developing technology provides constant flowing of real-time data from various sources, for this reason the analysis on decision-making would improve through "ongoing conversation" for policy research. By adding more datasets over time a possibility of incremental improvement over older designs are expected (Cook, 2014).

[9] Bridgewater, A (Apr 5, 2016) Why You Need Operational Intelligence for Big Data retrieved 11.06.2018 from: https://www.splunk.com/pdfs/why-you-need-improved-operational-intelligence-for-big-data.pdf

Similar to Janssen et al.'s (2017) process on decision-making, the policy process affecting the big data decision-making quality is explained by Bizer et al. (2012) by six steps, data capturing, data storage, data searching, data sharing, data analysis, and data visualization. It should be noted that, this process also shows similarities with big data maturity models (BDMM)s that draws a general outlook of organizations' development of the process. Although governments are constantly collecting data, reaching conclusions via analytics, they have never presented realized potential of datasets which would be very much helpful to diverse and deepen understanding of the policy actors. Another advantage is sharing data, meaning the interoperability between different public organizations, which gives an opportunity to the organizations to benefit from wider datasets of various sources. This process would result in a better decision-making process, ending mostly with executive side of big data analytics, which is the visualization of data.

Evidence-based policy making is improved via data-driven insights by using this technology. Although governments fall behind compared to businesses about data-driven analytics, they're still the biggest data collecting organizations. Höchtl (2016) discusses that big data may benefit from evidence-based policymaking process. Taken from the theoretically-rooted perspective, O'Reilly (1982) suggests that quality of data affects the quality of decision making. Henceforth, using big data in public policymaking increasingly influenced by research and data-based intelligence gathering of government agencies (Höchtl, 2016, p. 150). Raghuanthan (1999) adds another perspective on O'Reilly's argument and says "the decision quality improves with higher information quality for decision-makers with accurate knowledge of the relationships."(p. 276), a more current comment comes from Smith et al. (2009) by noting that "the success of evidence-based policymaking depends on the quality of the evidence that underlies it." (p. 59). Big data at this point reaches its goal by giving an opportunity to bear various relations between variables, breaks the common barriers of causality, and gives real-time insights on what is happening through the analysis (Kitchin, 2015). It is suggested that, by doing so previously a dataset might have been

called as "noise" a meaningless data package can be beneficial under different circumstances. An outlier example is data gathered from Greenland's ice sheet representing the historical lead data in the atmosphere is relatable with devaluation of Roman Empire's money after Ceasar's death year by year, because silver mining leads high amount of lead exhaust from mines to the atmosphere and this data buried under the ice throughout the years.[10] This analysis, could not have been possible without bringing different 25,000 datasets collected since 1999 and pattern discovery. As Hammer et al. (2017) suggests, "In some cases, big data can allow policy analysis to move beyond aggregates and look at what lies beneath to better inform policy responses." (p. 16) Like this example, what has not been given a value, a meaning now can be meaningful both for policy analysts, scientists and practitioners. However, it should be noted that, there are studies which proposes about evidence based policy making that scientific results do not always turn into better policies (Sanderson, 2002; Simmons, 2015).

2.10.3 Crowdsourcing and Public Talent

Through ICT usage, decision-making and innovation can be harnessed by citizens which can result in new opportunities by using "wisdom of people"(Giest, S., 2016; Janssen et al. 2017). For example, in the US, citizens can engage through using online petitions via whitehouse.gov. If petitions signed by 100,000 citizens, the petition is answered directly by the President's Office within 60 days.

This increasing citizen engagement also effects what is called as e-democracy, because the concept refers to using ICT to involve citizens in political debate and further for policy-making process. This process is seen as an answer to reducing democratic deficit (Macintosh, et al., 2005) because of apathy against political debate, and refuting

[10] For further inquiry on this subject: https://www.theatlantic.com/science/archive/2018/05/scientists-reclaim-the-long-lost-economic-history-of-rome/560339/

to voting leads lowered citizen input into policy-making process. Giest (2017) suggests similarly through adding citizens to this process via opening public data, citizens can contribute for development.

Morabito (2015) takes the benefit of using public talent from new public management context and, a possible usage of citizen engagement in data analytics through crowdsourcing explained citizens as consumer becoming prosumers. Desouza (2014) suggests crowdsourcing can be useful for labor-intensive applications, such as picture marking, language translation, and speech recognition free or voluntarily, however there is also possible ways of engaging citizens to labor market via freelancing them by governments.

For example, Google is using hard to machine readable book scanning for its captcha technology. By doing so, significant amounts of books are digitized. Governments have massive amounts of offline data, which are scannable and uploadable to a government cloud. Some businesses, such as Appen and Lionbridge uses crowdsourcing for similar jobs on freelance markets in developing countries, such as improving Facebook, Yandex and Bing's search results via using people skills, paying them to assess the results on their mobile phones. Turkey has one of the biggest mobile markets in the world. If government would like to digitize its content, under normal circumstances immense amounts of work hours for this documents to be digitized can be achieved in short time via using public resources.

Another example of crowdsourcing is population's usage in disaster reporting, UN Global Pulse Report (2012) and Norheim-Hagtun and Meier (2010) give example of using crowdsourcing following the Haiti Earthquake, where centralized text messaging system was used to allow cell-phone users to report on people trapped under rubble. Similarly, crowdsourcing is used for reporting problems in local governments.

2.10.4 Service Delivery

West (2012) states that "big data make it possible to study different areas for insights regarding student performance, health care, energy efficiency, and public sector performance."(p. 4) Increasing personalized treatment and personalized education by using artificial intelligence provides new advantages via data analytics. (UN, 2018) It is also important to understand that the patterns and correlations increase effectiveness and productivity and creates better tailored and targeted services, policies and programs. (Höchtl, 2016).

Education is one of these areas. It is the duty of the government to provide the high-quality education facilities to the children, big data can help government to screen the educational needs for the students instantly. West (2015) and Zainal and Hussin (2016) suggest that, tailoring education to the individual levels of each student`s learning style and opening online classes according to students needs is one of the greatest benefits of technology, in which big data can be very helpful to teachers in terms of personalizing the learning process. For example "development of computerized learning modules enables assessment of students in systematic, real-time ways."(West, 2015) By using data mining and data analytics software it is possible to generate feedback and assess the scholar aptitude of students via using anonymized records.
In terms of health policies, the nation-wide collected representative longitudinal data on over 3,000 individual including a genetic profile with various biological samples, different social, health, and biological variables for US National Social Life, Health, and Aging Project. The inclusion of bio measures with traditional policy variables and controls can, at a bare minimum, illuminate possible health problems in the future with better diagnostics, therapy and personalized medication as being a trending subject of medicine (Pirog, 2014).

When, local governments are taken into consideration there are various advantages of big data, especially in the use of Internet of Things, which is popularized with the

Smart City concept. Grid planning is a well-known example of it. By analyzing data in Smart Grid, the regions can be identified whether they have excessive high electrical load or power outage frequencies, transmission lines with high failure possibility (Chen & Mao, 2014). Besides grid planning, infrastructure, energy management, waste management, public transportation are other well-known beneficial areas of the big data exploitation.

2.11 Main Challenges Related to Big Data Applications

There are plethora of advantages of big data applications in public organizations, as suggested by the academic literature, various policymaking actors and by the governments applications. On the other hand, there is not a deus ex machina like solution which can solve everything with a finger snap, since because each decision comes with its own problems. Problems about big data are not limited to public sector and applicable to any benefactor. This section focuses on challenges of big data applications from public administrations` point of view.

The data production is growing exponentially day-by-day, and "past, unknown or private data becoming revealed to different interests, and decision-making within government and business is becoming more data-driven, evidence-based and technocratic" (Kitchin, 2017, p. 206). This inferring leads the discussion of challenges in big data applications in Public Administration.

2.11.1. Technical Challenges

There are technical challenges like preservation of data, lack of skilled data analysts in organizations, and then there are direct public administration challenges both as theory and application (Maciejewski, 2017). Moreover different maturity measures of countries depending on economic infrastructure to cultural adoption and digital divide (Klievink, 2016; Uçkan, 2003)) cause challenges. Almost all of the public

administration datasets are taken from the public, for this reason legal regulations and privacy is another issue. Mayer-Schönberger and Cukier (2013) mentioned another two examples which they defined as "dictatorship of data" and "governing the (data) barons". Lourenco et al.. (2017) suggest organizational and institutional challenges of big data applications in Public Administration. The list by Lourenco et al. (2017) privacy, organizational culture, civil collaboration, legacy system[11], recruitment of talents are part organizational challenge; data ownership, data security policies, civil liberties, equality, civil collaboration and biased data are institutional challenges. In this section a general outline of challenges of big data usage in public administration will be explained.

Applications of geospatial data used for long time to examine the impacts of policy areas (Pirog, 2014) However increasing usage of mobile devices with Global Positioning System (GPS) geospatial privacy becomes a concern. Strava, a GPS enabled sport app keeps track of its users on a global heat map that they designed. On January 2018, there were news about the about the open global heat map. Since sport is a crucial part of armies, secret military bases in the middle of deserts or underground bases and the names of the soldiers in them using the app mostly of US origin appeared. Although the users are giving their personal data with their consent, this creates a national security problem through breaching governmental security without even realizing the problems it would cost.[12] Michael and Miller predicted this in 2013,

> …as big multimedia datasets become commonplace, the boundaries between public and private space will blur. Emerging online apps will not only enable users to upload video via mobile social networking but will soon incorporate

[11] Legacy system is using outdated computer technologies due to structural usage of them for a long time. For example, PENTAGON is still using floppy disks to maintain old Nuclear Missile Facilities.

[12] Accessed 18.06.2018 http://digg.com/2018/strava-secret-military-bases-soldier-names the maps seems to be in deserts. Global heatmap is reachable through strava's page: https://www.strava.com/heatmap#7.00/-120.90000/38.36000/hot/all

> wearable devices in the form of a digital watch or glasses to allow for continuous audiovisual capture. People will essentially become a camera (Michael, K., & Miller, K., 2013, p. 23).

These kinds of problems are rooted in technical structure of big data analytics. Analysts do not conduct well-designed appropriateness analysis but rather gives examples without a code of ethics it is possible to breach privacy (Zuiderwijk & Janssen, 2012) For this reason it is extremely important to create proper privacy policies for both citizens, governments and businesses and their interoperability works through Government to Business (G2B), Government to Government (G2G), Government to Citizen (G2C). US Big Data report also underlines this by stating, "Maintaining our privacy values by protecting personal information in the marketplace, both in the United States and through interoperable global privacy frameworks.

2.11.2 Privacy

Privacy is the first and recurring concern from data analysts (IBM, 2014) academics (Shah, 2013; Höchtl, 2016; Das & Joseph, 2017) and policymakers (EU GDPR, 2017). Unfortunately, the literature offers few insights for public managers and policymakers on how to mitigate the potential privacy issues around big data efforts (Alexapoulos et al.., 2016). Both UK, USA, Australian governments big data action plans include and emphasize the importance of the privacy in their data policies. One of the most common issue about privacy is the trust of citizens to the government. There are many examples of sentences like "The public trust in government agencies and systems needs to be maintained." on the privacy section of the country big data policy reports and white papers. It is suggested that, states should be open to the citizens while applying the policies so that a double sided trust may appear. Makowski (2016) in his spectacularly named article: "From Weber to the Web… Can ICT Reduce Bureaucratic Corruption?" Suggests some examples about building trust and

accountability through crowdsourcing and using open government. Based on the open data issued by the governments in EU, citizens created crowdsourced systems to check government expenditures. Transparency International Hungary and K-Monitor, in Hungary funded redflags.eu portal to analyze announcements of public procurements. The portal is a real-time working portal that can be used by any citizen to analyze government expenditure, this system leads citizens to have a vigilant outlook to check government. The low open data quality prevents better public participation, EU reports show low quality open data, and it is not available to the public enough. This problem persists both in public and private organizations.

The Open data policies would lead a double-sided trust through democracy for citizens and businesses. Controlling big data may lead monopoly of the analysis by the governments. Macintosh (2004) described this as "The commonly held idea about policymaking is it is transferred to the official actors such as state, but the e-government policies making e-democracy a part of the policymaking process."

European Union enforced the data protection policies earlier than its counterparts by first declaring through enacting the 2000's Lisbon Treaty, the Charter of Fundamental Rights of the European Union in 2009. 8th article openly declares as:

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

It should be reminded that since EU is consisting of different constitutional entities the application of the law sees different conditions. For example, in 2016 the State of

Emergency in France provided by using IBM's i2b predictive analytical tools, looser approach to the privacy taken for counter-terrorism policies, through wiretapping and electronic surveillance (Kubler, 2017). For this reason it should be kept mind that this privacy policy action was not as effective as expected.

Furthermore, while developing data protection as a fundamental right, privacy, autonomy, transparency and nondiscrimination should be main concerns (McDermott, 2010) but under the state of emergency in France most of the raids were made to the neighborhoods with high population of Muslim and Black citizens, big data is open to profiling. Lyon (2002) also points out that this sort of profiling has hidden danger of perpetuating discrimination and assumptions about certain strands of populations. US Big Data report's one of the 5 suggestions was about challenges is: "Big Data and Discrimination: Preventing new modes of discrimination that some uses of big data may enable." McCue (2014) in his analysis of big data usage in crime prevention suggests that for this reason data accuracy, reliability and validity is extremely important to stop profiling before even the policy process begins.

European Union developed a law related to McCue's suggestions and EU General Data Protection Regulation created on 14 April 2016 to become enforced in 24 May 2018. In a close time to enforcement of the regulations all web pages using single user information in any means renewed their privacy policies, especially all of the web pages using cookies in their pages asked its users to allow its cookies by various means, sometimes as "if you continue to use this page you are allowing our new cookies policies" or sometimes asked before using the webpage. Buttarelli (2016) suggests that the law has two two strategic consequences. First it sets up a platform for partnership, which means that since other countries are also working in accordance with the EU countries, this law would have a "multiplier effect" on other governments and second it affects businesses to create a privacy policy which would cause a globalized standardization of the data protection.

The importance of privacy is also mentioned in the earlier in Lisbon Treaty, with the enforcement of the new law it is not an extra or a supranational regulation but rather a highly effective law. This would lead accountability of any entity using data. Article 59 of EU General Data Protection Regulation states as:

> Each supervisory authority shall draw up an annual report on its activities, which may include a list of types of infringement notified and types of measures taken in accordance with Article 58(2). Those reports shall be transmitted to the national parliament, the government and other authorities as designated by Member State law. They shall be made available to the public, to the Commission and to the Board.

The governments are responsible to open their data especially under the infringement of the data owners' rights, the regulation gives rights to citizens to use European Human Rights Court under any infringement since Data Protection is a human right defined by the EU via Lisbon Treaty (2007).

One of the most important thing about it is that the right to be forgotten (Article no.17) is given to the citizens. If the data of citizen is not useful for the reason it is taken, if the subject removes consent, data is processed unlawfully, or data is collected for child services and there is no use of it anymore, the subject has right to erase the data.

Under the ongoing EU membership process (officially Accession of Turkey to the European Union) Turkey is making some structural transformations. On the tenth chapter of the accession process under the name of Information Society and Media, EU Ministry of Turkey defined a policy about updating Privacy Policies. According to this process two important things realized from privacy protection point of view. Turkish Personal Data Protection Law no. 6698 enforced and this law also made it mandatory to establish Personal Data Protection Board to enforce the law accordingly. The law has similar properties like EU GDPR. This shows that, the expectancy of EU

GDPR to have global effect is a valid argument. Another regulation published on October 28, 2017 Official Gazette of Turkey about erasing the private data. Article no.7 of the regulation is step by step recreation of the EU GDPR's "right to be forgotten" section.

On the other hand, securing privacy is not enough since the use of big data will introduce an additional layer of complexity in terms of management of information security risks. The governments have biggest data silos, thus enforcing certain measures through laws would not be enough, as Snowden/Whistleblower & Wikileaks examples showed to the society. As reported by the NIST, understanding and addressing citizens' expectations on privacy and security is critical for governments to implement big data solutions successfully. What was called as intuition is becoming more of a business intelligence unit now.

2.11.3 Data Mining and Analysis Challenges

Earliest examples of data analysis and integration through public services were underlined as expansion of the quantity and scope of factual data analysis and data mining practices. Another problem about data mining given by Seifert (2004) is that, while it can identify connections between behaviors or variables, it does not necessarily identify a causal relationship. "Correlation does not mean causation" for data analytics has always been an explanation about statistical sources. On the other hand Seifert's predictions does not consist of current big data analysis practices because through big data analytics the systems can analyze causal relationships, on the other hand there are many technological disadvantages. These challenges include but not limited to data storage, network, data integrity, interoperability, metadata creation, open access, very long-term data preservation, migration of data from one system to another, technology. From staffing point of view the most crucial example is finding a data analyst who is both qualified and also can work with public institutions.

2.11.4 Staffing

The high quality data analysts are hard to come by and expensive to hire even in Silicon Valley. A 2011 McKinsey Global Institute study on big data analytics predicted a coming shortfall of around 150,000 people with deep analytical skills, and a lack of 1.5 million business people with the know-how to put big data analytics in use(Manyika et al., 2011) For the staffing policies of a government agency, this process can be described as an external factor challenging the implementation of the policy. Mayer-Schönberger (2016) while analyzing the big data and market relation warns that, "Government needs professionals with data-analytics expertise, the "quants," if it does not want to risk market failure." Since the governments work with businesses in today`s information era Mayer-Schönberger's analysis is an important one to underline the situation by "do not call for a bureaucratic expansion lightly. But without organizational enforcement, data-rich markets will be vulnerable to a dangerous concentration of decision-making power and control." On the other hand, data analytics without human interference seems to be a problematic due to garbage in - garbage out possibilities of big data.

2.11.5 Machine Readability and Interoperability

Massive offline data is staying in state archives around the world and if any government wants to benefit from this historical data and analyze them they should be digitized and categorized for machine readability. Desozua and Jacob (2017), explains this by saying "even in cases where public organizations collect significant amounts of data, it tends to be fragmented. More precisely, public agencies often operate in silos when it comes to their information technologies there is limited, if any, interoperability among information systems used across agencies."(Desouza&Jacob, 2017, p. 6) This may lead to asymmetrical distribution of information between public organizations.

2.11.6 Dictatorship of Data

Defined by Cukier and Mayer-Schönberger dictatorship of data is the domination of the data-driven probability analysis over free-will of the citizens. Cukier & Mayer-Schönberger (2013) say that the danger of data shifts from privacy to probability. Crime prediction and advanced arrestment, or automatic mortgage loan denial may lead to data dominance over citizens behavioral decision-making processes. This may cause trusting the numbers more than actual people. Although big data is open to use for various opportunities, it is also open to abuse on different occasions. Robert McNamara, the US Secretary of State throughout the Vietnam War, gave high number of body count data on newspapers to show the success and progression of the US Military. The US Generals argued that the body count is a mere statistics when it comes to war of will(Cukier & Schonberger-Mayer, 2013, p. 254). US, ended up losing the war. The author argues greatly about data usage here, we underlined Ackoff's definition of Data to the Wisdom, similar to his statement Cukier and Mayer-Schönberger says about McNamara, "a man of intelligence but not wisdom." Furthermore Beer (2016) discusses this same point by stating:

> Big Data brings with it a force to comply and a rationality that is hard to critique or resist. This can potentially be seen to have a kind of neoliberal reasoning or rationality at its core, one based upon the use of data as the mechanism by which the model of the market may be rolled out across the social world. (Beer, D., 2016, p. 6)

On this occasion, a synthesis of two scholars shows that the introduction of the analytics enforces very high rationality which may cause a trust in data but a direct distrust to the citizens which may cause a data/digital totalitarianism as a policy.

2.11.7 Threat of Data Analysis Company Monopolization

In addition to the dictatorship of data, there is a suggestion of a related concept, data barons. This description is directly linked with an old Latin proverb "Quis custodiet ipsos custodes?"/ "Who will guard the guardians?". Considering that big data are seen as a great advantage, there are now big companies like Google and Facebook which controls clusters of data on the world as a natural result of the current technology. Similar to the Mills (1956) description of power-elites, Cukier and Mayer-Schönberger (2013) suggest that anti-trust laws need to be put into place not to let data companies enjoy limitless privileges on citizens and become a non-fictional "Skynetlike" Data Elites themselves. A very recent example is the Facebook-Cambridge Analytica controversy, which affected 87 million users in 2016, 70.6 million of which were users from US. By using an app on Facebook page called "thisisyourdigitallife" the online survey system on the surface seems to work as an academic research, but when an individual user decided to use the system, the system not only reached the user's data but also reached the personal information of the people on his/her network. By using different variables such as page likes, locations, current living city, birthday, the system creates a psychographical profile. Through using this profile, tailor-made advertisements for the users are applicable for both market behavior and also for political campaigns to skew electoral behavior. Here, Facebook, a data baron, could not properly control its own internal algorithms and this lead to a scandal.

2.11.8 Lack of Big Data Readiness in Organizations

Oxford Dictionary defines maturity as, "the state of being complete, perfect or ready"(Simpson & Weiner, 1989). Beginning from 1970's maturity models for adoption ICT in organizations are suggested. Possible lack of readiness on various areas cause the policies to fail. Different maturity steps in public organizations are one of the greatest structural challenges because there is not a fixed government wide

solution for adapting big data policies. Each organization's culture, structure, need for big data, staff, alignment varies.(Klievink, 2017) The low levels of maturity levels signifies unawareness of the value of big data, lack of proper budgeting, managerial resistance, ICT integration challenges.

2.11.9 Other Challenges

Methodological questions arise about the big data usage. Gong (2016) suggests that big data usage is a reversal of traditional research methods, because conventional method for traditional data analysis is asking the research question and then collect the data needed. On the other hand, big data analyses and comes up with unexpected networks and results where traditional methodology is no enough to suggest or infer. For this reason there is a danger of "the end of theory" according the Gong et al. (2016) and Anderson (2008). If the institution using the system does not implement the system would end up with varying results without even a proper content.

Even though compressing algorithms are developing there is still limited amount of storage when compared the information output, Cukier (2010) interviewing with The Economist describes this by explaining that there is a divergent relationship between data created and available storage, explaining that data output is increasing exponentially than increasing rate of available storage. Although developing cloud technologies surpassed The Economist's predictions about future resulted with another problem. Increasing carbon footprint of the data centers. There are distributed data servers under the North Sea, and some arctic areas (Starosielski, 2015). Higher heat means the wasted energy both for the operating centers, and for the global effect. Velkova (2016) gives some solution examples to use the wasted heat potential. According to her research, now in Helsinki, Stockholm and Paris these heat is transferred to public areas such as churches or underground transportation to heat up the areas. The energy input problem used as a decoy by industries to hide the real problem of heat output according the Velkova.

2.12 Critical Success Factors in Big Data Use

Asay (2014) summarizes the reasons for the failure of big data implications as management resistance, selecting the wrong uses, asking the wrong questions, Unanticipated problems beyond big data technology, disagreement on enterprise strategy, big data silos, and problem avoidance, as shown in the Table 1.

Table 1: Reasons of Big Data Project's Failure

Reasons of Failure	Definition
Management resistance	Managers say that they trust their guts, and real-world insight is better than hard analytics when making decisions.
Selecting the wrong uses	Companies either start with an overly ambitious project that they're not yet ready to tackle, or they attempt to solve big data problems using traditional data technologies.
Asking the wrong questions	Organizations hire high skilled data scientists without the knowledge of business for this reason output and expectations does not meet
Lacking the right skills	Too many big data projects stall or fail due to the insufficient skills of those involved.
Unanticipated problems beyond big data technology	Analyzing data is just one component of a big data project. Being able to access and process the data is critical, but that can be thwarted by such things as network congestion, training of personnel, and more.
Disagreement on enterprise strategy	Big data projects succeed when they're not really isolated "projects" at all but rather core to how a company uses its data. The problem is exacerbated if different groups value cloud or other strategic priorities more highly than big data.

Table 1: Reasons of Big Data Project's Failure (continued)

Big data silos	Big data vendors are fond of talking about "data lakes" and "data hubs," but the reality is that many businesses attempt to build the equivalent of data puddles, with sharp boundaries between the marketing data puddle, the manufacturing data puddle, and so on. Big data is more valuable to an organization if the walls between groups come down and their data flows together. Politics or policies often stymie this promise.
Problem avoidance	Sometimes we know or suspect the data will require us to take action that we don't really want to do, like the pharmaceutical industry not running sentiment analysis because it wants to avoid the subsequent legal obligation to report adverse side effects to the U.S. Food and Drug Administration.

Compiled from: https://www.informationweek.com/big-data/big-data-analytics/8-reasons-big-data-projects-fail/a/d-id/1297842

Similarly, Comuzzi (2011) finds that big data projects fail because technology required to process big data (Hadoop, NoSql etc.) is either relatively new or becoming widespread recently; there are shortage of data scientists, or they are not aware of business implications of big data initiatives. In addition, when data and decisions are generated at all levels and all processes of an organization, this makes things hard to govern.

2.13 Big Data Maturity Models

Paradigm changing potential of big data for organizations is highly acknowledged by practitioners and academics (Comuzzi, 2016), and is used by businesses, governments and academia (Mayer-Schönberger, 2013). This revokes the question of, how an organization adapts big data policies and which point it is on adapting the policies for fully benefiting from it. It is suggested that, public organizations may be uncertain of whether and how to implement big data, and they lack the tools to determine if they are ready to fully engage in big data use (Klievink et al., 2017). When big data

applications are adopted to achieve full-fledged opportunities, a roadmap needs to be drawn. Big data maturity assessment becomes significant in this process.

Half of the big data projects of big organizations never had a chance to get completed (Marr, 2015). Heeks (2016) states that since 75% of policies fail, it is beneficial to understand at which point they've failed. Maturity models show where an organization is as it is, and define where to improve and control to get better results (Iversen et al., 1999). Big data maturity models aim to create a capability evaluation tool, which specifically aims to understand big data applications in organizations, and assist these organizations to develop turning points and avoid them from complications. There are Big Data Maturity Models (BDDM) for organizations to follow a track, understand their place in adoption process and compare. Pöppelbuß and Röglinger (2011) suggest that the maturity models assume predictable patterns of transformation and change, and include stages of levels to form a desired path to reach for organizations. For these reasons, the existence of big data maturity models is important from theoretical and practical point of view for developing proper policy-making.

As Giest (2017) suggests, big data readiness / maturity concept is generally used to evaluate public capacities by looking at organizational alignment, capabilities and maturity in connection to big data, and raises complementary points to Digital Era Governance. In other words, big data maturity process adopts a view about maturity similar to other e-government adoption models.

While there are hundreds of maturity models for ICT's (Pöppelbuß & Röglinger, 2011), more than 120 maturity levels for business analytics via ICT's (Mettler et al, 2010, p. 333), and 25 models for e-government adoption maturity models (Heeks, 2015), there are also specific models developed to evaluate big data maturity. In the following sections, first a short history of the maturity models will be discussed. Next, the most well-known and commonly used big data maturity models will be reviewed to provide a theoretical framework for the assessment of big data maturity in Turkey.

2.13.1 History of the Maturity Models

The earliest example of maturity models for using ICT in organizations for decision-making was proposed by Richard Nolan, which dates back to 1973. He developed a model named as stages-of-growth-model basing it on "stage theory" by identifying the elements and their growth depending on time. At the earliest ages of this developmental model, Nolan conceptualized 4 stages: 1). Initiation which means computer acquisition, 2). Contagion meaning intense system development, 3). Control meaning proliferation of controls, and 4). Integration meaning user/service orientation. What is important about this system is that, it drew an outline for all the maturity models about ICT usage after its development. 5th and 6th stages developed in 1979 named as data administration and maturity was the earliest mentioning of subject. The data administration process includes, "integration of applications", "shared data and common systems" which still correlates with the later levels of big data maturity levels used in today. He underlines consecutively the importance of data resource management from computer management; importance of managing data economically. Current Big Data Maturity Models (BDMMs) are using similar methods. Nolan's article is the primary example of a flock of models and articles about maturity.

Building upon Nolan's model, the most commonly used maturity model for organizations for applying ICT's is Capability Maturity Model (CMM) funded by US Secretary of Defense for improving businesses adoption of policies. Heeks (2015) states that since there are various models, before getting into big data maturity models, Layne and Lee's four stage model (2001) being one of the most commonly used and highly cited e-government related papers, draws a stage-model of e-government adoption as shown in the Figure 2.

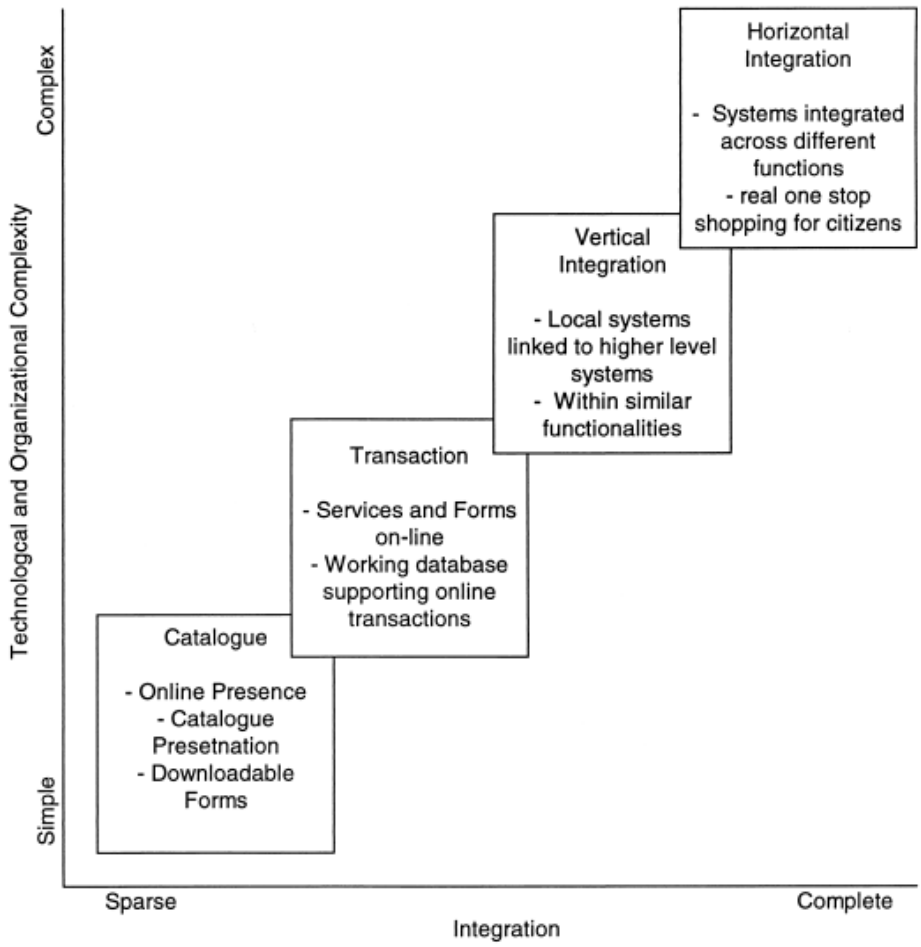


Figure 2: Layne and Lee e-Government Development

Source: Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. *Government information quarterly*, *18*(2), p. 124

Layne and Lee's model, when taken retrospectively as a process, shows that the services are becoming more interactive with each level for citizens. On the other hand, when considered from organization point of view, it also suggests "linking local systems" and "integrating systems across different functions". These levels represent the needs of big data applications in government. Integrated data warehouses and linked data interoperability is a common theme on big data applications in academic articles.

2.13.2 Categorization of Big Data Maturity Models

Table 2: 3 Types of BDMM's

Descriptive Models	Big Data & Analytics Maturity Model(IBM), SAP Big Data Maturity Model, Hortonworks Big Data Maturity Model, Knowledgent Big Data Maturity Assessment, Zaloni Maturity Model, Infotech Maturity Model, Kuraeva's Model, Klievink et al.s Model
Prescriptive Models	Info-Tech Big Data Maturity Assessment Tool El-Darwiche's Model, Radcliffe Big Data Maturity Model, Dell Maturity Model, Booz&Company Model, Van Veenstra's Model,
Comparative Models	TDWI Big Data Maturity Tool, CSC Big Data Maturity Tool

Compiled from Braun, H. (2015). Evaluation of Big Data Maturity Models -a Bench- Marking Study To Support Big Data Maturity Assessment in Organizations.

Big Data Maturity Models are usually classified into three groups: descriptive, prescriptive, and comparative (De Bruin et al. 2005; Comuzzi, 2016). Descriptive models show an organization's maturity in relation to a specific technology or capability, and do not draw a specific roadmap as a diagnostic device. Prescriptive models define the current level of organization and draw a roadmap for a better maturity. Comparative model analyzes similar organizations' position in big data maturity model and draws a roadmap by comparing its counterparts. For achieving comparative big data maturity model, there has to be a pool of other organizations that can be compared for benchmarking (Maier et al., 2009). Table 2 shows the categorization of some maturity models according to this classification.

Some of the maturity levels are ICT company based, such as Dell, SAP, IBM; some other are business analysis companies, TDWI, CSC, Hortonworks are as such, and some of them are developed by the academics, such as El-Darwiche, Van Veenstra's,

Klievink, Kuraeva's model. Among these, Big Data Maturity Models intended for public administration are the models proposed by Klievink et.al (2017) and Kuraeva (2016). The following section reviews the models developed by Klievink et al. and Kuraeva to provide a framework to evaluate where big data in Turkish Public Administration is standing.

2.13.3 Models Specifically Designed for Evaluating Big Data Maturity in Public Administration Context

Although there are various examples of maturity models especially for businesses, when it comes to big data maturity an exhaustive search process ended up with two results which are specifically designed for measuring big data maturity in Public Administration. Klievink et. al. (2017) and Kuraeva (2016) designed the big data maturity models for public Administration. These models are described below. Klievink et al. model provides an evaluation framework at the organizational level, whereas Kuraeva's model is more comprehensive and broader to allow for an analysis at the country level. Since this thesis aims to draw an overall picture of where Turkish Public Administration is at in big data applications in the world, Kuraeva's holistic model will be applied in the analysis section to shed light on the country's big data maturity level, then on the basis of these results recommendations will be made for developing big data applications in public sector in Turkey.

2.13.3.1 Klievink et. al. Big Data Maturity Model

Table 3: Klievink et al. Organizational Alignment: The Theoretical Optimum

Organization Type	1	2	3	4
Main statuory task	Coordination, project-based task, no data used	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 dat use characteristics		-Internal & external datasets -Structured & unstructured 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
Best aligned big data application type		Research	Object Evaluation	Continuous monitroing

Source: Klievink, B., Romijn, B.-J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. Information Systems Frontiers, 19(2), p.273

The system developed by Klievink et al. (2017) consists of three sub groups which measures big data readiness; organizational alignment, capability and maturity. Klievink et al. describes, IT governance, IT resources, internal attitude, external

attitude, legal compliance, data governance and data science expertise as relevant factors for dealing with big data. The first table on organizational aligenment explained in table no. 3

Organizational alignment draws a general framework of awareness of data related issues in various organizations. A coordinative, project-based working organization collects data slower then a registration and documentation organization. For example, census bureau needs constant collection of data in high intensity in real time or near real time bases, the datasets provide basis for other organizations for this reason such organizations are on the fourth type of the organizations. On the other hand, Klievink et. al. describes in the model that public organizations are limited to performing statutory tasks and activities, even if big data analytics provided some of the organizations would be prohibited to do so due to organizational culture and structure. The designers are trying to understand whether organizational strategy and infrastructure is in alignment with specific IT Strategy and Infrastructure.

Table 4: Klievink et al. Organizational Capability

Capability	**Explanation**
IT governance	Capability to design and develop IT strategy, decision-making and responsibility structures, supporting the organization, including integration of new
IT resources	IT systems capability to design, develop and maintain suitable IT infrastructure and expertise to faciliatate current and new IT systems
Internal attitude	Capability to develop internal commitment and vision for new processes and systems, especially openness towards data-driven decision-making

Table 4: Klievink et al. Organizational Capability (continued)

External attitude	Capability to develop external commitment and support for new processes and systems with important stakeholders
Legal compliance	Capability to design and develop a compliance strategy including process design, monitoring and redesign of processes, epecially regarding privacy protection, security and data ownership regulations
Data governance	Capability to design and develop a data strategy including collection, acquisition, quality control and data partnership
Data science expertise	Capability to bundle.'acquire. develop and retain data science knowledge in the organization, especially bundling knowledge on IT. business, statistics and mathematics

Source: Klievink, B., Romijn, B.-J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. Information Systems Frontiers, 19(2), p.275

Organizational maturity dimension, on the other hand, gives e-government development approach. The first level stove pipe organizations differs databases like data silos and restricts the in organization activities on data usage and does not mandatorily suggest IT usage but uses “where possible” this level is the initial digital transformation process where inept data entry is possible. Integrated organizations suggest early models of using in governmental data, such as using Security General Directorate’s traffic tickets usage possibility via Revenue Administration web-page. The third level nationwide portal brings those different infrastructural systems and creates a portal made out of modules. Turkey using www.turkey.gov.tr for this purpose, it is comprised of various modules. Also on this part changing various data via online forms is suggested. The fourth level suggests inter-organizational integration and open data policies. This is where Turkish organizations usually fail to

achieve so. Turkish Statistical Institute's 2017-2021 Statistics Strategy Plan suggests to integrating these datasets and also suggests opening public data. The last level is integrated part of maturity level describing demand-driven proactive model for big data projects.

2.13.3.2 Kuraeva Big Data Processing Maturity Model in Government

This model provides a holistic approach, assessing big data maturity as an overall governmental adoption process rather than an institution based approach.

The Rubric consists of six dimensions to analyze: vision and strategy, open data initiatives, R & D institutions and initiatives, big data maturity level in business sector, data governance, and big data projects experience in public sector.

Vision and Strategy, describes the governments' ICT strategy and big data vision which provide general framework and guidance for the maturity of governmental big data initiatives.

Open Data Initiatives are also one of the most crucial parts of the implementation since they are linked to the development of Big Data.

Research and Development Institutions are important for creating national big data research and developing universal services. The big data policies without a scholarly institutional awareness would lead importing technologies and policies and reintegrating them without a prepared know how might lead incompatibility, which would end up failed policy.

Since one of the major usage of big data in public administration is developing a suitable well developed market and linking big and open data policies with businesses (EU Big Data Report, 2016), business analytics becomes another important part stated by the rubric. Also, public institutions base their models firstly developed by private organizations because earlier adopters of the technology are private institutions.

Table 5: Kuraeva's Big Data Maturity Model for Public Administration

	Aware	**Exploring**	**Optimizing**	**Transforming**
Vision and strategy	• Big data is discussed but not reflected in government strategy. • Government strategy use of data extends simply to financial and regulatory reporting or simply data analytics in.	• Government has IT (ICT) strategy in whole. Big data vision and strategy is not clearly defined or not defined at all. • Big data application is largely experimental and is not clearly defined in strategies.	• Existence of short-term and middle term IT strategies and existence of insights from big data application in public administrations. • Big data vision and strategy is not clearly defined or defined partly.	• Existence of clear vision for a long-term period of using big data. • Publication of regularly reports by governmental institutions encourages the use of insight from big data processing within government processes. • Vision and strategy are developing based on the experience and lessons learned already available.
Open Data initiatives	• There are no government Open Data initiatives in a country.	• There are some Open Data portals but they do not have significant values. Quality and credibility of data are under concern. • Data updates are not regularly or data are not relevant. • Municipal Open Data portals are not presented or presented in small quantities.	• Variety of Open Data portals. • Municipal Open Data portals are essential. However, they are not cover all public agencies or not aggregated into one integrally hub. • There is no single access point for federal and municipal (regional) data. • There are gaps in data consistency and homogeneity.	• Variety of Open Data portals. • Municipal Open Data portals are essential. • Share research results with the public. • There is single access point for federal and municipal (regional) data. • Open Data portals cover all sectors of life. • Information is used as a strategic asset.
R&D institutions and initiatives	• Science background is not developed. • There are no big data educational programs and courses in public or private universities and schools. • There are no research and development initiatives in public institutes or research centers.	• Science background is developed only slightly. There are several commercial big data educational programs and courses in private universities or schools. • There are no research and development initiatives in public institutes or research centers.	• Science background is developed. • There are various big data educational programs and courses in public and private universities or schools. • There are several research and development initiatives in public institutes or research centers.	• Science background is developed. • There are various big data educational programs and courses in public and private universities or schools. • There is big variety of research and development initiatives in public institutes or research centers.

Table 5: Kuraeva's Big Data Maturity Model for Public Administration (continued)

Big data maturity level in business sector	Majority of local companies form Fortune 100 (local rating) have no big data capabilities.	• Just several local big players-companies have big data capabilities. • Business understands the overall big picture from all available data. • Big data providers are mostly overseas companies.	• High level of big data implementations in the following fields: banks, retail, oil and gas, telecommunications. • Approximately half of local big players-companies have big data capabilities.	• High level of big data implementations in all fields. • Majority of local big players-companies have big data capabilities. • Business uses big data to predict outcomes or to adjust processes accordingly.
Data Governance	• There is no clear data ownership assigned. • No standards tools nor documentation is available for use data across the whole public sector. • Data governance is largely manual.	• A government data governance model does not exist or is immature; data owners are commissioned for short-term projects and initiatives. • Understanding of data and its ownership are defined and managed in a piecemeal fashion. • Limited collaboration.	• Data governance model is implemented for the major data entities. • Collaboration between stakeholders is in place. • Governance process regularly reviews this model and its applications; updating and improving as needed. • Government begins to realize benefits of enterprise-wide consistency of data.	• A government-wide data governance model extends active stewardship to the majority of data assets; effective data governance processes are employed by stakeholders and stewards; well-defined standards are adopted.
Big data projects experience in public sector	• There is no big data projects implementations in public agencies.	• There are only several big data projects implementations in public agencies.	• There are variety of big data projects implementation but in specific fields: tax or financial.	• There are variety of big data projects are implemented in different fields.

Source: Kuraeva, A (2016) (p. 41, 42). Big Data Analysis Influence on Public Administration Processes (Master's thesis). Retrieved from https://www.hse.ru/en/edu/vkr/182647584

Data governance is ensuring ongoing production of high quality data (Janssen et al, 2012, p. 259). This would help policy-making feedback loop problems through constant high quality data generating (McCord, 2002, p. 6).

The last part "Big Data projects experiences in public sector" provides a general outlook for the success of the big data implementation process. A government wide

application results in integrated and transformative usage of big data in the later stages of big data maturity.

Similar to other stageway models, the model is dispersed into four levels: awareness, exploring, optimizing, and transforming. Kuraeva's model is easy to apply for any specific model of big data maturity readiness. Kuraeva gives each model a point from 1(aware) to 4 (Transforming) and suggests a general outlook of a country by results.

Table 6: Kuraeva's General Outlook on Maturity.

Total value	**Maturity level**
0-6	Aware
7-12	Exploring
13-18	Optimizing
19-24	Transforming

Source: Kuraeva, A (2016) (p. 41, 42). Big Data Analysis Influence on Public Administration Processes (Master's thesis). Retrieved from https://www.hse.ru/en/edu/vkr/182647584 p. 45

2.13.4 Summary and Evaluation of the Big Data Maturity Models

As organizations or countries increasingly utilize big data, there arises a need to evaluate how well they are doing in their implementation, and how they should proceed to develop further in this area. Overall, the big data maturity models provide a useful framework to evaluate which point an organization or government is at in adopting big data, and what the next steps need to be. There are several big data

maturity models, which are usually developed by ICT companies or business analysis companies. Along with the growing interest in using big data in public administration, recently, there have been attempts to apply these models to evaluate big data readiness / maturity in public organizations. Some models specific for the public sector have also been developed (Kuraeva, 2016; Klievink et al., 2017). The development stages that the models include tend to be similar to each other, with some analysis dimensions that show variance.

It should also be noted that although there are various advantages of these models, there are limitations and critics on them, such as poor theoretical foundation and lack of documentation (Mettler, 2009, p.3). DeBruin (2005) suggests that there is limited documentation on how to prepare a theoretically sound, vigorously tested and extensively accepted model. The models show incremental design process, and each model puts on brick upon another or change some characteristic without thinking about the design decision (Kohlegger et al., 2009).

Measurement of maturity / readiness has always been challenging; issues such as mature enough/ready enough compared to what, how to calculate different dimensions of readiness exactly in social sciences are still vague, including the area of public affairs. Suggesting a new model for assessing big data maturity is beyond the limits of this thesis study. The present study aims to use the model developed by Kuraeva (2016) to conduct a preliminary analysis of big data use in public administration in Turkey, as this model is the most comprehensive one that is specifically designed for evaluating the public sector big data maturity levels.

CHAPTER 3

BIG DATA USE IN PUBLIC ADMINISTRATION AROUND THE WORLD

The fourth industrial revolution and convergence of innovative technologies such as Big Data, Internet of Things, cloud computing, geo-spatial data and broadband, AI and machine learning, are promoting a dramatic shift towards more data and machine-driven societies (UN, 2018). As the importance of data driven policies and data-driven decision-making increase, although the public sector seems to be falling behind compared to the private sector and science, the governments are realizing the promise of big data each passing day (Mullich, 2013). This chapter provides examples on big data use from various countries that use big data initiatives in the world, to provide an insight on how countries utilize this tool in real-life public administration practices.

Therea are several global metrics measuring e-government development, such as latest UN Survey in 2018 placed Turkey 53th among 193 countries. However, unlike the UN e-government Survey or global Open Data Index, which standardizes and creates a global metrics measuring countries, big data readiness and applications differ on various levels and sectors. Hence, it is hard to achieve proper readiness measure for big data. To have a general outlook of big data applications, the other countries are given to understand as a resource for what are they achieving via using big data insights. While choosing the countries developed country examples on e-government chosen due to understand premier examples and developing BRIC countries (Brazil, Russia, India and China) are explained due to their comparability with Turkey.

3.1 United States of America

The United States is one of the prominent appliers of science and technology, which dates back to President Roosevelt's Science Advisory Board. Technological developments lead the George W. Bush Administration to establish an Office of E-Government and Information Technology under the Office of Management and Budget with the E-Government Act of 2002. The United States became the first country in the World in 2009 to create a Chief Technology Officer (CTO) under the Office of Science and Technology. In 2009, USA authorities launched data.gov for citizens, businesses and NGO's use. The website provides 285,795 datasets ranging from education, agriculture, health, finance, climate to local government. Furthermore in 2015 Chief Data Scientist office was created and the first appointee DJ Patill was announced for the mission "to unleash data to benefit all Americans". As can be seen from the pattern, beginning from science advisors of Roosevelt, now more specific officers are employed.

National Institute of Standards and Technology (NIST) has a big data working group named as NBD-PWG (abbreviation of NIST Big Data Working Group) since 2013, after the establishment of Big Data Research and Development initiative, they come to conclusion that the US government has access to varying stream of data from sensors, satellites, social media, mobile communications, e-mail, RFID and enterprise applications.(Mills, et al., 2012: 9) These data has a meaningful end when leaders transform them into meaningful, valuable information

In 2010, President's Council of Advisors on Science and Technology issued a report to the President and the Congress, named "Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology" The report mainly underlines the ever increasing importance of Networking and Information Technologies (NIT) by stating "research explore bold, unconventional ideas that would have enormous impact if they could be realized." (Kearns, 2010, p. xi). Among five of the main policies of the report the first one of them suggests that

since the data volumes are growing exponentially, "Every federal agency needs to have a "big data" strategy" (Kearns, 2010, p. xvii) but it also underlines the importance of securing privacy of the citizens while using these data. The importance of Big Data policies were given on subhead "Large-Scale Data Management and Analysis" (Kearns, 2010: 49), since the term "big data" was not a hype in 2010, it is not strange to see an interchangeable word instead of it. On the Report the Board gives examples about collecting health data from cancer patients from around the country and analyzing it, FBI collecting data from different sources for prediction in crime. Unsurprisingly, The National Cancer Institute is using the exact same policy through Genomic Data Commons and via Cloud Resources (Hinkson et al., 2017) since 2015 and FBI is now hiring Big Data Technologists through LinkedIn.[13] The Emory University Hospital in Atlanta, for example, has worked with IBM to develop bedside monitors in their Intensive Care Unit that collect and analyze over 100,000 data points for each patient each second.

The first report specifically focused on big data was demanded by President Obama for a comprehensive review of big data on different effects about economic, social and government impact. His emphasis on big data Implementation Policies lead him to be called as "Big Data President" by Washington Post.[14] On May 2014, a White House Paper entitled Big Data: Seizing Opportunities Preserving Values is presented to the President's Office. The paper draws a general outline of big data, links it with open data, suggests possible usages in public and private sector and tries to come up with a policy framework for big data. Since the earliest current definition of big data by Roger Mougalas in 2005, the US companies were using the opportunities of unstructured datasets to analyze and have positive outcomes for their own businesses. This White House Paper was the first to define a proper public policy outlook for the United States.

[13] FBI Big Data Technologist vacancy on LinkedIn retrieved 24.07.2018 from: https://www.linkedin.com/jobs/view/big-data-technologist-at-federal-bureau-of-investigation-fbi-738167904

[14] Scola, N. (Jun 14, 2013) Obama the Big Data President retrieved: 14.07.2018 from https://www.washingtonpost.com/opinions/obama-the-big-data-president/2013/06/14/1d71fe2e-d391-11e2-b05f-3ea3f0e7bb5a_story.html?noredirect=on&utm_term=.e14cb7b8099a

Since then, the U.S Administration increased its usage of big data on social, healthcare economic and public issues. State governments are employing Chief Data Officers since 2011, starting from the New York City. The State Department prepared Information Technology Strategic Plans in the US. In The Department of State's Information Technology Plan 2017-2019 it states that in order to benefit from the datasets the department should use big data to make informed decisions, use in problem-solving, and risk analysis (US Department of State, 2016, p. 4). Likewise other agencies, the Department of State is aware of the potential benefits of big data implications and although structured datasets are important for the department, the big data analytics would give results of business analytics and help in data-driven decision-making.

The Department of Homeland Security emphasizes the importance of data and data analysis in its 2015-2018 strategic plan. Especially "data-driven decision making" plays a big role in the plan (US Department of State, 2016: 9, 14). Data-driven decision-making is possible with big data analysis. Food & Drug Administration (FDA) states that the master data management strategy for handling business data and big data requirements policy is in process as of 2016 (FDA, 2016: 25). Federal Deposit Insurance Corporation (FDIC) likewise DHS gives importance on data-driven decision-making in its 2017-2020 Strategic Plan. Department of Health and Human Services CIO cover letter opens mentioning big data and Internet of Things. When taken in consideration all of the departments, the strategic plans of them always emphasize the importance of data-Driven decision-making.

On January 12, 2015, the President proposed the Student Digital Privacy Act: a new legislative proposal designed to provide teachers and parents the confidence they need to enhance teaching and learning with the best technology—by ensuring that data collected in the educational context is used only for educational purposes. (US Executive Office of the President, 2015)

As a government strategy, the importance of big data is observable, and there are reports and plans on departmental levels, there are U.S public administration usage of big data reported by the NIST, Big Data working group, some of the examples are taken from the Case report which consists of 51 case examples (Chang, 2015, p. 2). All these cases are accessible through NIST's webpage.[15] The Cases are divided into 9 subcategories, Government Operations, Commercial, Defense, Healthcare and Life Sciences, Deep Learning and Social Media, The Ecosystem for Research, Astronomy and Physics; Earth, Environmental and Polar Science; Energy. Some examples under each category will be given here. Although more advanced systems are used in the commercial data of business and other cases are not mentioned if they do not have a direct relation with the public administration.

Some of the policies are toward data preservation for future. Under the government operations, for example, 380 terabytes of structured census data will be preserved for 75 years. In this case, there is not much multidimensional approach but rather the government preserves the data for scientific purposes. The US Census Bureau uses big data for statistical survey response improvement by analyzing streaming survey data from administrative resources (Mills, et al., 2012) The US Census Bureau uses, statistical survey response improvement by analyzing streaming survey data from administrative resources (Mills et al., 2012: 7) the Bureau uses (HADOOP, Spark, Hive, R, Cassandra and MYSQL) to analyze these survey data. Similar to this, National Archives and Records Administration preserves search, retrieve and prepare for long-term preservation of the data, through digitizing and making them machine readable. According to Cook (2015), political will for a national registry is lacking in the US due to privacy and administrative concerns, and s/he emphasizes that linking data across agencies has a long way to go due to lack of trust.

[15] The page is available on: https://bigdatawg.nist.gov/usecases.php

Under the defense title, David Boyd, an expert worked on Big Data Frameworks for the Army, DARPA and DHS cloud computing research efforts[16], gives examples about Cloud Large-Scale Geospatial Analysis and Visualization since the traditional Geospatial Information Systems are slow to visualize the constant fluctuating of millions of data, Boyd suggests a secure system which even works on low-bandwidth systems and gives proper analysis through cloud storage. For a security policy, intelligence data analysis and processing, Boyd mentions that although the big data is operable on HADOOP and similar map reduce systems, those systems can only work for mid-size clusters to analyze. Since the huge data that identifies relationships between entities, identifying tracking the location of hostile actors, and reasoning from diverse disconnected and unstructured data sources is hard to work on with these, these data should be withdrawn from differentiating data silos of governments and should be used in semantically integrated data space.

In addition to defense, for security purposes CIA, DHS, NSA and FBI works with PALANTIR Technologies, a software company specialized in big data analytics for counter-terrorism and fraud detection. The Company's one of two assets, Palantir Gotham is used for integrating structured data into unstructured data, for example to detect people likely to commit crime. In this system, "information from rap sheets, parole reports, police interviews, and other sources is fed into the system to generate a list of people the department defines as chronic offenders." Other examples consist of Marines using big data to detect roadside bombs (Marr, 2015), coordinating disaster relief, finding missing and abused children.

In the Boston Marathon Bombing incidence, big data enabled the analysis of half a million images. These images of unstructured data were analyzed by algorithms on image processing to look for anomalies and patterns (Helms, 2015). By doing so, sensor information for criminal behavior enabled real-time analysis, reduced the time to decide and interfering people based on data. On the other hand, there was a

[16] Retrieved from Boyds LinkedIn professional account on 20.07.2018: https://www.linkedin.com/in/david-boyd-43a83439

scandalous privacy issue kept on the headlines for a while. The deceased perpetrator was using an Apple iPhone and the phone was linked to Apple's cloud system. Since the data encryption was made by the system itself Apple did not want to share the encryption codes in its system to unlock the phone. FBI, without the help of Apple, employed data scientists to break the code in 3 trials and successfully did so. But this raised another question on Apple's encryption system. A year after the decryption of this specific phone's key, a huge leakage of celebrity photos from the cloud system was served to the Internet in 2014 and caused increasing concerns about cloud server encryption security of the company.

A similar security measure was employed by the Los Angeles Police Department (LAPD). The department created a Real-Time Analysis and Critical Response Division, in collaboration with researchers from University of California, Los Angeles (UCLA). The system uses historical and real-time data to "predict" where future crime might occur. These data allow LAPD to concentrate dataflow from specific areas. One of the common usage is geospatial analysis. As Desouza (2012) said the data used in this case provide a rare but important example of how "big data" with high volume, velocity, variety, and complexity can support public efforts.

Like many other countries (Salas-Vega, 2015), the US government also began to use big data analysis on health policies. Medicare and Medicaid in the US now uses fraud prevention system for predictive analysis. Before using Big Data Analytics, selected papers were monitored from a pool of documents to detect the fraud (Mills, et al., 2012). By using Big Data Analytics Tool, since 2011 $1.5 billion was saved by the government. [17] The system was first used against fraudulent healthcare providers. If a doctor files more patients than he can serve, such as a radiologist filing forms he never produced, the system detects this activity by comparing it to another practitioner in the

[17] Agrawal et al. (May 24, 2016) Medicare's Big Data Tools to Fight and Prevent Fraud Yield over £1,5 billions in Savings retrieved 11.07.2018 from: http://www.modernhealthcare.com/article/20160524/NEWS/160529960/commentary-medicares-big-data-tools-to-fight-and-prevent-fraud-yieldShare

area. Also in health policy, the Centre for Disease control uses big data to analyze to collect and analyze spatial data to find spreading of the epidemics. Another point is made by Helm (2011) when he mentions that the Social Security Administration (SSA) is using a big data strategy to analyze unstructured data in the form of disability claims. The SSA is now able to process medical taxonomies and expected diagnoses more rapidly and efficiently to recreate the decision-making process better identify suspected fraudulent claims. Also, social media analysis is used by the Centers for Disease Control (CDC) to track the spread of epidemics and other public health threats (Marr. B, 231). The National Institutes of Health (NIH) started the Big Data to Knowledge (BD2K) in 2012 to stimulate-data driven discovery.

The US Immigration & Customs Department is one of the heavily emphasized departments by the government having cutting edge technologies since 9/11. Department of Homeland Security (DHS) developed a system with Arizona University called Avatar (Marr, 2016, p. 11) and similarly, CIA is working with Palantir to screen air travelers.[18] The AVATAR system uses, facial recognition, voice and body language analysis to compare people entering the US Customs with suspicious records. If the data matches, then for further investigation the person is assisted by a human agent.

Metropolitan Transportation Commission (MTC) of San Francisco saved over 250k per year in direct data collection and contraction costs. Like the MTC of San Francisco, in Miami a county uses big data analytics to create public transport traffic patterns (Jalote,2013). By doing so, both demand for public transportation and service quality increased without using more budget for new vehicles.

Responding to the correct issue with on time data processing methods would lead to the increasing value of data. Taking action on time and in correct place increases the importance of big data. This is not limited to crimes, such as crimes against city and

[18] For Bloomberg Businessweek's Analysis on the company: https://www.bloomberg.com/features/2018-palantir-peter-thiel/

health, which can be prevented through using analytics. In 2012, the New York City Department of Environmental Protection used the power of big data to prevent illegal sewer dumping by restaurants. The city was experiencing clogged drains due to illegal dumping of cooking oil into neighborhood sewers. Traditionally, to catch illegal dumping, the health department would send inspectors to restaurants that reported problems associated with backed-up sewers.

However, by using big data analytics, the ICT team working at City Hall, combined data on restaurants which did not obtain a carting service certificate from the City of New York Business Integrity Commission with geospatial location tools to derive a list of restaurants that were statistically likely suspects. As a result, city officials achieved a 95% success rate in tracking down illegal dumping of cooking oil by restaurants. Office of Policy and Strategic Planning of the New York City's Mayor's function is not limited to clogged drains, but knows how to harness big data by employing analysts who are mining data from various city agencies ranging from building and development issues, infrastructure to detecting the selling of bootleg cigarettes. (Howard, 2012)

In the US cities, use of big data analysis by local governments is a measure of being a smart city. For example, by incorporating crowdsourcing and geospatial mapping Boston Street Bump App prevented the city government to lose $200k per year by developing a $80k app. (Zook, 2017) On the other hand Rampton (2014) suggests that after launch of the application, the crew members were mostly fixing wealthy neighborhoods due to smart phones are more prevalent in those neighborhoods. Although Desouza (2015) criticizes this by saying that the ICT devices favors upper-classes in the US since not everyone is online, on the other hand it should also be noted that since this incident the prices of online connection is falling and the number of mobile devices is increasing tremendously.[19] At the beginning of 2018, 4 billion people were connected to the Internet.

[19]For a predictive analysis, see: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/

The policies ranges from local to global, an uncommon example is from Women's Olympic Cycling Team of the US. Although using computerized data analytics for sports goes back to 1980's, this team uses big data analytics, by collecting data on the athletes such as, diet, sleep patterns and blood test and analyzing them individually (Franks & Nagelkerke, 1988) resulted with a silver medal in 2012, a never before achieved success by a US women cycling team. The data was analyzed through Hadoop system. It is easy to say that like Ivan Drago's performance analytics in Rocky IV, now by using more unstructured data, athletes also enjoy higher performance without the help of performance enhancing drugs.

3.2 United Kingdom

One way of understanding the promise of e-government UK is to observe the gradual rising trend of the country in the UN E-Government Surveys since 2005. In 2016 UK championed as the top contender of E-Government Development index. Such emphasis on e-Government policies is relatable with early adoption of big data application in the UK. UK suggests that potential public sector applications of big data include real time management of information, countering non-compliance, fraud and error, transforming and personalizing the citizen experience, improving healthcare and patient outcomes and delivering more timely population estimates as lower cost (Yiu, 2012).

A report entitled "Seizing the Data Opportunity" refers big data as one of the 8 great technologies (Willetz, D., 2013). Actually, it is also the first great technology to be named by the State Minister of Universities and Science. Analogous with many other countries using big data as a policy instrument, UK also implemented an open data structure similar to USA's data.gov.

The UK version of the data structure was named and linked as data.gov.uk, which now currently hosts 45,000 datasets in the web page. UK policymakers see data as a whole concept and do not separate it as big or open data. There were several open data

predecessor applications developed such as Open Government License or legislation.gov.uk before the big data application studies have started.

Regarding its big data policy, UK focuses on three subjects, (i) developing the human capital according to data hungry era (skills), (ii) providing suitable tools and structures to benefit from data analytics and (iii) data capability (infrastructure), as enabling the data to be used by/among academics, businesses, government properly (shared data). The first condition signifies the problem of all countries and businesses that try to implement big data policies as well as the shortage of skilled Data Analysts[20]. The third condition also works hand in hand with the open data policies. It can be deduced that UK sees Big Data as a policy window to catch while defining one of the chapters in the report "data opportunity" (Willetz, 2016). In addition to that, data format, privacy of citizens, security of the data issues was mentioned in the paper like many other examples.

UK, as the most important and earliest adopters for big data analytics, understands the importance of data analytics for detecting certain measures regarding to budgeting, collecting taxes, and analyzing the financial markets. Due to the importance of the previously mentioned measures for public administration, early adopters are the ones who avoid a possible future digital divide. One of the early adaptors of big data analytics is called "British Connect", which is enforced by the British Revenue and Customs Office to avoid tax avoidance and evasion and cost £45 million (Maciejewski 2016: 12) The system initial investment amount only in a year with £1,4 billion additional tax revenue for services. The system aroused the interest of the British government and received £150 million further investment. With the help of this system £35 billions of unpaid taxes have been collected during the following year (Caldwell, 2014). Like British Connect, HMRC worked with a consulting company, Capgemini, to create another system named as Analytics for Debtor Profiling and Targeting

[20] An analysis of the shortage is done by Columbus, L (May 13, 2017) retrieved: 15.052018 from https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#5f64211d7e3b

(ADEPT) to collect debts in a proactive manner and the project successfully realized £3 billion of revenue by reminding citizens their debts through big data analytics.

Another financial system example from UK is the use of big data by Royal Scotland Bank. Even though the bank has no similar characteristics to the central bank, it has a privilege to issue money and works hand in hand with the UK Government. Their system personalizes the customer data for creating better plans and offers.
Healthcare is one of the top priority areas in big data usage, National Health Services (NHS) works with Mastodon C and Open Healthcare UK, analyzed statin prescriptions (common name of lipid lowering medications) and as a result found out that the licensed drug prescriptions are so common in some areas and it costs the NHS 200m £[21]. If the doctors did prescribe the generic form of the drug (which is 11 times cheaper than licensed version) NHS could have saved median of £22.96 millions per month.

Milton Keynes, a town of 200k population, bordering with London, uses big data to place itself a championing of Smart Cities (Marr, 2016). The city council introduced MK Smart application, to assess its effectiveness and impacts for the residents and engaged citizens to give opinion on which policies are better working. This signifies the e-democracy part of the project. Community Action Plan for Energy (CAPE) is built to detect thermal leakage in houses by using satellite data to suggest residents to use better insulation or improve their homes to reduce their carbon footprint and their energy expenditures. Also MK Smart detected usage of smart street lighting for better electric expenditure plan for fiscal advantage.

Like Milton Keynes town's smart city measures, London also introduced Transport for London to map customer journeys, manage unexpected events and provide personalized travel information. (Marr, 2016) Through analyzing e-transportation ticket Oyster, the system successfully implemented a better public transportation network and schedule.

[21] For a detailed UK prescription analysis, see: http://www.prescribinganalytics.com/

Another is the “CityOps” (city operations) as a service: The digital solutions driving improvements in physical efficiency are increasingly available “as a service”, thus transferring upfront capital investment into operational expenses. This enables city administrators to take more financial risks and implement solutions rapidly. As an example; Norfolk County Council (UK) was facing budgetary constraints with existing services. The council used a cloud-based model to transform municipal service delivery and achieved an overall saving of $10 million. The solution introduced technologies, such as big data and the cloud, to transform how internal departments collaborate with each other

3.3 Australia

The first version of Big Data for Better Practice Guide is prepared in 2013 and revised with the 2.0 version in 2015 by the Australian Government. The first version includes the definition of big data and a general introduction and several opportunities of the system, such as service delivery and policy development. The second version, after two years of improvement, directly introduces the part “implementing big data capability”. The Australian government creates its step by step big data adoption model each step representing a more developed and intrinsic plan. The model also underlines the implementation of information and privacy management under the big data policies, then creates an applicable policy for the responsible parties (government, practitioners, decision makers, data analysts). Desouza (2016) mentions that the Australian Government Information Management Office, the agency who is responsible for developing policy and guidelines on important information technology issues, has set up a Whole of Government Data Analytics Centre of Excellence, co-working with the Australian Taxation Office. Among the countries observed, Australia develops a well-planned policy to adapt big data policies through a meticulously designed implementation process.

3.4 Japan

Since 2001, Japan's Ministry of Internal Affairs and Communications prepare white papers[22] on e-government technology. By acknowledging US's roadmap, the report states "The utilization of Big Data will greatly improve the potential power of ICT; while the U. S. has already started strategic activities, Japan is focusing on it as a national strategic resource." The Japanese Government sees using of big data as a means of developing commercialization (Kazuyuki, 2017). 2017 White Paper, specifically addresses this issue in reports, utilization of big data would lead economic growth and social reform.(Japan White Paper, 2017) Japan also developed the Act on the Protection of Personal Information to ensure the standards for users. Japanese decision-makers decided that the big data applications will be a part of growth strategy of the government. (British Embassy, Tokyo 2013). The government invested 90 million dollars for research and infrastructure of big data. The Council of Information and Communications and the ICT Strategy Committee, both branches of the Ministry of Internal Affairs and Communications, designated "big data applications" as a crucial mission for 2020 Japan. A big-data expert group was formed to search for technical solutions and manage institutional issues in deploying big data. (Kim, et al., 2014)

3.5 France

Emmanuel Macron, just like Barack Obama's emphasis in the U.S, has called on EU leaders to implement a wide big data policy plan in the beginning of 2018. To use the potential of developing ICT, France have established a Ministry for Digital Affairs in 2014. The ministry published a bill named Digital Republic (fr. Republic Numerique)

[22] A white paper is an authoritative report or guide that informs readers concisely about a complex issue and presents the issuing body of knowledge .
White Paper. (n.d.). Retrieved August 31st, 2018, from http://www.yourdictionary.com/white-paper

which defines three policies regarding wider data and knowledge dissemination, equal rights for internet users, and fraternity through an inclusive society.

The first policy is distributing the big data government collected through an open data portal[23]. Through Digital Republic, the Ministry, similar to other EU Nations, introduced a detailed privacy policy. On the other hand, France is using IBM's i2 (Ideas to Business) big data analytics system to prevent crime (Krubler, 2017). This system has similar properties with the PALANTIR system that CIA is using and has the same founding members. The security policies were one of the main concerns. Furthermore, almost all research about France's current application on big data results with cloud usage, market integration and focuses on business and government relations. Although the Digital Republic bill draws a sketchy big data implementation plan, "building a data-oriented state" shows that, as a government policy France is still at the earlier stages of development and takes a conservative approach to this concept.

3.6 Singapore

Singapore is one of the earliest adopters of the e-government practices (TBD, 2002). In 2004, the Risk Assessment and Horizon Scanning (RAHS) program is launched under the National Security Coordination Council by Singapore government, to address national security, infectious diseases and other concerns (Kim, et al., 2014). The system analyzes large-scale datasets, proactively manages national threats ranging from terrorist attacks to financial crises and infectious diseases. Kim states that: "the system developed further and there is an experimentation center opened under the RAHS which focuses on new technological means to support policy making for RAHS and it maintains RAHS through systematic upgrades of the big-data infrastructure." (Kim, et al., 2014, p. 84) The government worked with IBM to improve productivity and maritime safety in its ports (Zeng, 2015).

[23] Open Data Portal of France: www.data.gouv.fr

It means that by using big data, Singapore Government is using evidence-based policy making devices, and it achieved the proactive responsive model of big-data maturity.

One of the most common examples given about big-data usage is the Google flu trends.(Dugas, et al, 2012; Cook, 2011; Olson et al., 2013; Ortiz, 2013; Valdivia et al., 2010) It is even called "poster child of big data"[24] In 2013 it failed to predict to do so (Butler, 2013; Santilana, 2014) by 140%. This leads to 100% increase in doctor appointments compared to previous year. Since then the system is discontinued. According to Kim (2014), Experimentation Center of Singapore, similar to the Google Flu trends used for:

> …exploration of possible scenarios involving importation of avian influenza into Singapore and assessment of the threat of outbreaks occurring throughout southeast Asia. Aiming to create value through big data research, analysis, and applications, the Singapore government launched the portal site in 2011 http://data. gov.sg/ to provide access to publicly available government data gathered from more than 5,000 datasets from 50 ministries and agencies. (p. 84)

Helbing (2015) refers that big data usage in governments lead to "some elites increasingly become excited about" getting results similar to the approach of the Singaporean "big government" model which he compares as "something like an authoritarian democracy ruled according to the principle of a benevolent, wise king" (Helbing, 2015, p. 56).

One of the areas that the Singapore Government used big data is improving the public transport in the country. By using geospatial analysis and consumer modeling, the Land Transport Authority (LTA) invested to improve 6.3 million daily trips in Singapore (IBM, 2014). Since 2010, LTA's Planning for Land Transport Network (PLANET) gathers daily 12 million public transport transactions and analyzes

[24] For a detailed analysis of Google Flu Trends skewed analysis, see: https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/

passenger choices to reach their destination. Maciejewski (2012, 129) says that the system helped to identify and reduce crowding on buses by increasing the frequency of bus services on the most popular transport routes (Hartung, 2013)

3.7 China

China is adapting e-government technologies since 1980's. The country has a unique place among the other big data analytics implementing countries, considering the volume of data and also country's state and political structure. For instance, Great Firewall of China is a famous example while mentioning technological policies of the country. China is the highest populated country and the one of the most omnipresent technologically advanced surveillance applied states in the world[25]. For Chinese government, big data is declared as Emerging Industry. In 2015 government announced $787 million project to build a big database for international shipping, but the investment also consists of industrial park, research academy and a big data centre.[26] On the other hand the technology does not only mean increasing the business activity and use of technological innovation (Cheng, 2012) but using it for surveillance of citizens, security policies toward regimes trust is an important point (Zheng, 2017). For example, privacy laws are not directly related to the privacy of the citizens but the protection of the government.

From the government's perspective, collecting data to reach results has always been a problem since Mao Zedung's Great Leap Forward (Chen, 2015). According to Zheng (2017) due to officials ambition to improve their career prospects, they tend to send bloated numbers to the center. This lead to a miscalculation of the GDP of China. State officials also acknowledged this issue and in 2010 Primer Minister Keqiang created another way to understand the track of economy, the cargo volume on the provinces'

[25] For an example of a BBC reporter was found in 7 minutes via surveillance system retrieved: https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter/

[26] Port Technology, 2015 China to launch big database for shipping (www.porttechnology.org)

railways, electricity consumption and loans disbursed by bank. Chinese officials misdirected increase in numbers. Authoritarian regimes like China also tend to adapt the big data policies. The prime minister also encouraged the usage of open data in 2015 (Chen, 2015).

An interesting use of big data in China is that 2% of the National Science Foundation of China's big data budget is used on Marxism-Leninism and Scientific Socialism. According to Zheng (2017) this policy resulted with results such as: 'Ideological security in the era of big data' and 'Innovative approach and methods to foster socialist core values among youth in the era of big data'.

The latest usage of big data by the government is Social Credit System described in its 13th Development Plan in 2015 , which tracks and grades the citizens for their obedience and state and official ideology, an older version "dangan" which means "archives" were used in countries history since the establishment of the regime. By using big data analytics and incorporating with business moghuls such as alibaba, baidu the government collects the citizen data and analyze behavior.

China as a part of UNDP policies and business integration shows technological optimism like some of the policies in Western Democracies. The roadmap for big data usage in public administration for development includes two policies, to create big data environment, that described as data philanthropy, meaning sharing of data with government. Second policy is resembles the big data usage in other countries, tackling development challenges by using big data technologies such as: promoting sustainable e-waste disposal practices, improving productivity of the public sector, understanding socioeconomic development trend, mapping poverty, improving urban transport planning, Identify pollution hotspots in cities. On the other hand, Chinese big data policies are described as a repression technology (Diamond, 2010) such as surveillance systems and profiling through social credit system are showing the threatening side of big data applications in public administration.

3.8 Brazil

Brazil's e-government initiative began in 2000. The country aims to improve efficiency and public policymaking for citizens since then. However there is still lack of formal legislation about data protection (da Mota et al., 2016) In their 2016-2019 e-government Strategy plan, opportunity of big data usage underlined, in the plan formal legislative actions explained.

The country began to use big data analytics on various areas. The city government and IBM partnered for better service quality by building Rio Operations Center an integrated data analytics center. (Kitchin, 2014b) The system uses citywide analytics of data collected from thirty different agencies such as traffic, public transportation, weather feeds, information sent by citizens and employees, and emergency services, the whole system uses a quick responsive analytics to improve efficiency. As a country placed on tropical belt, a common illness Dengue Fever also prevented through reporting by Rio Citizens. The local governments are common users of big data analytics. On the local government level there are other applications, Sao Paulo government uses big data analytics to improve public transportation services and tackle crime. A system called Detecta, used by Sao Public Security Bureau of Sao Paulo and Microsoft, integrates cameras around the city to catch the criminals through visual data analytics.

The Detecta system is software made by the partnership between the Public Security Bureau of Sao Paulo and Microsoft, and the main idea of the project is connect all the security cameras and a huge database to help the Military Police of State to catch the criminal (da Mota, 2016). Usage of big data analytics helps to improve security, transparency and decision-making, but da Mota (2016) suggests that the country fall behind on legal structures and has to improve its privacy and security policies on data management of citizens.

Brazil government uses big data analytics not only on local governments level but state and federal governments uses big data analytics. For example, Ministry of Science and Technology uses big data analytics to decrease deforestation of rainforests in Brazil through combining various socio-economic, weather and physical databases with biodiversity. (Malhado & Ladle, 2016) (Horita et al., 2017)

The country also used big data analytics to improve government spending; they managed to save 40% of spending compared to before introduction of a system for student uniforms. The system helps public employees to make a decision on when and how much quantity they need to purchase uniforms.

Ministry of Science and Technology uses big data analytics to decrease deforestation of the Rainforests of Brazil through combining various socio-economic, weather and physical databases with biodiversity data.

Brazil is an active member of Open Government Partnership and prepared Open Government action plan for three times, latest being prepared for 2016-2019. Open Data played an active political role on anti-corruption in Brazil's current history. A non-governmental organization, Contas Abertas (Open Accounts) reviewed publicly available government data from transparency portals. They found irregularities of finances about Petrobras' semi-public petrol company of Brazil from 2003-2010 era, which was headed at the time by the current president Dilma Rousseff. Inevitably, a corruption case opened and through impeachment, Rousseff's presidency came to an end. The public data used as a means for anti-corruption means in this case. In 2013, DATAVIVA an open data portal providing economical data created by the state of Minas Gervas, adapted to whole country level in three months.

The Waseda University Digital Government Report (2017) claims although Brazil is still investing in infrastructure and trying to develop its e-government projects, it is on an enhanced level in open government initiatives. Their Open Data National Infrastructure aimed to improve transparency and for better policy-making.

3.9 Russian Federation

Public sector of Russia is a developing example according to Kuraeva (2015). In its 2011-2020 Information Society Plan, An officer from CROC an IT Consulting Company from Russia, explains that Russia is different from the Western countries due to lack of developed cloud computing in the country. Since big data and cloud computing goes hand in hand lack of cloud computing curbs development of the big data. Kuraeva (2016) suggests that although private firms are driven by profit seeking, public sectors tend to enter big data realm slower then them this leads being big data implementation rare in Russia.

A public sector example from Russia is Central Bank of Russian Federation's applications to integrate other commercial banks to report to a central database to the citizen's account activities.

Although there are examples of such as Open Budget Portal of Moscow or Portal of Budgeting System of Russian Federation, the country ended its Open Government Partnership membership on 2013 claiming it will develop a better model on its own, rather then integrating into an international model.

According to Kuraeva, the country does not spare it's budget on Information Technologies due to limited budget as a developing country. The 2011-2020 Information Society Plan's general framework assigns some issues on data, such as developing legal framework for data protection, and creating digital data management for healthcare industry but the research could not find implication of these plans as of 2018.

3.10 India

Second most populated country in the world is harnessing of big data analytics in its daily politics. The last elections showed the importance big data analytics, Narendra

Modi used big data analytics before 2014 elections and social media analysis to where his popularity is lower and increased campaigns around that area.

In India there are other innovative ways of using big data. For example, ipaidabribe.com provides citizens an online reporting tool for bribery and fraud in their daily relations with the government officers. (Chandan, 2016) This webpage spread around the world and now used in 15 different countries, the webpage also helps to report honest officers. Although this database is an inept version which can only be used for anti-corruption measures on citizen basis. The data can also be used for nationwide level to analyze which organizations and where the officials have higher tendency for bribery and possible action measures can be taken. India uses mygov.in open government crowdsourcing platform to analyze citizens' point of view on government. For example, the citizens can comment each declaration made by president on an official platform, the Prime Minister demands for flood donations, citizens can suggest subjects on upcoming prime minister speeches. When taken from the crowdsourcing and open government point of view, India shows a proactive way of achieving e-Democracy. The country also has an open government data platform sharing public datasets with citizens and other third parties.

The Indian government used data analysts, like many other countries to catch tax evaders, register deregistered firms, and used geo-tagging to improve agriculture, manage disasters. India, uses big data analytics for sustainable development goals (Sharma & Kaushik, 2018).

3.11 Evaluation of the Countries

The review above first includes the main examples of big data from developed countries. 2018 UN e-Government Survey lists USA, UK, Singapore, France and Australia as the developed examples of e-Government policies. The difference of these countries is that, historically, they provide strategic plans to develop a certain policy on e-Government process on a step-by-step approach and follow the tracks. Big data

usage awareness is prevalent earlier in all of them. Since there are developed businesses on data analytics, these countries work hand in hand with them to increase efficiency. For example, USA's strategic partners are spread from its Silicon Valley, or UK incorporates Business Analytics companies to improve their big data roadmap.

When all countries are taken into consideration, the earliest big data applications are aimed towards improving economic measures. For this reason, preventing tax evasion and tax fraud has a pioneering role in these countries. Big data analytics are used for better market integration, improving G2B applications and decrease budget spending. Moreover, big data applications are common in local governments, commonly smart city measures taken by the local governments. This can also be seen as pilot applications for greater projects. Examples such as crime prevention, public health related crowdsourcing applications and infrastructure improving from the local level also provides a lesson for statewide projects. Brazilian Dataviva is a good example, it was first an open government project for a state in the country, the project spread into federal level and also some other Latin American countries adopted the project.

Another priority is on security, both the USA, France, UK, France, Brazil, China, Singapore using big data analytics to improve their security measures.

USA's action plans and applications are so wide that there are plethora of policy documents and applications on various areas. The country analyzes the possible areas for big data usage and picks the vital examples to apply on earlier stages, then improves to different areas. The successful big data projects works along with good strategic planning documents.

BRIC countries entered big data analytics a short period after developed countries, by taking examples from them. Among the BRIC countries, Russia is giving the least importance to big data adoption. The reason for this explained by Kuraeva (2016) is that Russia having a vast land and lack of IT budget and countrywide cloud computing infrastructure. Another thing about BRIC's is that except Brazil, none of the countries

are in Open Government Partnership, either they left the partnership earlier stages like India and Russia or never attended like China.

Among these countries, Brazil is the prominent examples using big data from preventing deforestation, improving education to presenting better local government via streaming data analysis. China uses big data analytics heavily on surveillance and security reasons. Russia does not show various examples, and India's big data awareness increased after 2016 with the encouragement of the Prime Minister Modi. Indian Prime Minister what US government does with online petition to the president and interacts with the citizens through a portal. Turkey's CIMER(Turkish Republic Presidency Communication Center) is seen as a similar application by the UN E-Government 2018 survey.

When a general outlook is made, the institutional culture and infrastructures are what made the difference. For example, the US government works with IT Consultancy companies from silicon valley for policy making documents, from the other side countries like Russia doesn't have proper infrastructure to work upon.

Another important conclusion is about the development of legal regulations. Successful big data implications are developed along with a legal framework of the country upon data governance and ownership. Big data maturity models are also taken into consideration in this legal framework.

CHAPTER 4

EVALUATION OF BIG DATA USE IN PUBLIC ADMINISTRATION IN TURKEY

The rise of globalization and technological advances made the governments to realize the importance of new digital age and implement it to varying ends. Denhardt & Denhardt (2000) explains this by stating that after the mid-1980's the New Public Management's common theme has been the use of market mechanisms and terminology, in which the relationship between public agencies and their customers is understood as based on self-interest, involving transactions similar to those occurring in the marketplace. Public managers are urged to "steer, not row" their organizations but they are challenged to find innovative ways to achieve results or to privatize functions previously provided by government." (ibid: 550) One of these innovative ways is e-government transformation. Denhardt and Denhardt (2000) further states that EU created the e-Europe+ action plan in 2000 to make Europe a dynamic and competitive marketplace in the world. Similarly, understanding the advantages of ICT usage in many countries ranging from USA to Poland, e-government became a way to increase the efficiency of the governments in delivering the public services (Zheng, 2017).

The historical process of increasing technological advancement influenced the industrial, academicals and governmental usage. The first computer in Turkey was used in General Directorate of Motorways in 1960's. First Internet connection was made in Turkey via a tcp/ip connection from ODTÜ to National Science Foundation in the USA. Turkey has been producing strategic documents on e-government since 1996's (3376 paged mammoth report TUENA). Since then, e-government applications are getting broader and penetrating more areas day by day. This chapter aims to give

a brief overview on development of Turkey's big data related policies, and then evaluates its outlook on big data maturity.

4.1. Big Data Applications in Turkish Public Administration

Various actors are benefiting from big data in Turkey, for example a researcher Başıhoş (2016) from The Economic Policy Research Foundation of Turkey (TEPAV), discussed that since the State Statistics Agency does not publish province based gross national product and find a way through big data analytics. The researcher took night lights spreading from the provinces throughout the years and find that Statistics Agency miscalculated the GDP by %34,64 less than World Bank Real GDP data, and infers that night lights are a good indicator of development. This example is coming from an NGO, there are also usage examples of big data in Turkish government agencies.

The strategy documents, action plans and development plans raised an awareness of benefits of the data usage and also throughout the years the documents show early signs of big data awareness. These plans lead government agencies to take initiatives to use big data applications. For example, TÜBITAK-BİLGEM developed a big data and cloud computing lab B3LAB, Ministry of National Education developed MEBBIS both for practitioners, students and parents to use, Ministry of Health developed e-Nabız (literally e-Pulse), MHRS (Central Patient Appointment System), and in agency applications, Social Security Institution developed systems for health payments, UBYS (National Healthcare Information System) and vaccine tracking system. Some of these examples will be discussed here to explain detailed benefits of the systems and brought together to understand similar policies around the world . Official Statistics Plan 2017-2021 and 2018 Public ICT Investments Report suggests that the big data applications will increase in public agencies, mostly aiming to establish proper data warehouses in organizations and data standardization between organizations. However in this section some existing projects explained.

4.1.1 TUBITAK B3LAB (Cloud Computing and Big Data Laboratory)

TUBITAK BILGEM's current General Outlook on e-Government in Turkey 2017 report, states that, increasing data volume and data relations parallel with the e-government applications makes big data a primary research area in Turkey. (p. 19) By using these data, the organizations find the opportunities to use business strategies to fulfill their needs. B3LAB is the main supplier of the funds for nation-wide research projects on big data.

While developing the big data and cloud computing policies around the world, with the collaborating project of the TUBITAK BILGEM and Ministry of Development's investment funding a laboratory aiming to develop infrastructure and conduct research on big data and cloud computing era was founded. These two concepts are merged because according to B3LAB's web-page, "Cloud Computing and Big Data technologies manage the vast amount of data produced by various resources and provide cost and workforce efficiency by centralizing the online services that are increasingly becoming prevalent." As it has been mentioned, cloud computing and big data are working hand in hand. As claimed by the B3LAB, it aims to "construct a model for the transition to a national public cloud and to contribute to the data analytics towards the extraction of meaningful information for public welfare."

Other than educating the ministry personnel on what big data and cloud computing are, until now the laboratory tried to enrich big data and cloud computing by making Cloud Computing and Big Data Workshops organized in 2014 and 2015, The first National Cloud Computing and Big Data Symposium in 2017, and organizing the summer schools for big data and cloud computing realized in 2015 and 2016.

The Cloud Computing and Big Data Workshops are organized to determine where Turkey stands in cloud computing and big data technologies. The attendees are ranging from Ministry, Academia, NGO's and business sector representatives. The 2014 workshop worked under the Needs and Requirements, Solution to the Needs &

Requirements titles. The B3LAB working groups produced results about where organizations see themselves in big data; these results will be used under the discussion section as a secondary data.

B3LAB is using both machine learning and big data analytics to analyze Pendik Municipality Call Center Voice Records to get results on subjects and analyzed it through different needs of different neighborhoods.[27] The voice recognition system overlapped with the call center workers' categorizations. This sheds light on possible machine-learning based call center usage for local governments.

4.1.2 Ministry of Health

The vast amount of data is produced by the healthcare sector and how to use it via big data analytics is a recurring theme. (Groves et al., 2013; Raghupathi & Raghupathi, 2014; Jee & Kim, 2013; Ludvigsson, J. F., 2009) The strategic plans and documents of Turkey also show a tendency to merge health data since 2003 e-Transition Turkey project. Big Data Projects are varying from their, nascent state to proactive stage among the Turkish Public Administration System examples, one of the most developed is the Ministry of Health's multiple infrastructured, proactive databases.

4.1.2.1 Health Informatics Network

Ministry of Health is using the Health Informatics Network infrastructure to manage the influx of big data. The system has two datacenters with 426 databases of 330 terabytes of capacity in Ankara and Istanbul (Ülgü & Gökçay, 2017, p. 269) This system helps healthcare organizations and professionals to reach shared databases and communicate via a shared base. By doing so, the health databases are opened to multiple resources and big datasets are applicable through these systems. This network, also according to Mahir Ülgü (2017), the Chief Executive Health Informatics

[27] Yavuz, A.G (Oct, 2017) Retrieved 09.06.2018 from presentation: http://www.kamu-bib.org.tr/kamubib-19/sunumlar/ali-gokhan-yavuz-kamu-bib19.pdf

Officer, created a gateway for other public institutions that are using the e-government protocols. That means a network similar to DARPA-NET is created by the ministry and presented alternative infrastructure for other public institutions in case of emergency, if other ministries cannot use their systems they can also use this alternative network system to communicate. In this case, very progressive and multiple user approach is what creates the value of big data.

4.1.2.2 e-Pulse (e-Nabız)

The system is used by the patients to learn their past medical exams, diagnosis, treatments, lab test results and prescriptions. Moreover, they can track their health data on the system if they're using wearable technologies. Ministry of Health aims to improve health awareness among citizens by using this application. Moreover, both citizens and health professionals would easily reach personal health records so that retrospective analysis of a patient's health records can be seen by the health professionals. The system anonymizes the data for further analysis according to the ministry. The system also helps health professionals to see past medical treatments, controls and medical operations, if the citizen accepts the system also adds the data to a central database for further analysis. Ülgü and Gökçay (2017) explain that this central database would help both doctors and the citizens for a comparable analysis of medical records.

4.1.2.3 Health Management System

The system collects real-time data from data provider units and systems in order to maximize efficiency and productivity in monitoring, measurement, evaluation, policy making and decision making processes, and uses integrated platforms on executive levels to have meaningful ends (Ülgü & Gökçay, 2017, p. 269). That means the system is using data to have interactive analysis, gives executive public administrators meaningful information rather than presenting unstructured datasets, and also this analyses lead policy-makers create data-driven decision-making.

The system uses different databases such as, medical exam data, repeating application by patients, time analysis of medical exams, satisfaction levels, e-Nabız evaluations, birth and medical operation ages, service quality through the days of week, application levels in different cities, medical analysis count for each patient etc. These datasets lead the practitioners and public managers to have more evidence-based policies which in turn increase the health service quality according to Ülgü's analysis.

4.1.2.4 Spatial Work Intelligence Platform

A benefit of big data is using geospatial analysis to reach results (UN Survey, 2018). The system developed by the Ministry of Health gives results on maps to practitioners and executives for quick understanding of what big data analysis can get.

Ministry of Health's geospatial analytics is used on detailed levels; such as results from provinces, town, and village based dataset made out of numbers. Moreover, geospatial mapping of these data, similar to night lights example analysis of TEPAV (Başıhoş, 2016), creates new opportunities. Ranging from understanding effects of industrial zones and by combining geospatial temporal analysis other results can be reached.

Ülgü and Gökçay (2017, p.273) compares, monthly air pollution levels and their spatial work intelligence platform geospatial results and draws a correlation between the highest air pollution months and respiratory illness relation. On the other hand, it should be kept in mind that without analyzing differing datasets inferring results may not be meaningful. The system is also used to improve service quality, for example by analyzing patient migration from one province to another the executives can take action for specific districts.

For example, the system is used for Hypertension drug price analysis (Sarıyıldız & Akgündüz, 2017) the gives different results that in some districts hypertension medicine costs higher than counterparts, a global example was given from the UK;

some of the distributors may sell licensed drug to some districts which causes higher drug costs rather than generic drugs with same composition. Big Data analytics shows what there is not awareness of. This leads hoarding of databases to find better causation between events. As Dinov (2016) also underlined "Big healthcare data refers to complex datasets that have some unique characteristics, beyond their large size, that both facilitate and convolute the process of extraction of actionable knowledge about an observable phenomenon."

Whole systems are used for better health services as claimed by the developers. On the other hand, Adam Tanner, in his investigative book on health, "Our Bodies, Our Data" claims that healthcare data become a multi-billion business between governments, private healthcare institutions, and drug companies. Violation of privacy might lead to bio-identity thievery.

4.1.3 Social Security Institution

Social Security is one of the main areas of big data analytics use. Big data applications are used in many countries in order to prevent evasions by citizens, reduce costs and increase service efficiency. As explained throughout the history of the development of 2016-2019 e-Government Turkey Strategic plan, it puts SSI one of the agencies with priority.

According to the plan, SSI planned to use Data Silos to integrate its data, in the second phase planned to use data mining gave meaning to the inept data until using big data systems, and in the third phase it includes a big data infrastructure to use it on its systems (Şık, 2017; 285). According to the Şık (2017) data silos are prepared in 2008, but they're still in use. However, SSI did not pass to phase 2, data mining. The data silo is used for creating health, insurance, retirement report groups out of 500 standard reports (Köseoğlu, 2017, p.2231).

The major problem about big data analytics in Turkey is that creating data silos are akin to offline data of the past. The organizations do not use unstructured data to get better understanding and use the possibilities of data but rather store them in huge silos which has no interoperability other than creating merged reports out of databases. As described by TBD's Big Data report (2016), SSI has MEDULA, which shares transaction information of pharmacies, using e-Bildirge (e-Declaration) for businesses to conduct social security payments, and a direct phone line ALO 170 for information sharing with the citizens. MEDULA is a well-developed example of big data application due to gathering various databases and using for various purposes. ALO170, call line for getting information is at best a primitive e-Government project when compared to big data analytics. When compared with the Ministry of Health's various systems SSI understanding of the big data applications not as particularly developed from applications basis. The applications stay on a system of e-participation to reduce paperwork through internet governance. A good developed example could be using machine learning sound recognition algorithms to analyze ALO170 calls, similar to B3LAB's Pendik call line example.

4.1.4 Ministry of National Education

Education is one of the most common areas of big data policy development, because the area consists of huge data collection and traceability, and return rate of the data subjects are high. For example, the e-Okul (e-School) system is in use since 2007. It is used by the students to follow their report cards, exam results from kindergarten to graduation from high school. In the more developed portal MEBBIS, the objective is to integrate schools, students, teachers, parents, school administrators, local Meb Boards to use various web-page modules. MEBBIS plays an important role of a small ecosystem of turkey.gov.tr for Ministry of National Education. The system is a high in volume data silo for MEB e-government services consisting of different modules, producing report cards, human resources services, personnel education system, fiscal reports for schools and other various services. The advancement of MEBBIS, as expected by the Strategic Plan 2015-2018, incorporates possible e-government

applications under one portal. However, there is not a specific detailed academic output about MEBBIS's data mining properties or data analytics but it just works as an online portal and a data silo.

FATİH Project (Fırsatları Arttırma ve Teknolojiyi İyileştirme Hareketi/Increasing Opportunities and Improvement of Technology Project) started in 2011. Since then, 7 billion Turkish Liras has been spent on various hardware ranging from tablets, smart boards to printers for the digitization of the classroom environment. These devices do not simply presents big data but along with the project of School Information System, where a geospatial mapping made for the Ministry to follow development of the plan via this map. An Interactive Class Management (ESY) database is created for synchronized and asynchronized teaching with the students. Furthermore, likewise West (2013) mentioned in his article about personalized education possibilities, the system presents teachers online assessment capabilities of students and possibilities about student centric teaching. The Ministry of Education develops various databases and ICT plans under different projects. A current one suggests using big data analytics for assessing teacher and student performance by analyzing various datasets, such as exam results, average academic outlook of the students in class, teachers education etc.[28]

4.2 Analysis of Big-Data Related Policies in Turkey

4.2.1 A Data-Centric History of E-Government and Big Data in Turkey

Information Society Department under the Republic of Turkey Ministry of Development lists the official documents on e-government transformation on its webpage. There are national level action plans, specialist's theses, reports, specialization commission reports, development plans and there are international documents listed by EU, OECD, UNESCO, UNCTAD, ITU, World Economic Forum

[28] (n.d) (2018), MEB'den Big Data Projesi retrieved 27.07.2018 from https://www.turkiyeegitim.com/mebden-big-data-projesi-96726h.htm

reports related to Turkey. The history of the e-government in Turkey will be discussed through development process of these main policy documents via using subdocuments regarding their point of view on data even though big data have developed after 2010's in e-government applications (Altun, 2017). There are consistent development of legal framework and infrastructure on these policy documents (Köseoğlu, 2017).

4.2.1.1 Legal Framework for E-Government and Big Data Use in Turkey

Ministry of Development is responsible for developing strategic documents for integrating Information Society Strategy, for this reason in 2003 Department of Information Society was founded for the specific subject which creates the plans and implement them to the society. Although there is not any strict e-government legislation, in 2011 with the statutory decree no.655 Ministry of Transport, Maritime Affairs and Communication (now Ministry of Transport and Infrastructure) made responsible for developing and implementing e-government policies. (EU Report 2018, 12) The Department of Communications is responsible under the ministry to coordinate and supervise the objectives and strategies of the relevant public authorities. This divide is more obvious on strategic planning. To illustrate, 2015-2018 Information Society Strategy and Action Plan is prepared by the Development Ministry, 2016-2019 National e-Government Strategy and Action Plan is prepared by the Ministry of Transport, Maritime Affairs and Communications.

The legal regulations include, Right to Information Act(2003), Article 20 of Turkish Constitution, Law on Personal Protection of Data(2016), By-Law on Electronic Communication Security, By-Law on the Personal Information Processing and Privacy in the Telecommunications Sector (2004), Council of Europe's Convention on Cybercrime, Law No. 5070 on Electronic Signatures (2004), Ordinance on the Procedures and Principles Pertaining to the Implementation of Electronic Signature Law (2005), Law No. 4822 on Consumer Protection (2003), Electronic Communications Act (2008), Public Procurement Law No. 4734 (2002, 2008), Law on Regulating Broadcast in Internet and Combating Crimes Committed through Such Broadcast (2007)

4.2.1.2 Earlier examples 1990-2000

The first developments of e-government in Turkey begins with the technology transfer in 1980's. In 1993, with the help of World Bank, Turkey Informatics and Economic Modernization Report is prepared. This is the first official document to emphasize the information society.[29] The report underlines data management by stating "design and implementation of a National Database and Information Policy" (World Bank, 1993) and first to exemplify "poor dissemination of public data" .

According to Yıldız (2007, p. 647) "until the introduction of the internet and widespread use of personal computers, the main objectives of technology use in government were enhancing the managerial effectiveness of public administrators while increasing government productivity" but the introduction of the internet should be waited to benefit from the full-fledged capacities of e-government because IT use was internal and managerial before the internet (Yıldız, 2007: 648). Also earlier technology usage was rather peripheral than playing a core administration perspective (Yıldız, 2007b, p. 396)

This process also affected Turkey, although there are projects in Turkey such as MERNIS which started in 1972 for centralizing the population documents, the process took 30 years to complete transformed into a digital system in 1996 and most of the process was actualized from 1997 to 2002. VEDOP (Tax Office Automation Project) is one of the earliest examples of e-government in Turkey. The first incarnation of the system as a pilot project went online in 1998. VEDOP started as a country-level automation project for tax offices by the Ministry of Finance (EU 2018, 14).

In 1998 the Prime Minister Mesut Yılmaz declared the KAMU-NET project to centralize all government agencies systems to reduce the red tape work interoperable.[30]

[29] The Department of Information webpage verifies this, retrieved 28.06.2018 from http://www.bilgitoplumu.gov.tr/bilgi-toplumu/ulkemizde-bilgi-toplumuna-donusum/

[30] Bürokrasi Kamu-Net ile Tarihe Gömülüyor. (Feb, 23, 1998) retrieved 06.06.2018 from: http://www.hurriyet.com.tr/ekonomi/burokrasi-kamu-net-ile-tarihe-gomuluyor-39007462

The webpage went online in 1998, realized questionnaires about what the e-government portal's domain name should be.(DPT, 2002: 4) The project stopped in 2003 and e-government development initiative transferred to State Planning Agency according to Urgent Action Plan.(Yıldırım et al, 2006, 52) (EU 2018, 24) 2013-2014 National Cyber Security Strategic Plan re enforced the KAMU-NET to integrate all government agencies, this time the system transferred to Transportation and Infrastructure Ministry. The project is still under development as of 2018. Despite the early examples, the process developed throughout the years.

The concept of e-government likewise Yıldız's (2007) incorporation is realized through after the introduction of the internet, the first strategic plan on information structure that also concerns e-government policies was demanded by the Prime Minister's office in 1996 (TUENA) spanning 1996-1999 years was the first plan to use words "on-line government" The plan states various problems from administrative perspective. These are:

1. To use and integrate the IT in public organizations most important problem is regulations and redtape. This causes the intended technology to become dated. (Planning)
2. Each organization chooses the IT structure it finds the most suitable to its needs with given resources. (Organizing and Budgeting)
3. Training for the officers seems a useless expenditure, and the user experience develops through a trial and error process. (Staffing, budgeting)
4. Most of the data is still offline data. (Organizing)
5. The resource for skilled information technology officers is not enough and the transfer to the private sector is high, due to compared low wages and job satisfaction. (Staffing)
6. Interoperability is low. There should be a coordination center to create all public agencies work synchronized. (Coordinating)

The reason for TUENA's particular challenges given is because similar to the earliest policy problems, for example the staffing continues to be a problem in IT (Halper & Krishnan, 2018), structural problems persist under latest technologies (Lavertu, 2015). After 20 years of TUENA reports publishing, the report's finding on offline data is still a problem for Turkey's open data policies. Global open data index shows that among the 15 possible datasets, only 5 of them are in a machine-readable format as of 2018. [31]

4.2.1.3 e-Europe Plan and Policy Transfer to Turkey

The e-Europe Initiative, issued in December 1999, aims to establish the necessary infrastructure for the new economy, especially in the internet field in order to make Europe the most vibrant and competitive marketplace in the world. In 2000, along with Malta and Cyprus, Turkey is invited to be a part of the project (TBD, 2002, p. 22). According to EU Accession process of Turkey, in the National Plan of EU General Secretary, "establishment of e-Turkey project for information society" is stated.
In order to have a competitive, dynamic and a knowledge-based economy and to ensure the transition to the information society and to adapt the e-Europe Action Plans of Turkey, the Prime Minister's office launched e-Turkey initiative 10.09.2001 dated and 352 Circular No.

On May 2002, Undersecretariat of Prime Ministry, and various NGO's working on IT areas organized First Information Technologies Forum. The main titles from forum result report are infrastructure, e-economy, e-government, law, education, R&D.

According to these reports, e-government subtitle, "the state of Turkey's e-government is in a primitive stage" according to report "more than a hundred public agencies have webpages. These web pages include static information and there is not a proper informing of the citizens. Inter-organization communication, informing, organization

[31] https://index.okfn.org/place/tr/

is on very low level. The organizations are only good at giving information about the organization, but e-services are insufficient." (TBD, 2002, p. 230). The main founding on data is on "adapting to the new technologies" subtitle, the last article states as, "It is clear that the volume required for e-government services will be very large" (TBD, 2002, p. 245). Due to lack of technological development in the era the report projects that "the environment in which this data will be stored in a central database and the technological infrastructure required to provide distributed access should be prepared." Today's big data technology does not force to have centralized data centers. On the other hand mentioning the volume proponent of the big data in 2002 is earliest example of usage. The report through standardization and interoperability title mentions the transfer of unstructured and structured data, constant fluctuation of data with the systems and survivability of the systems.

The draft for e-Turkey Initiative Action Plan issued in August 2002. This document is the first official document specifically aims e-government policies (DPT, 2009, 3). According to the Information Society Department's webpage, the implementation of the plan was not possible due to economic and political instability of the era.

As part of Urgent Action Plan's Public Management Reform Section, coded as KYR-22 e-Transformation Turkey Project was declared as a high priority project. e-Turkey initiative action plan transformed into e-Transformation Turkey project in 2002. The main actor for e-government and Information Society transformation is declared as State Planning Agency in this plan and two processes are foreseen, first to create an action plan for e-transformation and transformation of legal regulations. Under the specific subject a detail is given as "each public agency will establish a data system and all the data will be open and the visibility of the system will increase." (T.C Başbakanlık, 2003, p. 27) This report restricts the open data only to state departments, does not mention opening datasets of government to the public. Uçkan (2003) states that Turkey should urgently revise its national policies on science and technology, launch a strong IT campaign and humanitarian and economic development must be on

a democratic basis. Each passing day causes both global and national level divides, and that this is not only a digital but also a humanitarian divide.

4.2.1.4 Short Term 2003-2004 Plan

e-Transformation Turkey Project Short Term Action Plan for 2003-2004 years is prepare on October 2003. This document lays 73 policies under 8 sections on information society strategy, legal infrastructure, technical infrastructure, education and human resources, standards, e-government, e-health, e-commerce. 23 of the articles are on e-government. When data protection laws are considered or when as a public policy the policies about "data" is under legal structure section and the other data words are aiming to centralize and give better health services.

In the second part of KYR-22 article legal infrastructure transformation, two things are done in the first year. The Consumer Protection Amendment Law came into force in 2003 and amendments regulated e-commerce (EU Commision 2014, p. 14). In October 2003, the Turkish Parliament enacted the Right to Information Act (Law No. 4982), on 26 April 2004 the law came into force aiming "to lay down the guidelines and procedures for individuals to exercise their right of information in accordance with the principles of equality, neutrality and openness which are the fundamentals of democratic and transparent administration.

The general result report prepared for the short-term action plan for the years 2003 and 2004 were published on January 2005. The report emphasizes that the projects on data sharing and metadata creation will continue for the agencies. Specifically on 2003-2004 action plan there is an action plan dedicated to The Economics and Feasibility of Internet Data Centers. The Reporter states that: "For large-scale projects such as e-Transformation Turkey; to combine, share and appropriately outsource their data on Internet data center services required to increase the efficiency and interoperability."(T.C UBAK, 2005, p. 28)

4.2.1.5 2005 Action Plan

On March 2005 e-Transformation Turkey Project 2005 Action Plan was issued. The projects about data management include creating data silos and integrating databases. For example, General Directorate of Security will "harmonize and integrate Car traffic and registration documents, tax and traffic fine databases" (DPT, 2005). By doing this Ministry of Finance, State Statistics Institute and General Directorate of Motorways work together. 10 out of 50 policies are about creating database for the state agencies. Article 35 "Data Sharing for Interoperability" to create a document "identifying e-government metadata standards on who can and how to access the information kept in public agencies and where information will be stored is going to be prepared. Data elements and data structures used in providing public services and necessary mechanisms for sharing data will be formed. Interoperability Framework Guide will be updated."(DPT 2005b, p. 12) The Result Report of the action plan states that a meta-data working group has begun (DPT 2006, p. 15).

4.2.1.6 2006-2010 Information Society Strategy and Action Plan

In 2006 a comprehensive plan for Information Society Strategy and Action Plan for the e-Transformation Turkey process is prepared spanning from 2006 to 2010. The report is published as two sections; first part is the Strategic Plan, which sets the common policies, and the second part is the Action Plan, which describes how to achieve these results by creating certain articles for state organizations. The general report merges and develops the previous plans. The Strategic plan foresees some data centric policies such as data silos for National Education Ministry (p. 31), data share between Justice Department and General Directorate of Security (p. 31), Census Bureau and General Directorate and Land Registry (p. 31), detecting the data ownership regulation(p. 35) This shows that the strategic plans are becoming more interested in the potential of using data.

Policies from the strategic plan show infrastructural developments whereas action plan has some proactive measures which interpolates with big data analytics. Action no.43 is a descriptive example of this, improving UYAP (National Judiciary Informatics Systems) for establishing decision support systems and taking protective measures by using the database and data analysis. There are common policies from the action plan (plans 48, 53 and 54) showing the awareness of using the benefits of big datasets and to create a hand over of data between the state agencies.

Article 81 under the Data and Document Management section states that: "A policy will be set for the reuse of public information generated by the public source by the state, businesses and individuals to create added value outside of commercial or for commercial purpose, and regulations will be made taking into account the EU Legislation." This is the first declaration of using open data among other action plans. In addition, the action document states that Turkey, as a policymaking process is aware of the possible advantages of open data for citizens, businesses and government. In 2012, the Final Report of 2006-2010 Information Society Action Plan prepared. Aforementioned policies such as, using predictive data analysis UYAP, or opening public information to the society is not mentioned in the final report.

4.2.1.7 2015-2018 Information Society Action Plan & 2016- 2019 e-Government Action Plan

Increasing importance of global usage of big data showed itself directly as a policy on these plans. The strategic importance of big data has been realized after the initiation of the plan. E-government Action Plan states a clear action "E1.3.2- Developing and Extending Big Data and Internet of Things Policies in Public Sector".

The action plan underlines the importance of e-Government Research Program to use at better decision making and to do so draws this roadmap "In first step, service models, data-driven decision making mechanisms, institutional memory/information management, Public innovation, open source software usage, big data applications and

green informatics will be primary areas."(p.56) This plan also overlaps with 412.plan in 10th 5 year Development Plan of Turkey, similar article stated in the development plan. Also, evidence-based action and policy making for education and employment is planned. To do this, the report foresees huge integrated big data centers for agencies. The report under the 240th article, emphasizes "a roadmap for the digital transformation of the industry" and tries to adapt Turkey Industry 4.0 standards by expecting it a nationalized production strategy for technologies such as AI, IoT, Robotics, additive manufacturing, augmented reality.

2016-2019 e-Government Action Plan anticipates the development of Big Data and Internet of Things policies via analyzing different service areas such as education, health, social security, transportation, taxation, work life and security are the earliest mentioned policy areas. In contrast to the mentioned countries, these policy areas seem to be well chosen.

Kitchin (2017) defines these areas for governments as, improvements to public administration and cost savings through enhanced 'operational efficiency, a reduction in the cost of errors and fraud in benefit administration, and an increase in tax receipts by narrowing the tax gap improved allocation of funding into programs, higher-quality services, increased public sector accountability, [and] a better-informed citizenry' and "concerns state security and the tackling of crime. All states are involved in surveillance for the purposes of security, safety and crime prevention and apprehension through policing and wider intelligence-gathering" (p. 154)

The 2016-2019 report incorporate Kitchin's arguments and expects results for big data usage under six categories: improving the public services via data analytics, using structured and unstructured data from the public organizations to increase efficiency, integrating public data and analyzing information security problems by data analytics, using IOT's to improve efficiency in service sector for e-government process, improving public fiscal management through big data analytics. Preventing informal economy, tax fraud through big data analytics.

While creating big data policies in Turkey, the report focuses on three purposes; creating the projects with a big data analysis driven decision making, staffing combined teams working on these projects and informing citizens and creating public opinion awareness on how big data analytics will be beneficial for them.

2015-2018 Information Society Plan article 193 specifies lack of public sector initiatives, finds that some process is achieved for productivity, reducing tax loss and improving service quality by using high volume data, but it is not as developed as the business sectors big data usage in Turkey. By referencing Japan in article 100, it suggests to use preventive data analytics for disaster management. Most importantly, it understands how big data's market value increasing in article 98, market value will reach 53,4 billion USD in 2017 and plans to use it on new employment possibilities. One catch is a structural problem for IT staff everywhere in the world, because qualified analysts hard to find and sustain their job satisfaction, the article 63 comprehends this problem. Turkey, as planning wise states common themes about big data implications, on the other hand foresees 10k datasets to be open to public on an open data portal, also in 2016-2019 E.421 foresees open data portal to be opened. As of July 2018, the open data portal http://www.resmiistatistik.gov.tr does not have the merged statistics on the webpage. State Statistics Agency webpage hosts these statistics.

The Information Society Action Plan gives responsible agencies on big data public implementation agencies as, Social Security Institution, Ministry of Health, Scientific and Technological Research Council of Turkey. The projects of these organizations on big data will be given.
There are also technical reports such as National Broadband Strategy and Action Plan 2017, that specifically plans to build big data centers to develop broadband networks (Strategy 3, 17) to do this, legislations will be changed.

4.3 Assessment of Big Data Maturity in Public Administration in Turkey

The following section first evaluates existing studies on big data maturity in Turkey, and then evaluates its place in each of the six dimensions of Kuraeva model on the basis of the policies reviewed in the previous section.

4.3.1 Evaluation of the Existing Studies Conducted by TBD and B3LAB

TUBITAK B3LAB made a workshop in 2015 and conducted a survey including 70 public employees from various agencies. Details of this study are provided in the Appendix 1. One of the questions they asked was, "Do you need data analysis in your organization?". 28 of them gave "yes" answer, 18 gave "no" and 24 of them said "I don't know". According to Asay (2014), Head of Developer Ecosystem in Adobe, one of the reasons of big data failure is the "management resistance". More than 60% of managers says that they trust their guts and real-world insight is better than data-analytics. This means that value awareness of big data is low on the organization.
In another question, "Is there any planned data analytics project in your organization?" 39 of the correspondents answered "YES", 14 of them answered "NO" and 17 of them answered "I don't know."

On big data maturity assessment, the most important part about starting to use analytics is being aware of the technology. As explained in previous chapter about Turkey, both current strategy, action and development plans are aware of the possibilities of big data. The catch is the organizational adoption process. As can be referred from the B3LAB's results from 2015, organizational awareness stays on 40% level.
According to B3LAB research, when the needs of the organizations considered, there are 7 problems stated by the public employees,

1. Lack of skilled personnel (55 people)
2. Emerging data from different sources (54 people)
3. Awareness (44 people)
4. Data Analysis speed & performance (43 people)

5. Data query performance (35 people)
6. Infrastructural Costs (29 people)
7. Other (14 people)

The highest chosen answer coincides with Asay's (2014) "8 reasons of failed big data policies". "Lacking the right skills", the inadequate number of data analysts and "asking the wrong questions" means organizations cannot find suitable skilled personnel even though they find skilled data analysts. Second highest choice "emerging data from different sources" also overlaps with Asay's model.

Similar to B3LAB's research, TBD (Informatics Association of Turkey) made a research on public organizations. Details of the study are provided in Appendix 2. The third question of their questionnaire asks "Are you interested in big data as an organization?"

13.8% answers as "no interest in big data", 10.3% answers "interest in research level" , 27,6% answers as "beyond researching level there are ideas about projects", 31% answered as "planning on big data applications", 17,2% of answers includes the "project realization and data investment" level interests. This shows that among public institutions awareness of data analytics is increasing. It should be kept in mind that all of these answers actually only on the "Awareness" level of big data maturity models. A further question about whether the organization has a big data strategy returns with different results, 4.1% answers there is not a strategic planning, 34.5% answers "emphasizing the necessity of strategic work on big data", 20.7% answers "big data is one of the long term goals", and 17.2% answers as "big data analytics is one of the short term goals". A quarter of the organizations are complete darkness for Radcliffe (2014), rest of the attending organizations are exploring, and understanding their place in big data analytics.

For example a data company Zaloni's data maturity model the first section "Ignore" consists of "Data Warehouse" part, this means that the organization uses it's own data

silo's and benefits from only structured data. Let us remind that, public organizations in Turkey uses the exact same wording with the first level of data warehouse, data silo literally "veri ambarı" Department of Justice, Department of Finance has these data warehouses. This level of big data maturity level means increased storage costs and low level of analytics. One of the prominent examples from Turkey is Ministry of Health, as described in previous chapter, by using data centers, the ICT Directorate of Ministry of Health, created an alternative data storage facility which also can be used by other public agencies, also by using different data analytics and geospatial mapping they can use it for getting better decision-making and efficiency. Turkish Ministry of Health's level is on the "Governed Data Lake" level since machine learning of Responsive Data lake is not possible as of now (Ülgü & Gökçay, 2017) and rather than responsive analytics the system shows descriptive analytics through structured datasets.

Related to this developmental process TBD's research results on data storage shows that 69% of organizations are using data warehouses in one organization, 51.7% is using its own stratified system and in need integrates the data in warehouse. Which means 20.7% uses both techniques. As big data maturity models explained, both of these techniques are not enough to benefit from the usage of data analytics.

The organizations want to benefit from big data but, they want to do it without sharing information, having strict boundaries in and between other public organizations. However, Asay (2014) says "big data is more valuable to an organization if the walls between groups come down and their data flows together. Politics or policies often hinder this promise." As explained in Turkish Public Administration chapter, what happened to shared database of KAMU-NET since 1998 is another microcosm of these events. Bysgtad et al. (2017) suggests creation of integrated solutions rather than central IT Silos, because standardization and integration of data leads decreased expenses. When the amount of budgets spared on big data analytics asked to the personnel, TBD Study results show that there is a significant amount of budgeting for more than public organizations in Turkey for big data projects.

As a result from these analytics, it can be deduced that Turkey, on the level of awareness reaches a well-developed point via strategic and action plans, but the researches suggests that there are still organizations who are unaware of the data analytics applications. Also, stacking data into data silos are common policies by the government organizations, which results inefficient usage of highly valuable and collected data. Even though Action Plans suggests developing big data strategies, there is not a coherent and open big data policy to take in action.

Turkish private sector likewise its counterparts in the world developed asymmetrically compared to the public organizations. The spreading of R&D institutions are increasing this means an institutionalized awareness of big data. If a state agency wants to develop a big data analytics plan, or if it is imposed by central administration suitable project payments made handsomely.

4. 3. 2 Analysis of Big Data Maturity in Turkey on the Basis of Kuraeva Model

Among the big data maturity models reviewed in the earlier section, this study adopts the one that was developed by Kuraeva (2016) to be used in public administration context. According to the model, big data maturity is assessed on the basis of six dimensions: vision and strategy, open data initiatives, R & D institutions and initiatives, big data maturity level in business sector, data governance, and big data projects experience in public sector. Similar to other stage-way models, the model is dispersed into four levels: awareness, exploring, optimizing, and *transforming.*

4.3.2.1 Vision and strategy

Turkey has both Information Society Strategy and Action Plan and National e-government Strategy and Action Plan which draws a general outline of ICT plan on Exploring stage. The awareness is set by 2015-2018 Information Society Action Plan no.45:

45. Volume of digital data is expected to increase 44-fold in the next ten years. Such a big volume of data has an enormous potential for advancements in productivity, cost minimization, service provision and product development. In fact, OECD considers big data driven innovation as a new source of growth. Big data market is expected to grow with an annual average rate of 60 percent and reach 53.4 billion USD in 2016. Demand will increase in areas such as nonrelational database systems, business intelligence and data analytics applications. Developed countries are working on new educational and employment policies in order to meet the demand for qualified labor required by big data and investing in new R&D programs(15-18 Action Plan,p. 20).

Big Data usage entered to the 2016-2019 National e-government action plan but there is no "existence of insights from big data applications" since there were limited usage of big data applications at the time of action plans development.

Official Statistics Program 2017-2021 emphasizes using databases for natural geospatial mapping via Corine in Ministry of Forest and Water Management which provides an integration to the EU data standardization enforcement. As a whole big data strategy is defined in very abstract manner in latest Information Society Action Plans and there were limited applications. For this reason Turkey is on Optimizing (3 points) part of the vision and strategy, since there is no indication of transformation policies explained by the model, such as existence of clear usage of big data.

4.3.2.2 Open Data Initiatives

There are some government open data[32] such as air quality, national statistics, government budget etc. Waseda University a leading organization in e-government stated in its report that TURKSTAT posts government data regularly on its website. These statistics come from a variety of government ministries, and can be downloaded in Excel format. It also criticizes that while the site hosts a large amount of data,

[32] For the portal the link is: https://index.okfn.org/place/tr/

particularly economic data, it does not have advanced searching, charting, or organizational features. This means that it is not developed into a wide array of datasets and/or integrated to a nationwide open data portal. Some of the past open data is not updated and machine readability is a problem. On Municipal databases, Şahinbey Municipality developed a project on open data which earned them an award in eTurkey Awards in but the web-page is down since 2017, there is not a nationwide open data on municipality manner. For this reason Turkey is Exploring (2 points) in open data initiatives.

Although Turkey in 2012 Turkey prepared Open Government Action plan for the Open Government Initiative, in September 21, 2016 Turkey defined as inactive by the Open Government Partnership[33] due to being unable to deliver an Open Government Action plan since 2014, a year later Turkey's participation ended.[34]

4.3.2.3 R & D Initatives

There are Big Data/Data Analytics M.S programs provided by universities such as İstanbul Technical University, Sabancı University, Bahçeşehir University, TED University but except TED all programs are only limited to non-thesis paid M.S programs. There is not a government backed data analytics or big data program. There are research initiatives such as B3LAB directly created for cloud computing and big data analytics by government's TUBITAK, and Gazi University founded a Big Data and Information Security Center Research Lab. Although these research initiatives founded by government prepares workshops, academic research and projects they're limited to these two institutes there are not variety of institutions. As can be inferred from these areas Turkey is still exploring (2 points) in its databases.

[33] Letter for Inactivity on OGP, retrieved 04.04.2018 from: https://www.opengovpartnership.org/stories/turkey-made-inactive-open-government-partnership

[34] Letter for ending of OGP participation retrieved 04.04.2018 from: https://www.opengovpartnership.org/documents/september-2017-letter-informing-ending-of-turkeys-participation-ogp

4.3.2.4. Big Data Maturity in Business Sector

There are examples from communication sector such as Turkcell and Anadolu Jet Geospatial mapping example, by using anonymized customer travelling signals, Turkcell provides Anadolujet rush hour times for better flight timings. TTNET Intellimap[35] using big data analytics to provide business sector automated SMS sending in certain geospatial hot places. The system used for better shop placement for the company. Also the company opened the biggest datawarehouse in Turkey. Koç's information system company Koçsis, provides IoT and Big Data analytics solutions. Also KoçSistem integrated it's infrastructure for e-bill,e-defter and other Revenue Administration policies. Tüpraş funded a data analytics center in ODTÜ for academic industrial collaboratioN A subsidiary of Anadolu Grubu, Anadolu Bilişim A.Ş provides its companies such as Efes Pilsen A.Ş big data analytics for better services. Ereteam, a big data analytics company provides services for Akbank, Allianz, British American Tobacco, Garanti Bank, Koçtaş, Migros, Superonline, TEB(a subsidiary of BNP Paribas) Vakıfbank, Yapı Kredi, Zorlu business intelligence and big data applications.[36] A general outlook gives, likewise the rest of the world businesses are aware of the importance of big data analytics, they're working with consulting firms to develop solutions. When taken as a whole, the pioneering big data analytics, solutions and predictions are done by businesses. Also, companies like SAMPAŞ, Universal, provides smart city and other big data solutions for public agencies. For this reason Big Data Maturity in Business Sector is in optimizing (4 points) situation. The big companies are aware and producer of the analytics, they have solutions for predicting outcomes and the big data analytics ranges from Petrol, Food, Communication, Banking and retail sectors. Their corporate maturity levels are not assessed since whole business sector is reviewed.

[35] https://kurumsal.turktelekom.com.tr/mobil/servisler/sayfalar/intellimap-big-data.aspx

[36] ftp://public.dhe.ibm.com/software/pdf/tr/events/2013/p1/Ereteam/Ereteam_Brosur_2.pdf

4.3.2.5. Data Governance

According to EU Accession process, Data ownership and rights of the data owner is described by the Turkish Personal Data Protection law no.6698. Data standardization process is continuing according to 2017-2021 National Statistics Program (NSP) in various areas, such as forest metadata is planned improve into EU standards. The report states that:

"The third NSP (2017-2021), aims data integration, increase the use of administrative records in statistical production, data security and data confidentiality, opening administrative registers of real and legal entities to TÜİK for standardizing the data identifiers and variables."

Data from sources are still created and entered via piecemeal fashion, time schedules are not fixed. For example Census Bureau's birth records are fixed retrospectively each year because of distributed data usage. Social Security Institution's data input is irregular and fragmental for this reason proper analysis is not possible on real time basis. There are limited collaborations between public agencies, for example, it is still hard to receive data from Ministry of Justice for General Directorate of Security and plans fall apart after integration. When taken in consideration all of these points it can be said that state agencies are exploring (2 points) data governance.

4.3.2.6. Big Data Projects in Public Sector

There are projects by Ministry of Health, Ministry of Finance, Ministry of National Education as explained in Turkish Public Administration example section. But except the action plans a wide usage of big data applications is not consistent. However big data applications still stays in most of the strategic and action documents but is not used widely for this reason it is still on exploring (2 points) part.

4.3.2.7 Overall Evaluation

Kuraeva's model gives each dimension a point from 1 (aware) to 4 (transforming) and suggests a general outlook of a country by results.Accordingly, results from the above analysis *when added together* suggest a total 15 point*s* of "Optimizing" level for big data in Turkey. The highest points come from business models like many other countries.

It should be noted that Kuraeva's model comes with its disadvantages. For example, to have a better understanding of where an institution is, IT Staffing Norms should be kept in mind. As mentioned before it is hard to find capable data scientists for government organizations both as salary satisfaction and public sector integrity. Another problem is Legal Framework. Existence of legal framework in a country is one of the foremost signifiers of Vision and Strategy. The state agencies differentiated budgets for data analytic strategy is another part of the problem

CHAPTER 5

CONCLUSION AND DISCUSSION

5.1 Summary of the Findings and Discussion

This study is a preliminary attempt to evaluate where Turkey stands in its use of big data applications. To this end, a literature review is done and general outline of what big data is and its techniques are listed, big data maturity models for organizations are explained. Then different examples are listed from different countries and Turkey regarding use of big data analytics for the purpose of public agencies. Big data maturity level of Turkey is also evaluated according to Kuraeva model, using a review of big data policies in Turkey. In addition, data taken from B3LAB and TBD studies are evaluated to draw a general outline of where Turkey stands on its big data policies and implementations.

There is a saying in Turkish, which is attributed to the Germans "Start like a Turk, finish like a German" to emphasize starting doing something with high enthusiasm but also to do it in an organized manner up until the end. Turkish legal documents, action plans and project documents are all prepared with top-notch policy designs and academic integrity. They are on par with their worldwide counterparts. There are thousands and thousands of pages of analysis and recommendations (TUENA report being 3397 pages long is a good example). Then there are yearly Project Development Reports of these action plans, which are again prepared with the highest academic integrity. There is an interesting problem with these reports, year by year they tend to repeat what has not been done due to some conditions or they tend to change the wording and use the same explanations next year. At the end of the short-term or middle-term action plan they merged into one general outlook of what has not been

done and goodwill are explained for the next plans. However this is not a problem only in Turkish Public Administration since Heeks (2005) and Marr (2015) explained more than half of the e-government policies are failed to achieve success. There are some examples which are executed well and became a successful project, such as MERNIS of the Census Bureau of Turkey. MERNIS in itself has a potential to draw a policy feedback loop for many other big data applications projects, but the policy documents does not draw a meaningful line of process. The documents assign a general distribution of pilot projects on organizations they find and try to achieve success. However there are organizations that need big data analytics more than others. The public agencies should be picked in case of big data analytics needs and applications possibilities by comparing international counterparts, other than investing each department a big data budget.

5.2. Recommendations

One of the striking things about well-developed policies is that, in either the U.S Federal Big Data Research and Development Strategic Plan or in Australia's Big Data Action Plan, the problems are defined precisely and who will achieve what and how is described precisely. When the 2015-2018 Information Society Strategy and Action Plan of Turkey is considered, it is clear that the committee preparing the plan is aware of the value of big data and importance of IoT. Then specific plan no.45 tries to achieve many things while creating pilot applications that are scattered across different areas. This approach affects implication of the projects because everything is planned to be achieved in a short time. Similar to the creation of the National Broadband Strategy and Action Plan (2017-2020) specifically a Big Data Strategy and Action Plan creation is suggested.

Another important point about big data is that, since it is the principal device of the data-driven decision-making, there are low levels of awareness by the institutions in Turkey, and there are examples of managers both in Turkey and in the world who trust their intuition. Although intuition and experience is effective, they tend to anchor on

human infallibilities. For this reason importance of data-driven decision-making should be emphasized for better results in public organizations.

Since benefits of big data are achievable through data integrity and interoperability, one of the problems of Turkey is the lack of a National Data Center (Afyonluoğlu, 2018). Turkey's Integrated Public Data Center is proposed in 2013, as of 2018 it has not been created. Being a latecomer in digital era causes both economical disadvantages each passing day and also digital gap between the world and Turkey widens.

Research showed that, Turkish public organizations are hesitant about sharing their data. There might be security reasons, and institutional culture also may be a part of it, but interoperability is on very low levels and as described by the questionnaires conducted by TBD, even public data collection and measurement is not standardized. Some organizations uses data warehouses, others uses different stratified data analysis systems. For both achieving maturity and integrating these data, public data complex is needed.

Open data is another important part of big data projects as explained in opportunities of using it. Unfortunately Turkey is not a member of Open Government Initiative since September 21, 2017. Even though TÜİK releases mostly financial data a need for integrated open data portal still continues if Turkey wants to achieve it's open data policy goals explained in the 2015-2018 Program. Most of the big data initiatives are described with counterpart open data but Turkey has limited itself with strictly big data usage. For this reason, only on the opportunities area and mandatory places it is repeated.

The legal framework for Turkey is developed well through the EU Accession process. Even structural steps are made and a Personal Data Protection Board established. This is another signifier of Turkey being better at the legal transformation basis was the 2003 Emergency Action Plan described the importance of legal transformations to

develop an e-government plan and in a year's time planned actions have been taken on legal basis.

The lack of skilled IT personnel is the most consistent problem of public organizations. Data analysts are harder to employ since lack of the personnel supply due to data management being a developing area, and due to higher private sector salaries, public organizations are suffering from proper staffing.

The maturity models developed by various scholars and countries suggest a step-by-step approach to define where a country is with regard to big data use. The most developed big data maturity models end up becoming automatized constantly analyzing, adapting systems, which create continuous evaluation in the policy cycle. Although there are private IT firms like Facebook and Google which have this kind of developed policies; there are very few governmental organizations achieving this in the world, one of them being US Army Automated Continuous Evaluation System and UK Government Program on Performance Data (Höchtl, 2016). For this reason most of the governments are still adapting big data technologies.

This study showed that, Turkey has awareness of big data but it also has an advanced private sector to provide useful means for big data applications. It is possible for Turkish government to achieve better ends in this situation. To do so, again the importance of big data action and strategy plan has to be prepared for each organization with a strict and clearly defined plan.

5.3 Limitations of the Study

One of the challenges in this study was that, the topic of big data covers wide variety of disciplines on different areas. For example from computer science perspective, Hadoop, NoSql and similar technologies are explained but MongoDB, Cassandra, Pig and many others could not be covered. Because the thesis aims to draw a general framework for public administration, most commonly used are mentioned.

Another limitation is that there are many other countries which are using big data in public administration such as South Korea and Romania, but counting up all of the countries one by one would lead a flock of examples, rather most developed examples and comparable examples are given from the world. At some point the applications begin to repeat between countries and also developing countries began to follow similar approaches, for this reason country examples are limited both by availability of the documents, academic output about big data application about that country and specific language barrier of the country also affected. Chinese example is given because of its outlier type of application on privacy of its citizens and constant media discussions created enough literature to explain about China.

To apply Kuraeva's model to analyze big data awareness in Turkey, secondary data is used. More diverse and deeper analysis could be made through conducting interviews with the informatics officers of organizations in Turkey.

5.4 Future Research

There are various maturity models under descriptive, prescriptive and comparative models. In this thesis, one of the most commonly used examples is adopted to have a general outlook. A more detailed analysis using different maturity models or developing a new maturity model could be some ways to advance this research.

This thesis tries to draw a general outlook of Turkey in the world via informing the reader about technologies, policies and adoption possibilities on big data. A further comprehensive field research conducted with the IT department heads of public organizations via using comparative maturity model developed by the researcher accordingly would give a detailed point of view about big data maturity picture about Turkey.

Another future research on this subject can be about comparative analysis of different countries to develop better policies for similar looking countries. Since there is an

increasing interest in big data applications in public administration in Turkey, detailed policymaking procedures on open data is also possible to be studied.

Customer data is already used for improving sales of companies. Loyalty Cards of retail stores analyzes particular consumers' habits to suggest personalized discount rates and campaigns to them. This is called nudging, a concept related with decision-making, uses behavioral science to improve decision making of the citizens. UK's Ministry of Health started to use nudging to improve citizens' health choices. Governments are nudging citizens for a quite long time via various ways. The majority of public policies aim to frame and facilitate citizens behavior. Citizens, communities and policymakers, stop 'bad behaviors, steering taxation is one of the oldest ways of nudging. Increasing tax on unhealthy foods or decreasing tax levels on electric cars to encourage citizens are some examples, via using big data analytics behavioral science creates consent and steer people from libertarian paternalistic point of view. A further study on big data analytics and behavioral science relation with the public administration would shed a light on government analytics.

REFERENCES

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, *16*(1), 3-9.

Altun T., Şaşin, F., & Öztaş, N. (2017). Kamu Politikalarının Belirlenmesinde ve Uygulanmasında Büyük Veri, *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, *22*.

Anna, K., & Nikolay, K. (2015). Survey on Big Data Analytics in Public Sector of Russian Federation. Procedia Computer Science, 55, 905-911.

Anderson, C. (2008), "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", Wired Magazine, www.wired.com/science/discoveries/magazine/16-07/pb_theory/

Afyonluoglu, M. (2018). Ulusal Veri Merkezi ve Türkiye: Nasıl?, IDC Kamu'da Dijital Dönüşüm ve Veri Merkezi Konferansı 27.03.2018 Ankara

Alexopoulos, C., Loukis, E., & Charalabidis, Y. (2014). A Platform for Closing the Open Data Feedback Loop based on Web2.0 functionality. JeDEM, 6(1), 62–68.

Australian Government Information Management Office. (2013). The Australian Public Services Big Data Strategy, (August), 1–27.

Başbakanlık, T. C. (2002). 58. Hükümet Acil Eylem Planı. http://www.sbb.gov.tr/Lists/EylemVeDigerPlanlar/Attachments/13/58.H%C3%BCk%C3%BCmetAcilEylemPlan%C4%B1.pdf accessed: 19.07.2018

Başıhoş, S. (2016). Gelişmişlik Göstergesi Olarak Gece Işıkları: Ulusal Ölçekte ve İl Bazında GSYH Tahmini. Tepav Tartışma Metni, 23, 2017.

Bates, J. (2013). The domestication of open government data advocacy in the United Kingdom: A neo-Gramscian analysis. Policy and Internet, 5(1), 118–137

Batty, M. (2013). Big data, smart cities and city planning. Dialogues in Human Geography, 3(3), 274-279.

Bekkers, V., & Homburg, V. (2007). The myths of e-government: Looking beyond the assumptions of a new and better government. *The Information Society*, *23*(5), 373-382.

Bensghir. TK,(2017) Türkiye'de Büyük Açık Veri: Politik ve Yasal Çerçeve. In Sağıroğlu, Ş & Koç, O., (Eds.), Türkiye'de Büyük ve Açık Veri Analitiği: Yöntemler ve Uygulamalar (1.Ed) Ankara, 2017

Bertot, J. C., & Choi, H. (2013). Big data and e-government. Proceedings of the 14th Annual International Conference on Digital Government Research. https://doi.org/10.1145/2479724.2479730

Bizer, C. (2009). The emerging web of linked data. *IEEE intelligent systems*, (5), 87-92.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: Four perspectives — Four challenges. SIGMOD Record, 40(4), 56–60. http://dx.doi.org/10. 1145/2094114.2094129

Boyd, D. & Crawford, K. (2012). Critical Questions for Big Data. Information, Communication & Society, 15(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Braun, H. (2015). Evaluation of Big Data Maturity Models -a Bench- Marking Study To Support Big Data Maturity Assessment in Organizations.

Buchholz, S., Bukowski, M., & Sniegocki, A. (2014). Big and Open Data in Europe: A growth engine or a missed opportunity?, 116.

Buttarelli, G. (2016). The EU GDPR as a clarion call for a new global digital gold standard.

Caldwell K (2014) Are you next on the taxman's hitlist? The Telegraph. 10 December. Available at: www.telegraph.co.uk/finance/personalfinance/tax/11092959/HMRC-targets-Are-you-next-on-the-taxmans-hitlist.html (accessed 14 July 2018).

Cavoukian, A., & Jonas, J. (2012). *Privacy by design in the age of big data* (pp. 1-17). Information and Privacy Commissioner of Ontario, Canada.

Chandan, H. C. (2016). Corruption, organisations and culture in contemporary India. In *Indian Culture and Work Organisations in Transition* (pp. 145-161). Routledge India.

Chang, W. L. (2015). *NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*(No. Special Publication (NIST SP)-1500-3).

Chen, C.L., & Zhang, C. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf. Sci., 275*, 314-347.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Cook, T. D. (2014). "Big data" in research on social policy. *Journal of Policy Analysis and Management*, *33*(2), 544-547.

Cox, M., & Ellsworth, D. (1997, August). Managing big data for scientific visualization. In *ACM Siggraph* (Vol. 97, pp. 21-38).

Crosier, S. (2001). John Snow, The London Cholera Epidemic of 1854. CSISS Classics.

Cukier, K. (2010). Data, data everywhere: A special report on managing information. Economist Newspaper.

da Mota Ueti, R., Espinosa, D. F., Rafferty, L., & Hung, P. C. (2016). Case studies of government use of big data in Latin America: Brazil and Mexico. In *Big Data Applications and Use Cases* (pp. 197-214). Springer, Cham.

Das, A., Singh, H., & Joseph, D. (2017). A longitudinal study of e-government maturity. Information and Management, 54(4), 415–426. https://doi.org/10.1016/j.im.2016.09.006

Davenport, T., & Harris, J. (2017). *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. Harvard Business Press.

Desouza, K. C., & Jacob, B. (2017). Big data in the public sector: Lessons for practitioners and scholars. *Administration & Society*, *49*(7), 1043-1064.

Dinov, I. D. (2016). Volume and value of big healthcare data. *Journal of medical statistics and informatics*, *4*.

DPT(2002), e-Devlet'e Geçiş Sürecinde KamuNet Çalışmaları
http://www.bilgitoplumu.gov.tr/wp-content/uploads/2014/04/e-Devlete_Gecis_Surecinde_KAMU-NET_Calismalari.pdf accessed: 17.07.2018

DPT (2003) eAvrupa Girişimi, Bilgi Toplumu Dairesi Başkanlığı http://www.bilgitoplumu.gov.tr/Documents/3/Diger/000323_eAvrupaGirisimiDetaylari.pdf accessed: 17.07.2018

DPT (2005), e-Dönüşüm Türkiye Projesi, 2003-2004 KDEP Uygulama Sonuçları

DPT(2005), e-Transformation Turkey Project 2005 Action Plan, March 2005
http://www.bilgitoplumu.gov.tr/wp-content/uploads/2014/04/Action_Plan_2005.pdf

DPT (2006), e-Dönüşüm Türkiye Projesi 2005 Eylem Planı Sonuç Raporu

DPT. (2009), E-Dönüşüm Türkiye Projesi, Bilgi Toplumu Dairesi Başkanlığı http://bilgitoplumu.gov.tr/Documents/1/Icra_Kurulu/090715_IK27.ToplantisiE-DonusumTurkiyeProjesiGelismeler.pdf

Dunleavy, P., Margetts, H., Bastow, Tinkler, J.; New Public Management Is Dead—Long Live Digital-Era Governance, *Journal of Public Administration Research and Theory*, Volume 16, Issue 3, 1 July 2006, Pages 467–494, https://doi.org/10.1093/jopart/mui057

Dwivedi, Y. K., Weerakkody, V., & Janssen, M. (2011). Moving Towards Maturity : Challenges to Successful Implementation and Diffusion. The DATA BASE for Advances in Information Systems, 42(4), 11–22. https://doi.org/10.1145/2096140.2096142

El-Darwiche, B., Koch, V., Meer, D., Shehadi, R. T., & Tohme, W. (2014). Big data maturity: An action plan for policymakers and executives. *The global information technology report*, *43*, 51.

EU Commision (2014). *Digital agenda for Europe : rebooting Europe's economy*. Luxembourg: Publications Office

EU Commision. (2014) Turkey e-Government Report retrieved 17.06.2017 from: https://joinup.ec.europa.eu/sites/default/files/document/201406/eGov%20in%20TR%20May%202014%20v.11.0.pdf

EU Commision. (2018). Turkey e-Government Report. https://joinup.ec.europa.eu/sites/default/files/inline-files/eGovernment_in_Turkey_2018_1.pdf accessed: 20.07.2018

Fernandes, L. M., O'Connor, M., & Weaver, V. (2012). Big data, bigger outcomes. *Journal of AHIMA*, *83*(10), 38-43.

Fisher, D., Drucker, S. M., & König, A. C. (2012). Exploratory visualization involving incremental, approximate database queries and uncertainty. *IEEE computer graphics and applications*, *32*(4), 55-62.

Franks, I. M., & Nagelkerke, P. (1988). The use of computer interactive video in sport analysis. *Ergonomics*, *31*(11), 1593-1603.

Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (2017). Big data and social science : a practical guide to methods and tools. Agri Marketing. https://doi.org/10.1002/ajh.23643.

Fredriksson, C., Mubarak, F., Tuohimaa, M., & Zhan, M. (2017). Big data in the public sector: A systematic literature review. *Scandinavian Journal of Public*

Administration, *21*(3), 39-62. Retrieved from http://130.241.16.45/ojs/index.php/sjpa/article/view/3452

Gamage, P. (2016). New development: Leveraging 'big data'analytics in the public sector. *Public Money & Management*, *36*(5), 385-390.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137-144.

Gänßlen, S., & Losbichler, H. (2014). Big Data-ein Segen für das Controlling. *White Paper des Internationalen Controller Vereins, Wörthsee*. (via Google translate)

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, *2007*(2012), 1-16.

Giest, S. (2017). Big data for policymaking: fad or fasttrack? Policy Sciences, 50(3), 367–382. https://doi.org/10.1007/s11077-017-9293-1

Gong, J., Heaslip, K., McNeil, S., Farzan, F., & Brink, S. (2016). *Big data: opportunities and challenges in asset management* (No. CAIT-UTC-030). Department of Transportation In cooperation with Rutgers, The State University of New Jersey.

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'big data'revolution in healthcare. *McKinsey Quarterly*, *2*(3).

Halper, F., & Krishnan, K. (2013). TDWI big data maturity model guide interpreting your assessment score. *TDWI Benchmark Guide*, *2014*, 2013.

Hammer, C., Kostroch, M. D. C., & Quiros, M. G. (2017). *Big Data: Potential, Challenges and Statistical Implications*. International Monetary Fund.

Heeks, R. (2003). Most eGovernment-for-Development Projects Fail: How Can Risks be Reduced? IGovernment Working Paper Series, (14), 1–30. https://doi.org/10.1017/CBO9781107415324.004

Heeks, R. (2009). The ICT4D 2.0 manifesto: Where next for ICTs and international development? Development Informatics, Working Paper Series. https://doi.org/10.1016/0736-5853(84)90003-0

Heeks, R. B. (1999). Software Strategies in Developing Countries. Communications of the ACM, 42(6), 15–17. https://doi.org/10.1145/303849.303853

Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., ... & Zwitter, A. (2017). Will democracy survive big data and artificial intelligence. *Scientific American*, *25*.

Helms, J. (2015). Five examples of how federal agencies use big data. IBM Center for the Business of Government blog. Retrieved from http://www. businessofgovernment.org/blog/business-government/five-examples-how-federalagencies-use-big-data

Hinkson, I. V., Davidsen, T. M., Klemm, J. D., Chandramouliswaran, I., Kerlavage, A. R., & Kibbe, W. A. (2017). A comprehensive infrastructure for big data in cancer research: accelerating cancer research and precision medicine. *Frontiers in cell and developmental biology*, *5*, 83.

Ho, A. T. K., & Mccall, B. (2016). *Ten actions to implement big data initiatives: A study of 65 cities*. Center for The Business of Government.

Howard, A., (2011) How Data and Open Government Transforming the NYC, O'Reilly Radar
http://radar.oreilly.com/2011/10/data-new-york-city.html acessed: 20.07.2018

Höchtl, J., Parycek, & Schollhammer. (2016). Big data in the policy cycle: Policy decision making in the digital era. Journal of Organizational Computing.

Janowski, T. (2015). Digital government evolution: From transformation to contextualization.

Janssen, M., Estevez, E., & Janowski, T. (2014). Interoperability in big, open, and linked data-organizational maturity, capabilities, and data portfolios. Computer, 47(10), 44–49. https://doi.org/10.1109/MC.2014.290

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?.

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. Journal of Business Research, 70, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Jarmin, R. S., & O'Hara, A. B. (2016). Big data and the transformation of public policy analysis. *Journal of Policy Analysis and Management*, *35*(3), 715-721.

Jee, K., & Kim, G. H. (2013). Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthcare informatics research*, *19*(2), 79-85.

Kalkınma Bakanlığı (2006) 2006-2010 Bilgi Toplumu ve Stratejisi Eylem Planı

Kalkınma Bakanlığı (2012) 2006-2010 Bilgi Toplumu ve Stratejisi Eylem Planı Sonuç Raporu

Kalkınma Bakanlığı (2013). Onuncu kalkınma planı (2014-2018). *Ankara: Kalkınma Bakanlığı*.

Kalkınma Bakanlığı (2015). Bilgi Toplumu Stratejisi ve Eylem Planı 2015-2018, http://www.bilgitoplumustratejisi.org/t r/doc/8a9481984680deca014bea4232490005

Kalkınma Bakanlığı (2017) 2017-2021 Ulusal Genişbant Stratejisi ve Eylem Planı

Raghupathi, Wullianallur, and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." *Health information science and systems* 2.1 (2014): 3.

Kazuyuki, Motohashi (2017). *Survey of Big Data Use and Innovation in Japanese Manufacturing Firms* (No. 17027). Research Institute of Economy, Trade and Industry (RIETI).

Kearns, M. J. (2010). Designing a digital future: Federally funded research and development in networking and information technology.

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3), 78–85. https://doi.org/10.1145/2500873

Kim H, P., & Dunnigan, R. (2014). Big Data Analytics: A practical Guide for Managers. https://doi.org/10.1017/CBO9781107415324.004

Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences, 1–192. https://doi.org/10.4135/9781473909472

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1-14. https://link.springer.com/article/10.1007/S10708-013-9516-8

Klievink, B., Romijn, B.-J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. Information Systems Frontiers, 19(2), 267–283. https://doi.org/10.1007/s10796-016-9686-2

Köseoğlu, Ö., & Demirci, Y. (2017). Türkiye'de Büyük Veri Madenciliğine İlişkin Politika ve Stratejiler: Ulusal Politika Belgelerinin İçerik *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, *22*.

Kubler, K. (2017). State of urgency: Surveillance, power, and algorithms in France's state of emergency. *Big Data & Society*, *4*(2), 2053951717736338

Kuhn, J. (2012). Open government : Who participates and why ?, 72, 1–72.

Kuraeva, A (2016). Big Data Analysis Influence on Public Administration Processes (Master's thesis). Retrieved from https://www.hse.ru/en/edu/vkr/182647584

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, *52*(2), 21.

Lavertu, S. (2015). We All Need Help:"Big Data" and the Mismeasure of Public Administration. Public Administration Review, 76(6), 864–872. https://doi.org/10.1111/puar.12436

Lo, F. (2015). Big Data Technology: What is Hadoop? What is MapReduce? What is NoSQL.

Lourenco, W., Author, B., & Manoelbatista, A. (2017). Analysis of Studies on Applications and Challenges in Implementation of Big Data in the Public Administration. International Journal on Recent and Innovation Trends in Computing and Communication, 5(5), 751–759. Retrieved from: http://www.ijritcc.org/download/browse/Volume_5_Issues/May_17_Volume_5_Issue_5/1496295494_01-06-2017.pdf%0A

Lowman, M. (2017). *A practical guide to analytics for governments : using big data for good*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Lyon, D. (2002). Everyday surveillance: Personal data and social classifications. *Information, Communication & Society*, *5*(2), 242-257.

Maciejewski, M. (2017). To do more, better, faster and more cheaply: using big data in public administration. International Review of Administrative Sciences, 83(1_suppl), 120–135. https://doi.org/10.1177/0020852316640058

Macintosh, A., Coleman, S., & Lalljee, M. (2005). E-methods for public engagement: helping local authorities communicate with citizens. *The Locale-Democracy National Project, Bristol City Council. Disponible en< http://www. iidi. napier. ac. uk/c/grants/grantid/7191263*.

Makowski, G. (2017). From Weber to the Web… Can ICT Reduce Bureaucratic Corruption?. In *Beyond Bureaucracy*(pp. 291-312). Springer, Cham.

Malhado, A. C., & Ladle, R. J. (2010). New data system to galvanize Brazil's conservation efforts. *Nature*, *465*(7300), 869

Manning, P. (2013). Big data in history. Big Data in History. https://doi.org/10.1057/9781137378972

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.

Marr, B. (2015). *Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance*. John Wiley & Sons.

Marr, B. (2016). *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*. John Wiley & Sons.

Mayer-Schönberger, V., & Lazer, D. (2007). Governance and Information Technology: From Electronic Government to Information Government (Chapter 1). Governance and Information Technology : From Electronic Government to Information Government, 329. https://doi.org/10.1080/19331680802428630

Mayer-Schönberger, V., & Ramge, T. (2018). *Reinventing Capitalism in the Age of Big Data*. Basic Books.

Mayer-Schönberger, V., Ramge, T., (n.d.). Viktor Mayer-Schönberger, Thomas Ramge - Reinventing Capitalism in the Age of Big Data (2018, Basic Books).

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.

McCue, C. (2014). Data mining and predictive analysis: Intelligence gathering and crime analysis. Butterworth-Heinemann.

McNulty, E. (2014). Understanding big data: the seven V's. *Dataconomy, May*, *22*.

Michael, K., & Miller, K. (2013). Big data: New opportunities and new challenges. Computer, 46(6), 22–24. https://doi.org/10.1109/MC.2013.196

Milakovich, M. (2012). *Digital governance: New technologies for improving public service and participation*. Routledge.

Mills, S., Lucas, S., Irakliotis, L., Rappa, M., Carlson, T., & Perlowitz, B. (2012). Demystifying big data: a practical guide to transforming the business of government. *TechAmerica Foundation, Washington*.

Morabito, V. (2015). Big data and analytics: Strategic and organizational impacts. Big Data and Analytics: Strategic and Organizational Impacts, 1–183. https://doi.org/10.1007/978-3-319-10665-6

Mullich, J. (2013). Closing the big data gap in public sector. *SURVEY REPORT—Real-Time Enterprise,(Sep. 2013)*.

Munné R. (2016) Big Data in the Public Sector. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham

Nolan, R. L. (1979). Managing Crises of Data Processing. *Harvard business review*, *3*(4).

Nolan, R. L. (1973). Managing the computer resource: a stage hypothesis. Communications of the ACM, 16(7), 399–405. https://doi.org/10.1145/362280.362284

Norheim-Hagtun, I., & Meier, P. (2010). Crowdsourcing for crisis mapping in Haiti. *Innovations: Technology, Governance, Globalization*, *5*(4), 81-89.

O'Reilly III, C. A. (1982). Variations in decision makers' use of information sources: The impact of quality and accessibility of information. *Academy of Management journal*, *25*(4), 756-771.

Pirog, M. A. (2014), Data Will Drive Innovation in Public Policy and Management Research in the Next Decade J. Pol. Anal. Manage., 33: 537-543. doi:10.1002/pam.21752

Podesta, J., Pritzker, P., Moniz, E. J., Holdren, J., & Zients, J. (2014). Big Data: Seizing Opportunities. Executive Office of the President of USA, (May), 1–79. https://doi.org/10.5121/ijgca.2012.3203

Pöppelbuß, J., & Röglinger, M. (2011). What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. Ecis, (August 2014), Paper28. Retrieved from http://aisel.aisnet.org/ecis2011/28/

Pulse, U. G. (2012). Big data for development: Challenges & opportunities. *Naciones Unidas, Nueva York, mayo*.

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), 1-34.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public administration*, *80*(1), 1-22.

Salas-Vega, S., Haimann, A., & Mossialos, E. (2015). Big data and health care: challenges and opportunities for coordinated policy development in the EU. *Health Systems & Reform*, *1*(4), 285-300.

Seifert, J. W. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. Government Information Quarterly, 21, 461−480.

Sengupta, R., Heeks, R., Chattapadhyay, S., & Foster, C. (2017). Development Informatics Working Paper Series Exploring Big Data for Development: An Electricity Sector Case Study from India. Retrieved from http://hummedia.manchester.ac.uk/institutes/gdi/publications/workingpapers/di/di_wp66.pdf

Simmons, R. (2015). Constraints on evidence-based policy: insights from government practices. *Building research & information*, *43*(4), 407-419.

Smith, J., & Sweetman, A. (2009, August). 4 Putting the evidence in evidence-based policy. In *Roundtable Proceedings*.

Starosielski, Nicole. *The undersea network*. Duke University Press, 2015.

Stough, R., & McBride, D. (2014). Big data and US public policy. *Review of Policy Research*, *31*(4), 339-342.

Şık, A., (2017). Sosyal Güvenlik Kurumu Büyük Veri Çalışmaları. In Sağıroğlu, Ş & Koç, O., (Eds.), Türkiye'de Büyük ve Açık Veri Analitiği: Yöntemler ve Uygulamalar (1.Ed) Ankara, 2017

TBD, Türkiye Bilişim Şurası (2002). Bilgi Toplumuna Doğru: Türkiye Bilişim Şurası Raporu. *Türkiye Bilişim Şurası*.

Tanner, Adam, Our Bodies, Our Data: How Companies Make Billions Selling Our Medical Records (Beacon Press, January 2017), http://www.beacon.org/Our-Bodies-Our-Data-P1249.aspx.

TechAmerica Foundation: Federal Big Data Commission. (2012). A Practical Guide To Transforming The Business of Government, 1–40. https://doi.org/10.1109/MCSE.2011.99

Tomar, L., Guicheney, W., Kyarisiima, H., & Zimani, T. (2016). *Big Data in the public sector: Selected applications and lessons learned*. Inter-American Development Bank

Townsend, Anthony M. *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company, 2013.

Uçkan, Ö. (2003). *E-devlet, E-demokrasi ve Türkiye* (Vol. 1). Literatür Yayıncılık.

UK Department for Business Innovation and Skills. (2013). Seizing the data opportunity: A strategy for UK data capability, (October), 1–51.

Ulaştırma ve Altyapı Bakanlığı (2005), E-DÖNÜŞÜM Türkiye KDEP-2004 7 Numaralı Eylem Raporu, Teknik Altyapı ve Bilgi Güvenliği Çalışma Grubu, Ankara

Ulaştırma Denizcilik ve Haberleşme Bakanlığı (2016). 2016-2019 Ulusal E-Devlet Strateji ve Eylem Planı

Ulaştırma Bakanlığı (1999) Türkiye Ulusal Enformasyon Altaypısı Anaplanı Sonuç Raporu

UN, E. (2018). government Survey 2018. *E-Government in Support of Sustainable Development/UN Department of Economic and Social Affairs*.

United States. Executive Office of the President, & Podesta, J. (2014). *Big data: Seizing opportunities, preserving values*. White House, Executive Office of the President.

United States Department of State Information Techology Strategic Plan 2017-2021

Ülgü, M. M., & Gökçay Ö. G. (2017). *Sağlık Bakanlığında Büyük Veri Çalışmaları*, In Sağıroğlu, Ş & Koç, O., (Eds.), Türkiye'de Büyük ve Açık Veri Analitiği: Yöntemler ve Uygulamalar (1.Ed) Ankara, 2017

Velkova, J. (2018). Studying Emerging Data Practices: Creating a Cultural Biography of Objects Through Using the Web as an Ethnographic Resource.

Viktor, M. S., & Kenneth, C. (2013). Big data: A revolution that will transform how we live, work, and think. *Houghton Mifflin Harcourt*.

Waseda University. (2017) The 13th Waseda - IAC International Digital Government Rankings 2017 Country Report, online, http://e-gov.waseda.ac.jp/pdf/2017_Country_Report.pdf

West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. *Governance studies at Brookings*, *4*(1).

West, D. M. (2015). Connected learning: How mobile technology can improve education. *Center for Technology Innovation at Brookings. Retrieved March*, *25*, 2016.

Willetts, D. (2013). *Eight great technologies*. Policy Exchange.

World Bank (1993), Turkey Information Technologies and Modernisation Report

Yildiz, M. (2007). E-government research: Reviewing the literature, limitations, and ways forward. Government Information Quarterly, 24(3), 646–665. https://doi.org/10.1016/j.giq.2007.01.002

Yiu, C. (2012). The Big Data Opportunity. Policy Exchange. Retrieved from http://www.policyexchange.org.uk/images/publications/the big data opportunity.pdf

Yu, H., & Robinson, D. G. (2012). The New Ambiguity of "Open Government." Ssrn, 178(2012), 178–208. https://doi.org/10.2139/ssrn.2012489

Yu, H., & Robinson, D. (2012). *The New Ambiguity of "Open Government"*. *Princeton CITP*. Yale ISP Working Paper.

Zainal, N. Z., Hussin, H., & Nazri, M. N. M. (2016, November). Big Data Initiatives by Governments--Issues and Challenges: A Review. In *Information and Communication Technology for The Muslim World (ICT4M), 2016 6th International Conference on* (pp. 304-309). IEEE.

Zheng, Y. (2017). Explaining Citizens' E-Participation Usage: Functionality of E-Participation Applications. *Administration & Society*, *49*(3), 423-442.

Zook, M. (2017). Crowd-sourcing the smart city: Using big geosocial media metrics in urban governance. *Big Data & Society*, *4*(1), 2053951717694384.

Zuiderwijk, A., & Janssen, M. (2014, June). The negative effects of open government data-investigating the dark side of open data. In *Proceedings of the 15th Annual International Conference on Digital Government Research* (pp. 147-152). ACM

.

APPENDICES

A: B3LAB 2015 BIG DATA WORKSHOP ATTENDEES PARTICIPATING IN THE SURVEY STUDY

1 Adnan Menderes Üniversitesi
2 Anadolu Ajansı
3 Atılım Üniversitesi
4 AVEA
5 Bankacılık Düzenleme ve Denetleme Kurumu
6 Bilgi Güvenliği Derneği
7 Bilgi Teknolojileri ve İletişim Kurumu
8 Bilkent Üniversitesi
9 Boğaziçi Üniversitesi
10 Datameer
11 Devlet Hava Meydanları İşletmesi Genel Müdürlüğü
12 Diyanet İşleri Başkanlığı
13 Doğuş Üniversitesi
14 Erciyes Üniversitesi
15 Ereteam
16 Fırat Üniversitesi
17 Forrester
18 Gebze Teknik Üniversitesi
19 Gediz Üniversitesi
20 Gelir İdaresi Başkanlığı
21 Hakimler ve Savcılar Yüksek Kurulu
22 Harran Üniversitesi
23 IDC
24 İstanbul Büyükşehir Belediyesi
25 İstanbul Teknik Üniversitesi
26 İŞKUR
27 İzmir Büyükşehir Belediyesi
28 Jandarma Genel Komutanlığı
29 Karayolları Genel Müdürlüğü
30 KAREL
31 Kocaeli Büyükşehir Belediyesi
32 Kocaeli Üniversitesi
33 Kredi ve Yurtlar Kurumu
34 Maden Tetkik ve Arama Genel Müdürlüğü
35 Netaş
36 Maliye Bakanlığı
37 Melikşah Üniversitesi

38 Necmettin Ebakan Üniversitesi
39 ÖSYM
40 Pamukkale Üniversitesi
41 Pendik Belediyesi
42 RTÜK
43 Sabancı Üniversitesi
44 Savunma Sanayii Müsteşarlığı
45 TEİAŞ
46 Teknopark İstanbul
47 Telekomünikasyon İletişim Başkanlığı
48 TÜBİTAK
49 Türk Hava Yolları
50 Türk Patent Enstitüsü
51 Türk Standartları Enstitüsü
52 Türk Telekom
53 Türkiye Büyük Millet Meclisi
54 Türkiye Cumhuriyeti Cumhurbaşkanlığı
55 Türkiye Cumhuriyeti Yargıtay Başkanlığı
56 Türkiye Cumhuriyeti Adalet Bakanlığı
57 Türkiye Cumhuriyeti Aile ve Sosyal Politikalar Bakanlığı
58 Türkiye Cumhuriyeti Başbakanlık Afet ve Acil Durum Yönetimi Başkanlığı
59 Türkiye Cumhuriyeti Başbakanlık Sermaye Piyasası Kurulu
60 Türkiye Cumhuriyeti Bilim Sanayi ve Ticaret Bakanlığı
61 Türkiye Cumhuriyeti Çalışma ve Sosyal Güvenlik Bakanlığı
62 Türkiye Cumhuriyeti Enerji Piyasası Düzenleme Kurumu
63 Türkiye Cumhuriyeti Gençlik ve Spor Bakanlığı
64 Türkiye Cumhuriyeti Gümrük ve Ticaret Bakanlığı
65 Türkiye Cumhuriyeti İçişleri Bakanlığı
66 Türkiye Cumhuriyeti İçişleri Bakanlığı Nüfus ve Vatandaşlık İşleri Genel Müdürlüğü
67 Türkiye Cumhuriyeti Kalkınma Bakanlığı
68 Türkiye Cumhuriyeti Kültür ve Turizm Bakanlığı
69 Türkiye Cumhuriyeti Merkez Bankası
70 Türkiye Cumhuriyeti Milli Savunma Bakanlığı
71 Türkiye Cumhuriyeti Sağlık Bakanlığı
72 Türkiye Cumhuriyeti Sayıştay Başkanlığı
73 Türkiye Cumhuriyeti Tarım Bakanlığı
74 Türkiye İstatistik Kurumu
75 Türkiye Kamu Hastaneleri Kurumu
76 Vakıfbank
77 Vodafone
78 Yıldırım Beyazıt Üniversitesi
79 Yıldız Teknik Üniversitesi
80 Ziraat Bankası

B: TBD SURVEY STUDY BIG DATA QUESTIONNARE AND RESULTS

Soru 1. Kurumunuzun çalışma alanı nedir?

Ankete katılan kurumların dağılımı aşağıdaki şekilde gösterilmiştir. Ankete, Ulaştırma, Belediye, Eğitim, Ekonomi ve Sağlık kurumları başta olmak üzere 18 kamu kurumu katılmıştır.

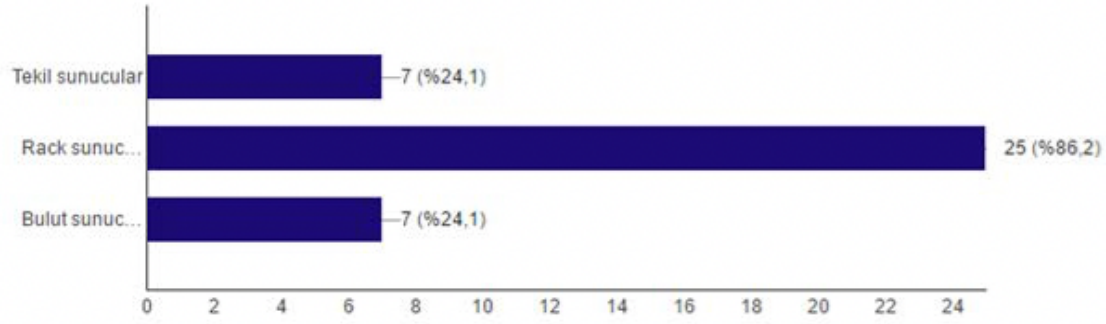
Soru 2. Verilerinizi nerede tutuyorsunuz?

Kamu kurumlarında uygulama verilerinin nerede tutulduğu ve çoktan seçmeli çok seçimli soruna alınan cevaplarda, kamu kurumlarının

- %86,8'i rack sunucularında,
- %24,1'inin tekil sunucularda,
- %24,1'inin ise bulut sunucularında

barındırıldığı sonucu ortaya çıkmıştır.

29



Soru 3. Kurum olarak “büyük veri” konusuna ilginiz var mı?

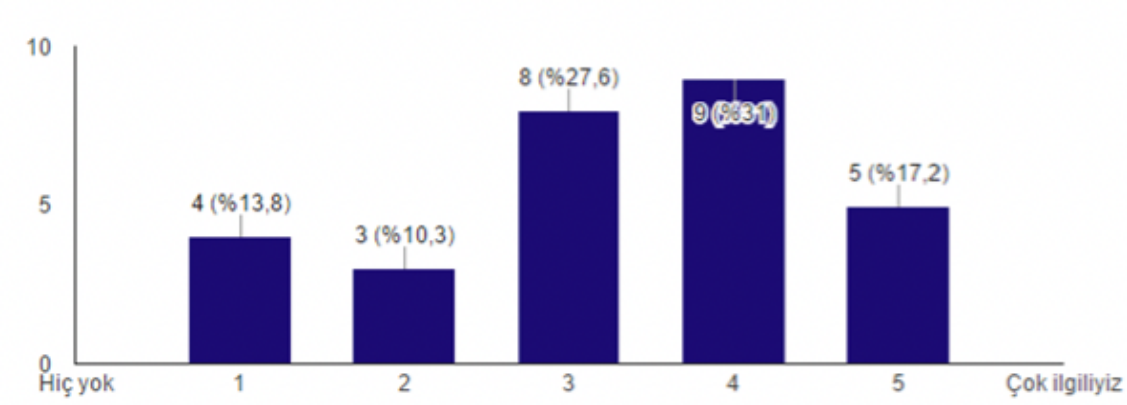
Kamu kurumlarının büyük verilerin işlenmesi ve büyük veri projelerinin gereksinim oluşturması noktasında kurumlardan aldığımız geri dönüşler sonucunda kurumların:

- %13.8’inin büyük veri konusuna ilgi duymadıkları,
- %10.3’ünün araştırma seviyesinde ilgi duydukları,
- %27,6’sının araştırma seviyesinin ilerisine gidip proje fikirlerini değerlendirme

seviyesinde ilgi duydukları,
- %31’inin büyük veri konusunda proje planlaması yapacak seviyede ilgi

duydukları,
- %17,2’sinin proje gerçekleştirme veya büyük veri yatırımı yapacak seviyede

ilgi duydukları sonucu ortaya çıkmıştır.



Soru 4. Kurum olarak “büyük veri” stratejiniz var mı?

Kamu kurumlarının büyük verilerin işlenmesi ve büyük veri projelerinin kurum iş süreçlerine etkisini ölçerek uzun vadeli bir strateji oluşturması noktasında kurumlardan aldığımız geri dönüşler sonucunda kurumların;

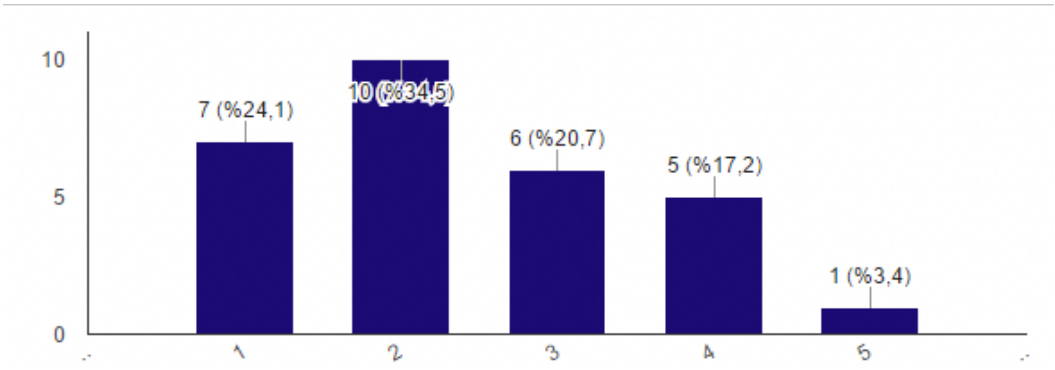
- %24.1’inin büyük veri konusunda stratejik bir çalışma yapmadığı,
- %34,5’inin büyük veri konusunda stratejik çalışma yapılmasının gerekliliği

üzerinde durduğu,
- %20,7’sinin büyük veri konusunda stratejik çalışma yapılmasını uzun vadede

planladığı,
- %17,2’sinin büyük veri konusunda stratejik çalışma yapılmasını kısa vadede

planladığı,

● %3,4'ünün büyük veri konusunda stratejik çalışmasının bulunduğu sonucu ortaya çıkmıştır.



Soru 5. Kurumunuzdaki verinin büyüklüğü nedir?

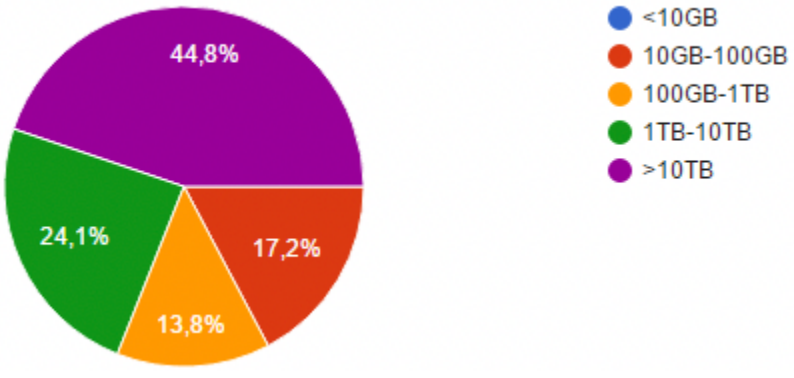
Kamu kurum uygulamalarında oluşan ve her geçen gün büyüyen verilerin anketin yapıldığı tarihi itibariyle büyüklüğünün tanımlanması istenmiştir. Kurumlardan alınan cevaplara göre kamu kurumların;

● %44,8'inin verileri 10 TB'ın üzerinde,
● %24,1'inin verileri 1 TB ile 10 TB arasında,
● %13,4'ünün verileri 100 GB ile 1 TB arasında, ● %17,2'sinin verileri 10 ile 100 GB arasında,
● %0'ının verileri 10 GB'ın altında

olduğu sonucu ortaya çıkmıştır. Kurumların verilerinin her geçen gün katlanarak büyümesi ve veri kaynaklarının çeşitlenmesi ile büyük veri projelerine ihtiyacı artacaktır.

Soru 6. Kurumunuzda yapısal ve yapısal olmayan veri büyüklüğü oranları nedir?

Kamu kurumlarında veri büyüklüğünün yanı sıra verilerin yapısal olup olmadığı da büyük veri uygulamaları için önemli bir faktördür. Yapısal verilerin işlenmesi



noktasında birçok kolaylık var iken yapısal veriler ile yapısal olmayan verilerin birlikte işlenmesinin gerektiği durumlarda devreye giren büyük veri uygulamalarının verileri anlık ve daha az maliyetle daha hızlı işleyebilme noktasındaki avantajları göz önünde bulundurulduğunda, kurumlar için daha fazla işlevsel uygulamaların geliştirilmesine olanak sağlayabilir.

Kamu kurumlarından alınan cevaplara göre;
● %3.4'ünün verilerinin tamamının yapısal olduğu, ● %34,5'inin verilerinin %80 yapısal olduğu,
● %17,2'sinin verilerinin %60 yapısal olduğu,
● %27,6'sının verilerinin %40 yapısal olduğu
● %17,2'sinin %20 yapısal olduğu,
● %0'ının %0 yapısal olduğu sonucu ortaya çıkmıştır.

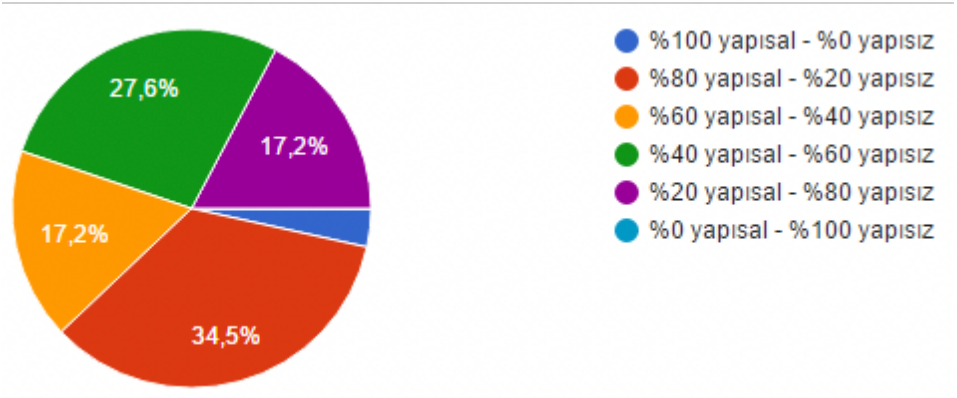
Kurum verilerinin yapısallık oranları incelendiğinde kamu kurum verilerinin büyük bir yüzdesinin yapısal olmadığı ve işlenmesinde zorluklar yaşandığı tespit edilmiştir. Yapısal ve yapısal olmayan verilerin uzun vadeli planlamalar ile büyük veri yapıları ile kullanılması, kurumların işleyişlerini daha etkin hale getirecektir.

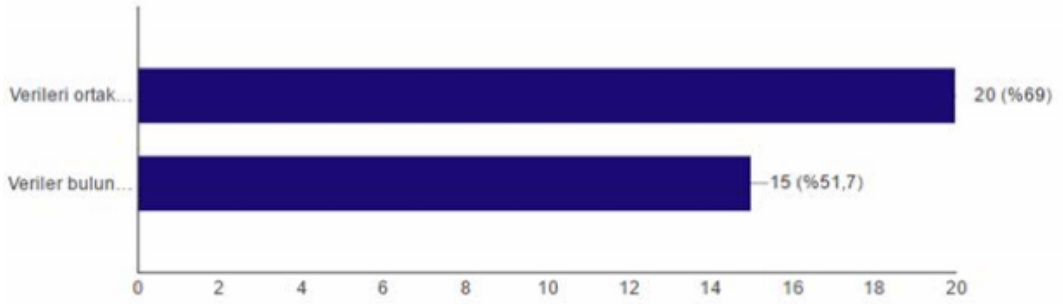
Soru 7. Kurumunuzda veri işlemede aşağıdaki hangi yaklaşımı uyguluyorsunuz?

Kamu kurumlarının verileri işleme yaklaşımların araştırılırken tüm proje verilerin ortak (merkezi) bir sistemde mi barındırdıklarını yoksa bulundukları sistemlerden gerektiğinde bütünleştirme ile mi alınarak işlem yapıldığı sorusundan alınan cevaplara göre kurumların;

● %69'u verileri ortak bir sistemde (warehouse, veritabanı gibi) toplayıp işliyor ● %51,7'si verileri bulundukları sistemlerde işleyip gerektiğinde bütünleştirme

(entegrasyon) yapıyor sonuçları ortaya çıkmıştır.



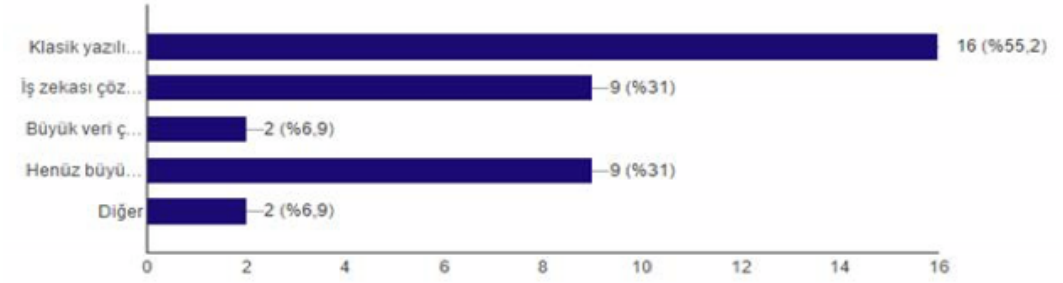


Soru 8. Büyük ölçekli veri işlemede hangi yöntemleri kullanıyorsunuz?

Kamu kurumlarının büyük miktardaki verilerini işleme yöntemlerine bakıldığında her ne kadar iş zekası çözümlerinin oranı artmış olsa da klasik yazılım geliştirme yöntemleri ile işleme en sık kullanılan yöntem olmaya devam etmektedir. Kurumlardan alınan cevaplara göre kurumların;

● %55,2'si klasik yazılım geliştirme yöntemleri ile
● %31'i iş zekası çözümleri ile
● %6,9'u büyük veri çözümleri ile
● %6,9'u bu üç yöntemin dışında farklı çözümler ile verilerin işler iken ● %31'i henüz büyük ölçekli veri işlemediğini belirtmiştir.

Bu istatistik incelendiğinde büyük veri kullanan kurum sayısının çok düşük bir oranda kaldığı, klasik yazılım geliştirme yöntemlerine olan alışkanlığın bir süre daha devam edeceği gözlemlenebilir.

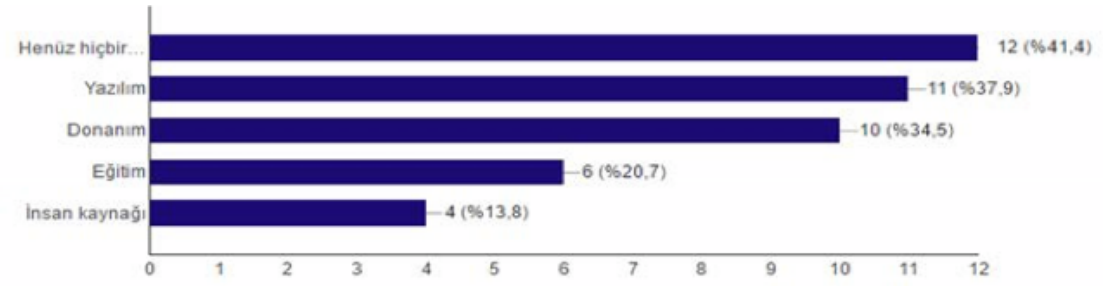


Soru 9. Büyük veri konusunda kurumunuzda hangi yatırımlar yapıldı?

Kamu kurumlarının büyük veri konusundaki yatırımları gelecek planlamaları için önemli bir yer tutmaktadır. Kurumların bu konudaki yatırımlarının sorulduğu soruya gelen cevaplara göre kurumların;

● %37,9'u yazılım yatırımı,

● %34,5'i donanım yatırımı,
● %20,7'si eğitim yatırımı,
● %13,8'i insan kaynağı yatırımı yaparken ● %29,9'u henüz hiçbir yatırım yapmadığını



Soru 10. Büyük veri yatırımlarınızın yaklaşık değeri nedir?

Büyük veri konusunda yapılan yatırımların değeri sorulduğunda gelen cevaplara göre kurumların;

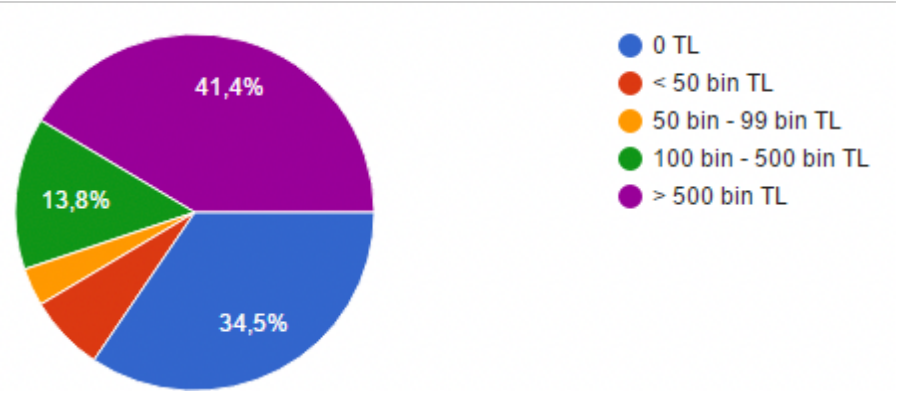
● %41.4'ü 500.000'in üzerinde,
● %13,8'i 100.000 ile 500.000 arasında, ● %3.4'ü 50.000 ile 99.000 arasında
● %6,9'u 50.000'den az yatırım yaparken ● %34.5'u hiç yatırım yapmadığını

belirtmiştir.

Soru 10. Büyük veri projeniz var mı?

Kamu kurumlarının büyük veri projelerinin sayılarını sorduğumuz soruya gelen cevaplara göre kurumların;

● %6,9'unun 5'in üzerinde,
● %17,2'sinin 2-5 arasında,
● %10,3'ünün 1 projesi var iken
● Kalan %65,5'inin ise hiç projesinin olmadığı belirtilmiştir.



Bu sorunun cevapları önceki sorularla karşılaştırıldığında kamu kurumlarının veri madenciliği, iş zekası, yüksek boyuttaki yapısal verilerden alınan raporların büyük veri

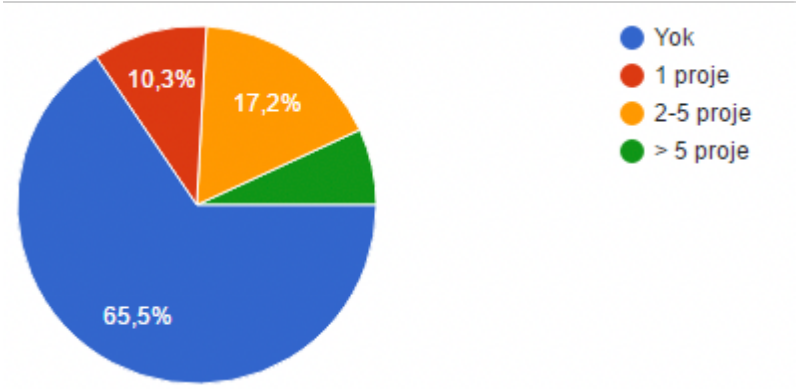
projeleri kapsamın alındığı sonucu ortaya çıkabilir. Bu noktada büyük veri tanımının net bir şekilde yapılması ihtiyacı oluşabilir.

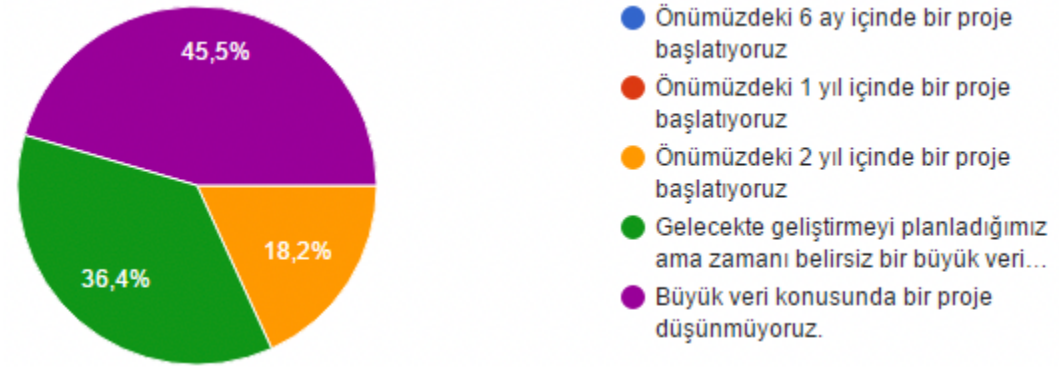
Soru 11. Büyük veri projeniz yoksa kurumun büyük veri konusunda gelecek planları nelerdir?

Büyük veri projesi bulunmayan %69 oranındaki kamu kurumlarının kısa ve uzun vadede büyük veri planları sorulduğunda alınan cevaplara göre kurumların;

- ● %18,2'si önümüzdeki 2 yıl içinde en az bir büyük veri projesine başlayacağını,
- ● %36,4'ü gelecekte geliştirmeyi planladığı ama zamanı belirsiz olan en az bir büyük veri projesinin olduğunu belirtirken
- ● Kalan % 45,4'lük dilimin hiçbiri büyük veri konusunda proje geliştirmeyi düşünmediklerini belirtmiştir.

Genel oranı aldığımızda kurumların %29,7'si büyük veri konusunda proje planlaması yapmaz iken önümüzdeki 2 yıl içinde büyük veri projesine başlamayı planlayan kurumların oranı %11,9, süre belirtmeksizin gelecekte bir büyük veri projesine başlamayı planlayan kurumların oranı ise 23,8'dir.



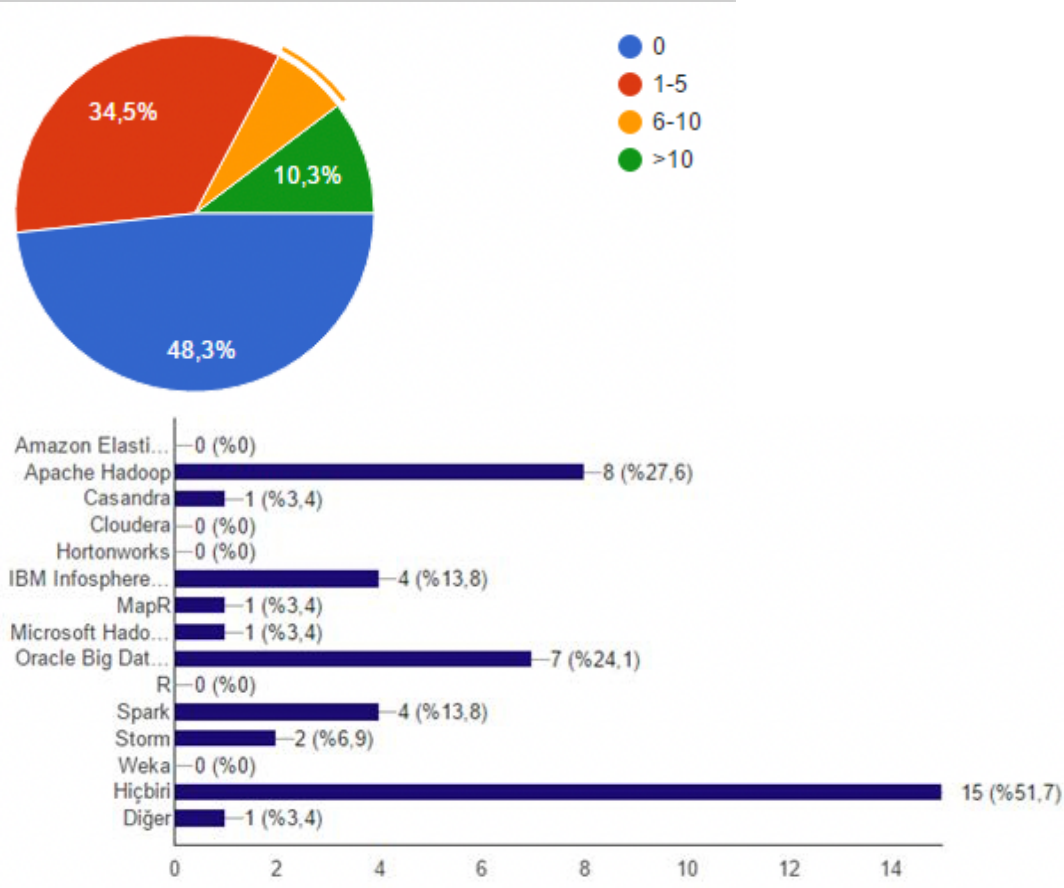


Soru 12. Büyük veri projelerinde çalışan sayınız kaçtır?

Büyük veri konusuna yapılan insan kaynağı yatırımının ölçüldüğü bu soruya verilen cevaplara göre kurumların büyük veri projeleri için gerçekleştirdiği istihdam sayıları:

- %10,3'ü 10'un üzerinde çalışan ile,
- %6,9'u 6-10 arası çalışan ile,
- %34,5'i 1-5 arası çalışan ile büyük veri projelerini geliştirirken
- %48,3'ü büyük veri konusunda henüz personel istihdamı yapmamıştır.

Soru 13. Büyük veri işlemede hangi teknolojileri kullanıyorsunuz ya da inceliyorsunuz?



Soru 14. Kurumunuzda "Büyük Veri Analitiği" konusunda aşağıdaki konularda yetkin çalışan sayısını belirtiniz.

Ankete katılan kurumlarında yarısında veya biraz daha fazlasında büyük veri analitiği konusunda yetkin çalışan olmadığı görülmektedir (mavi çubuklar - 0 çalışan). Ancak önemli bir kısmında (%40-50) en az 1 analitik konusunda çalışan bulunabilmektedir. Belirgin teknolojilerde, örneğin Hadoop'ta, 31 kurum içinde 9'unda bilen en az bir kişi bulunmaktadır (%33). Bu da büyük veri teknolojilerinden tamamen uzak olmadığımızı göstermektedir.

C: TURKISH SUMMARY/ TÜRKÇE ÖZET

TÜRKİYE'DE KAMU YÖNETİMİNDE BÜYÜK VERİ POLİTİKALARININ VE OLGUNLUĞUNUN DEĞERLENDİRİLMESİ

Geçmişte, filmlerde, bilim-kurgu romanlarında ve diğer medyada temsil edilen dijital çağ gündelik hayatın bir gerçeklik haline gelmektedir. Yalnızca teknolojinin basmakalıp gösterisinin yanı sıra, Bilişim ve İletişim Teknolojileri (BİT), yapay zeka ve nesnelerin interneti ile beraber de akıllı cihazlar gündelik hayatın bir parçası haline gelmiştir. Geçmişte, Kubrick'in Uzay Yolculuğu: 2001filminde tasvir ettiği HAL 9000 isimli bilgisayar günümüzde son kullanıcıya çeşitli isimlerde hitap ederek satılmaktadır. Dünya'nın birçok ülkesinde, zaman, maliyet analizi ve enerji tasarrufu gibi teknolojiler ile gerçek zamanlı bilgiyi kullanan akıllı şehirler yükseliştedir.

Bu teknolojik gelişmelerin tamamında uygarlığın başlangıcından beri en belirgin niteliklerinden biri olan veri üretimi ve kullanımı önemli bir rol oynamaktadır. (Maciejewski, 2016; Jansen & Kuk, 2016). Mağara duvarlarına bir taşın vurarak bir şeyleri kaydetmeye başlayışından beri veri üretiminde logaritmik bir artış gerçekleşti. Medeniyet tarihinde üretilmiş tüm kitapların neredeyse 15bin katı günlük olarak çeşitli veri kanallarıyla üretilmektedir. Bunların içinde salt metin bilgisi değil, aynı zamanla akış halinde sosyal medya, görsel ve işitsel yükleme verileri bulunmaktadır. Tüm bu, yüksek akış halindeki büyük hacimli ve çeşitlilik gösteren veriye, "büyük veri" denmektedir.

Yiu (2012) büyük veriyi temelde geleneksel veri tabanı yönetim uygulamaları kullanılarak işlenmesi güç olan veri türü olarak belirtmiştir. Bu, bir ülkenin anlık sağlık kayıtları (O'Reilly, 2015), eğitim politikalarında karar verme süreçleri (Lavertu, 2015), yıllık tarımsal sonuçlar gibi örnekleri içerir. Geçmişteki geleneksel veri tabanı

yönetim uygulamalarının yanına 2010'lara doğru çok katmanlı Hadoop, NoSQL gibi veri işleme uygulamaları eklenerek büyük veri analizi mümkün kılınmıştır.

Kamu sektöründe BİT hizmetleri ve e-devlet uygulamalarının gelişmesiyle, hükümetler büyük miktarlarda dijital veri toplamaya ve bunları işlemeye yönlenmektedir. Büyük verinin kamuda kullanımının verimliliğin arttırılması (Tomar, 2016), hizmet kalitesinin arttırılması (Bertot, 2013), erken suç tespiti (Ho & McCall, 2016), bürokratik süreçlerin azaltılması (Giest, 2017) gibi olumlu yanları bulunmaktadır. Büyük veri uygulamaları bürokratik işlemlerin ve kamu politikalarının iyileştirilmesiyle kalmamakta aynı zamanda hükümetlerin şeffaflıklarının arttırılması ve vatandaşların refahının iyileştirilmesi gibi diğer avantajlar sunmaktadır (Kimet vd., 2014).

Türkiye dahil olmak üzere Dünya'daki birçok ülke, büyük veri uygulamalarının benimsenmesi için politikalar geliştirmekte ve uygulamaktadır. Büyük verinin kullanımının avantajları ve dezvantajları birçok bilimsel makalede açıklanmıştır (Gietz, 2017; El-Darwiche, 2014; Hammer, Kostroch ve Quiros, 2017) Büyük veri uygulamalarının kamu yönetiminde uygulanması görece yeni olduğundan daha fazla teorik ve politik tartışmaya ihtiyaç bulunmaktadır. Bu uygulamaların kamu yönetiminde kullanımı yeteri derecede araştırılmamıştır ve gelişmekte olan bir akademik alandır. Türkiye'de kamu yönetiminde büyük veri kullanımının gelişmekte olmasından yola çıkarak, dijital uçurumun genişlememesi adına bu politikaların gelecekte nasıl geliştirilebileceğine ve edinebileceğine ilişkin bir analize ihtiyaç bulunmaktadır.

Bu çerçevede, bu tez çalışması Türkiye'de kamu yönetiminde büyük veri politikalarını değerlendirmek ve Dünya'daki yerini anlayabilmek adına büyük veri olgunluğunun ölçülmesine odaklanmaktadır. Bunları gerçekleştirmek için, kamuya sunulmuş çeşitli strateji ve eylem planları, uygulama örnekleri, ikincil anket verilerinden toplanan veriler incelenmiş ve büyük veri politikalarının ön analizini yapmayı amaçlamıştır.

Bu sebeple, tezin ikinci bölümünde genel bir literatür analizi yapılmış, konuya ilişkin teknik açıklamalar getirilmiş, akademik yönden çıktılar açıklanmış, büyük veri kullanımının fırsatları ve zorlukları incelenmiş ve büyük veri olgunluk modellerinin bir tartışması yapılmıştır. Tezin üçüncü bölümünde Dünya'da e-devlet yönünden gelişmiş ülkelere yer verilmiştir, bu yapılırken Birleşmiş Milletler'in iki yıllık e-devlet analiz raporlarından yola çıkılmıştır. Bunun yanı sıra 2050'ye dek gelişimleri öngörülen Brezilya, Çin, Hindistan ve Rusya (BRIC) ülkelerinden örnekler verilerek Türkiye benzeri gelişmekte olan ülkelerde büyük veri uygulamaları gösterilmiştir. Dördüncü bölümde, Türkiye'nin politika dokümanları incelenmiş ve veri merkezli bir bakışıcısından e-devlet'in gelişimi açıklanmıştır, bunun yanı sıra Türkiye'deki uygulamalara değinilerek genel bir bilgi verilmiş, neticede de Kuraeva'nın Kamu Yönetimi'nde Büyük Veri Olgunluk Değerlendirmesi (2016) yaklaşımı kullanılarak Türkiye'nin genel bir analizi yapılmıştır. Sonuç bölümünde bu çıkarımlar değerlendirilmiş, gelecekteki uygulamalar için öneriler getirilmiştir.

Büyük veri üzerine literatür taramasında öncelikle büyük verinin tanımı yapılmış ve özellikleri belirtilmiştir. Fisher, Drucker ve Konig (2012), büyük veriyi geleneksel veri yönetimi araçlarıyla depolamak, yakalamak ve analiz etmek için karmaşık yapıdaki veriler olduğunu söylemektedir. Örneğin, nitelikleri etiketlenmiş bir tablodaki yapılandırılmış veri ele alındığında bir vatandaşın, hangi ilde nüfusa kayıtlı olduğu, ikametgah adresi çerçevelerinin altından doğrudan veri çekilebilmektedir. Öte yandan yapılandırılmamış veri denilen teknik kavram, bu vatandaşın günlük sosyal medya üzerinden canlı olarak yapmış olduğu paylaşımlarının söylem analizi, sağlık hizmeti aldığında kendisinin yaş grubundaki kimselerin hastaneye hangi sıklıkta ve hangi sebeplerle gittiğinin analizi ya da bir afet durumunda kitle kaynaklı raporlama ile afet bölgesindeki kimselerin hizmetin götürülebilmesi gibi farklı türde verileri içermektedir.

Büyük verinin genel olarak kabul edilen üç özelliği bulunmaktadır. Bunlar hacim, çeşitlilik ve hızdır. Hacim, büyük veri dendiğinde akla ilk gelen özelliktir. Üretilen, toplanan ve işlenen verinin miktarını açıklamaktadır. (Buchholz vd., 2014) Verinin

hacmi giderek arttığı için göreliliği her zaman göz önünde bulundurulmalıdır. 20 yıl önceye dek yaygın bir şekilde kullanılan disket formatının milyonlarca kat üzerinde depolama alanına sahip hafıza kartları artık onlardan daha ucuza satılmaktadır. Verinin artışının yanı sıra, depolama ve işleme teknikleri de artmaktadır. Çeşitlilik ise, bahsedildiği gibi salt tablo halindeki veriyi değil aynı zamanda görsel ve işitsel veri gibi yapılandırılmamış veriyi içermektedir. Günümüzde ağ bağlantı hızlarının artışıyla beraber bu verilen doğru zamanda işlenebilmesi gerekliliği ortaya çıkmıştır. Veri ambarlarında depolanan verilerin yanı sıra bulut bilişimle anında ulaşılabilen "sıcak veri" kavramı ortaya çıkmıştır. Bu üç özelliğe İngilizce'de 3V şeklinde baş harfleri adlandırılmalarından yola çıkılarak İngilizce'de V harfiyle başlayan başka özellikler eklenmiştir, Örneğin değer (value), görselleştirme (visualization), geçerlilik (veracity) değişkenlik gibi diğer etkenler çeşitli araştırmacılar tarafından eklenmiştir.

Büyük veri ile ilişkilendirilebilecek üç önemli kavram bulunmaktadır. Bunlar i) bulut bilişim, ii) açık veri ve iii) nesnelerin internetidir. Bulut bilişm verinin akış hızına uygun olan, her yerde erişilebilirlik sağlayan altyapıyı sağlamaktadır. Açık veri, kamu tarafından toplanmış verilerin toplanan kişilere açılarak üzerinden uygulamaların gerçekleştirilebilmesini sağlamaktadır. Nesnelerin interneti ise eskiden internete bağlı olmayan cihazların, internet erişimi ve alıcılar aracılığıyla yeni fırsatlar sunmasına yardımcı olmaktadır.

Büyük verinin avantajları tezde şu şekilde sıralanmıştır, ekonomik avantajlar, veri güdümlü karar verme, topluluk kaynaklarının ve kamu yeteneği kullanımı, hizmet sunumu şeklindedir. Veri analitiği devletlerin vergi kaçakçılığının önlemesi, bütçe tahminlerinin iyileştirilmesi gibi örneklerde kullanılmaktadır. Veri güdümlü karar verme ise, içgüdüsel karar vermenin aksine kapsamlı analizler yaparak kamu görevlilerinin daha doğru kararları daha kısa sürede alabilmesine olanak tanımaktadır. Topluluk kaynaklarının ve kamu yeteneğinin kullanımı örneğin vatandaşların sahip oldukları mobil cihazlarla, afet raporlaması, ormansızlaşmanın önüne geçilmesi gibi uygulamalarda bulunmasına yardımcı olmaktadır. Hizmet sunumunun iyileştirilmesi ise, büyük veri analizi sayesinde örneğin sağlık hizmetlerinin daha iyi şekilde

sunulabilmesi, çeşitli analizin yapılarak kamu sağlığının iyileştirilmesi gibi alanlarda kullanılmaktadır.

Büyük veri uygulamalarının zorluğu açısından ise, öncelikli olarak teknik bir yapı olması sebebiyle, gerekli altyapının eksikliğinin yanı sıra Lourenco vd. (2017) gizlilik, örgüt kültürü, sivil işbirliği ve eski sisteme köklü adaptasyon gibi sebeplerle büyük veri uygulamalarının adaptasyonunun kamu kurumları açısından bir güçlük doğurabileceğinin altını çizmektedir. Bir diğer önemli zorluk ise mahremiyet/gizlilik konusundadır. Büyük verinin her yere her zaman ulaşan yapısı itibariyle mahremiyet konusu en öncelikli örneklerden biri olarak ortaya çıkmaktadır. Bu konudaki iyi uygulamalara, yasal altyapıya kendilerini uyarlayabilen ülkeler büyük veri edinimi konusunda diğer ülkelerden öncelikli hareket etmektedirler. Veri madenciliği ve analiz zorlukları, teknik zorlukların yanı sıra aynı zamanda hangi verinin nasıl işleneceği ve bağlantıların kurulabilmesine ilişkin sistemlerin ve sezgisel analizlerin üretilebilmesinin güçlüğünden söz etmektedir.

Kamu yönetiminin büyük veri uygulamalarını kullanımı açısından en önemli güçlüklerden birisi ise, istihdam konusundadır. Çünkü salt kamu kuruluşları için değil aynı zamanda özel kuruluşlar için de nitelikli veri analisti bulmak, çalışan piyasasının darlığı ve sektörel rekabet sebebiyle güçtür. Öte yandan kamunun maddi anlamda veri analistini tatmin edici maaşları verememesi oldukça yüksek maaşlarla çalışan bilişim sektöründe, özel sektöre kayılmasının önünü açmaktadır. Cukier ve Mayer-Schönberger'in (2013) tanımlamış olduğu verilerin diktatörlüğü, büyük veri analizinin insani karar verme mekanizmalarının önüne geçerek veri öncülüğünde karar verme ile insanı insan yapan özelliklerinden birini etkileyebileceğine ilişkindir, bu yazarlar aynı zamanda veri analizi şirketlerinin tekelleşmesinin de mahremiyetin önüne geçerek bilim kurgu filmlerindeki tekel bilişim şirketleri gibi rol oynama tehlikesinden bahsetmişlerdir. Tezin ana meselelerinden birisi olan, büyük veri hazırlığının eksikliği bir diğer güçlüktür, gerek kurumsal, altyapısal ve birçok içsel ve dışsal etkenler gerekçesiyle kurumlar büyük veri uygulamalarında olgunluğa erişemeyebilmektedirler.

Heeks'e (2016) göre kamu kuruluşlarının bilişim projelerini uygulamalarının yarısının tamamlanamamaktadır. Bunların iyileştirilmesi için büyük veri olgunluk modelleri önerilmektedir çünkü olgunluk modelleri bir kurumun bir politikada nerede durduğunu tanımlayabilmekte, yol gösterici çözümler sunmakta ve benzer kurumlarla karşılaştırarak hangi yönden eksik olduğunu göstererek kuruluş için bir tanı modeli rolünü oynayabilmektedir. Bu modeller 1973 yılında Richard Nolan'ın öne sürdüğünden beri uygulanmaktadır. BİT uyumuna ilişkin, 120'den fazla olgunluk modeli bulunmaktadır (Mettler vd., 2010) doğrudan e-devlet benimsenmesi için 25 model bulunmaktadır (Heeks, 2015) İş dünyasına ilişikin birçok olgunluk modeli bulunsa dahi, sadece kamu yönetiminde büyük veri olgunluğunu değerlendirmeye yönelmiş iki modele ulaşılabilmiştir. Bunlar Kuraeva (2016) ve Klievink vd. (2016) modelleridir. Bu modellerden Kuraeva'nın modeli daha genel bir çerçeve sunarak bir ülkenin genel olarak büyük veri uygulamalarında nerede olduğunun çerçevesini sunmaktadır. Bunu yaparken ülkenin büyük veri edinimine ilişkin olarak 6 kategori saptayarak bunların olgunluğunu değerlendirmektedir. Bu kategoriler i) vizyon ve strateji, ii) açık veri girişimleri iii) araştırma geliştirme enstitüleri ve girişimleri iv) özel sektörde büyük veri olgunluk düzeyleri v) veri yönetimi becerileri vi) kamuda büyük veri projeleri deneyimleri, şeklindedir. Bu uygulamaların olgunluk düzeylerini i) farkında ii) keşfediyor iii) uyum sağlıyor iv) dönüşüyor şeklinde dört olgunluk seviyesi atayarak çeşitli kriterlerle her bir kategoride nerede olduğuna puan vererek genel olarak hangi kategoride olunduğunun bir değerlendirilmesi yapılmaktadır.

Klievink vd.'nin modeli, kurum bazında organizasyonel uyum, kurumsal yetenek ve organizasyonel olgunluk alt değerlendirmelerine bölerek karmaşık bir analiz yapmaktadır. Tezde amaçlanan Türkiye'nin büyük veri uygulamaları açısından genel konumunu değerlendirmek olduğundan, Kuraeva'nın modeli kullanılmıştır.

Türkiye'deki uygulamaların değerlendirilmesinden önce, Dünya örnekleri değerlendirilmiş ve bu ülkelerin süreç açısından nerelerden geçmiş olduklarına değinilmiş, aynı zamanda ülkelerin gerçekleştirmiş olduğu genel, finansal, akıllı kent uygulamaları, güvenlik politikalarının yanı sıra, gerçekleştirilmiş bazı yaratıcı

örnekler de değerlendirilerek büyük veri analizinin hangi alanlara dek genişleyebileceği gösterilmiştir. Örneğin, Amerika Birleşik Devletleri Kadınlar Olimpik Bisiklet Takımı, büyük veri analizi ile gerçekleştirilen kişiselleştirilmiş antrenman, beslenme ve uyku programlarının yardımıyla kendi tarihinde gerçekleşmemiş bir başarıya imza atarak 2012 yılında kadınlar takip kategorisinde ilk kez gümüş madalyayı almıştır. Genel olarak büyük veri analitiğinin kullanıldığı öncelikli alanlardan birinin sağlık sektöründeki uygulamalar olduğu göz önünde bulundurulduğu, kişiselleştirilmiş ilaçlardan, tedaviden söz edildiği göz önünde bulundurulduğunda bu örnek öncül bir nitelik taşımaktadır.

ABD Dünya'da büyük veri politikaları açısından en erken harekete geçen ülkedir. 2010 yılında hazırlanan dijital gelecek raporunda belirtilen temel politikalardan biri her federal kurumun kendine ait bir büyük veri stratejisine sahip olması gerekliliğidir. (Kearns, 2010, p. xviii) Bunun yanı sıra, şimdiye dek bahsedilen mahremiyet konusuna da değinilerek yurttaşların veri politikaları açısından mahremiyetlerinin güvence altına alınması gerekliliğine değinilmiştir. Barack Obama'nın büyük verinin, ekonomik ve sosyal boyutunun devlet üzerindeki farklı etkilerine ilişkin vurgu neticesinde 2014 yılında büyük veri kullanımına ilişkin detaylı bir Başkanlık Raporu hazırlanmıştır. Buradan yola çıkarak yıllar içerisinde başta güvenlik olmak üzere birçok konuda uygulamalar geliştirilmiştir. Hem federal devlet düzeyinde hem de eyalet ve şehirler bazında birçok uygulamaya yer verilmiştir. Örneğin Los Angeles Emniyet Müdürlüğü, Los Angeles California Üniversitesi işbirliği ile geçmişteki verileri ve gerçek zamanlı akış halindeki verileri kullanarak gelecekte işlenebilecek suçlara ilişkin analiz gerçekleştirirek emniyet güçlerinin devriye sıklıklarını ve rotalarını yeniden düzenlemektedir. Desouza (2012) mekânsal ve coğrafi-mekansal analiz gerçekleştirilerek yapılan bu tip analizlerin büyük verinin kamu sektöründe uygulanışına önemli bir örnek olduğunu söylemektedir. ABD'nin diğer uygulamaları sağlıkta bildirim sahtekarlığının tespiti (Mills vd., 2012), sınır güvenliğinin iyileştirilmesi (Marr, 2016), açık veri portalı, yerel yönetimlerin hizmet olanaklarının iyileştirilmesi gibi yerlerde kullanılmaktadır.

Birleşik Krallık, veri analizini kullanmaya başlayan birçok ülke gibi öncelikli olarak bütçe, vergi toplama ve finansal piyasaları analiz etmek gibi uygulamaları devreye sokmuş. İngiliz Gelir ve Gümrük İdaresi tarafından gerçekleştirilen "British Connect" isimli uygulama 45 milyon pounda mal olmuş sistem bir yıl içerisinde 1,4 milyar poundluk bir vergi geliri kaçışının önüne büyük veri analizi sayesinde geçmiştir. Sisteme yapılan 150 milyon poundluk ek yatırım bir yıl içerisinde 35 milyar poundluk verginin tahsil edilmesine olanak tanımıştır. (Caldwell, 2014)

Ekonomik önlemlerin yanı sıra, sağlık uygulamalarının iyileştirilmesi de paralel devam eden bir süreç olmuştur. Çeşitli bölgelerde yüksek fiyatlı lisanslı ilaçların reçete edilmesinin Ulusal Sağlık Hizmetleri Kurumu'na (NHS) yıllık 200 milyon pound maliyeti olduğu görülmüş yerine, aynı yapıdaki muadil ilaçların reçete edilmesi teşvik edilerek aylık 30 milyar poundluk bir tasarruf edilmesinin önü açılmıştır.

Fransa büyük veri uygulamalarının kullanımlarında dikkatli olunması gereken bir nokta olan, fişleme açısından ele alınmıştır. Olağanüstü hal döneminde kullanmış oldukları sistemler algoritmaların yönlendirmesi aracılığıyla çeşitli etnik arka plandan gelenlerin yoğun olduğu mahallelerdeki kimselerin önlemler çerçevesinde fişlenebildiğini göstermiştir. (Kubler, 2017) Avustralya gelişmiş ülkeler arasında düzenli politika belgelerini en iyi hazırlayan ve bunların madde madde uygulamasını gerçekleştiren bir örnek olarak gösterilmiştir. Japonya Dünya'da teknoloji konusunda önde gelen bir ülke olarak altyapı yatırımlarına önem vermiş oldukça gelişmiş veri aktarımı sistemleri kurmaktadır. Singapur için çok daha detaylı uygulamaların varlığından söz edilebilir. Uluslararası kuruluşlarla işbirliği yaparak ülkelerini büyük veri açısından lider bir ülke konumuna getirmeyi amaçlamaktadırlar. Yerel yönetimler açısından da akıllı kent uygulamalarında başı çeken ülkelerdendir.

Türkiye'de kamu yönetiminde büyük verinin değerinin anlaşılması süreci açısından öncelikli olarak strateji ve eylem planları, kalkınma planları, bütçe raporları incelenerek veriye ilişkin bir süreç analizi gerçekleştirilerek veri politikalarının gelişimi gözlenmiştir. Bu incelemede, kamu yönetiminde verinin değerinin giderek

arttığı, kamu üst belgelerinde bahsedilme sıklığının da arttığı görülmüştür. 1998 yılında ilk kez sözü edilen2015-2018 Bilgi Toplumu Stratejisi ve Eylem Planı ile 2016-2019 e-Devlet Stratejisi ve Eylem Planları güncel dokümanları, doğrudan büyük veri uygulamalarının farkındalığının arttığı, çeşitli kamu kurumlarında pilot uygulamalarının gerçekleştirilmeye başlandığı geleceğe ilişkin büyük veri politikalarının planlandığı görülmüştür.

Bunun yanı sıra Avrupa Birliği'ne uyum sürecinde yasal düzenlemelerin de paralel olarak edinildiği görülmüştür. Bu yasal düzenlemeler, 2016 yılında oluşturulan Avrupa Birliği Genel Veri Koruma Düzenlemesi (EU GDPR) çerçevesinde kurumsal dönüşüme de olanak sağlamış, Kişisel Verileri Koruma Kurumu kurulmuştur. başlığında uygulamalardan, Sağlık Bakanlığı, Sosyal Güvenlik Kurumu (SGK) ve Milli Eğitim Bakanlığı'ndan (MEB) çeşitli uygulamaların örnekleri açıklanmıştır.

Türkiye'de Sağlık Bakanlığı, Sağlık Bilişim Ağı, Sağlık Yönetim Sistemi, e-Nabız, Mekansal İş Zekasıi sistemleri sayesinde "büyük verinin yönetilmesi ve analizi ile etkili sağlık hizmeti sunmak ve gerçek zamanlı karar desteği sağlama" (Ülgü & Gökçay, 2017, p. 281) mümkün olmuştur. Bu uygulamalardan, Sağlık Bilişim Ağı sayesinde, alternatif bir altyapı kurarak, ihtiyaç halinde diğer kamu kurumlarının da kullanabileceği bir ağ oluşturmuştur. Bu tip entegre edilmiş modeller, gelişmiş büyük veri uygulamalarına ait veri tabanı yapılarını içermektedir. SGK'nın uygulamalarından MEDULA hastane ve eczanelerin entegre bir provizyon mekanizması içerisinde çalışmasına yardımcı olmaktadır. Bunun yanı sıra Türkiye'de büyük verinin akıllı kent uygulamalarında, ulaşımın iyileştirilmesi, akıllı kart hizmetleri gibi yerel yönetimlerde uygulamaları bulunmaktadır.

Türkiye analiz edilirken, Kuraeva'nın modeli ele alındığında, vizyon ve strateji açısından Türkiye büyük veriye ilişkin çerçeveyi üst politika belgelerinde kurmuştur ancak spesifik bir büyük veri eylem planı hala bulunmamaktadır. Açık veri uygulamaları açısından Türkiye İstatistik Kurumu çeşitli verileri kendi sitesinde sunmaktadır ancak, bunlar yoğunlukla ekonomik verileri içermektedir. Aynı zamanda

Türkiye, güncel bir açık veri eylem planı hazırlamadığı için Açık Veri Partnerliği (OGP) üyeliği sonlandırılmıştır. Araştırma ve geliştirme girişimleri açısından, TÜBİTAK'ın kurmuş olduğu Büyük Veri ve Bulut Bilişim Laboratuvarı ve Gazi Üniversitesi'nin Büyük Veri ve Bilgi Güvenliği Araştırma Merkezi kurmuştur. Çeşitli Üniversiteler büyük veri ve veri analitiği programları yüksek lisans düzeyinde bulunmaktadır. Özel sektörde büyük veri olgunluğu ele alındığında, birçok ülkede tekrar etmekte olan özel sektörün kâr amacı gütmesi sebebiyle büyük veri analizine daha erken entegre olduğu ve Türkiye'deki birçok büyük şirketin bundan çeşitli uygulamalarla fayda sağladığı görülebilmektedir. Veri yönetimi becerileri, kurumların entegre edilmiş veri tabanı uygulamaları kullanımları ve veri eşgüdümlülüğü gibi konuları içermektedir. Kamu kurumlarında, güncel Ulusal İstatistik Programı bunların gerçekleştirilmesini öngörmektedir ancak günümüzde bu konuda yeterli uygulamalar bulunmamaktadır. Kamu kurumlarında büyük veri projeleri açısından, Sağlık Bakanlığı, MEB, SGK gibi kurumların çeşitli projeler gerçekleştirdiği örnekleri verilmiştir, ancak bu uygulamalar tüm kamu kurumlarına yayılmadığından, bu kriterde yeterlilik yüksek seviyede değildir. Genel çerçeve ele alındığında Türkiye'nin özel sektörünün büyük veri konusunda görece gelişmişliği ya da, vizyon ve planlama açısından politika belgelerinde varlığı gibi etkenler göze alındığında Türkiye'de kamu kurumlarında büyük veri farkındalığının varlığının olduğu ve genel olarak pilot uygulamalar çerçevesinde de keşfeden bakış açışını geçmiş olduğu söylenebilir. Aynı zamanda araştırmada bu farkındalık ikincil verilerle desteklenerek de analiz edilmiştir. Henüz kendi kendini optimize eden bir sürece girmemiş olsa da, analiz Türkiye'de kamu yönetiminin büyük veri uygulamalarına uyum sağlamaya başladığı kriterlerini yerine getirdiği görülmektedir.

Araştırmanın geneli ele alındığında, büyük verinin değeri ve uygulama alanları tartışılmıştır. Kamu yönetimi alanındaki akademik çıktılardan yola çıkarak bir literatür analizi gerçekleştirilmiş böylece kamu yönetimi açısından genel bir çerçeve çizmek mümkün olmuştur. Bunun yanı sıra e-devlet uygulamaları açısından gelişmiş ve gelişmekte olan ülkelerin örnekleri verilmiştir. Gelişmiş ülkelerde, yasal düzenlemeler erken yapılmakta, bilişim altyapısının öneminin farkındalığı yüksek olmakta bu

sebeple de çeşitli BİT özel sektör danışmanlık kurumlarıyla işbirliği yapılarak çıktılar alınmaktadır. Bu örnekler genel olarak ele alındığında görülmüştür ki, büyük verinin kamuda uygulamalarında öncelikli olarak Yeni Kamu İşletmeciliği çerçevesinde maddi güdülenme yüksektir bu sebeple vergi kaçırmanın önlenmesi, sağlık ve eğitim hizmetlerinin giderlerinin düzenlenmesi gibi etkenlerle kullanılması öncelikli örnekler arasındadır. Bunun yanı sıra yerel yönetimler açısından akıllı kentler aracılığıyla büyük veri analizi mümkün kılınmaktadır. Gelişmiş e-devlet ülkeleri açık veri uygulamalarına önem vermekte, bu şekilde olumlu çıktılar alabilmektedir.

BRIC ülkeleri göz önünde bulundurulduğunda, Brezilya haricindeki üç ülke Açık Veri Partnerliğinde ya kendileri çekilmiş ya da hiçbir zaman üye olmamışlardır. Bu ülkeler arasında Rusya, bilişim sektörüne yeterince yatırım yapmadığından uygun altyapısı bulunmamakta bu sebeple de Hindistan, Çin ve Brezilya'nın gerisinde kalmaktadır. Türkiye ile karşılaştırıldığında bu ülkeler arasında Hindistan ve Brezilya'nın çeşitli uygulamalarının benzerliği görülebilmektedir. Öte yandan bu ülkeler yasal düzenlemeler açısından Türkiye'den geri kalmaktadırlar. Politika belgelerinin takibi, gelişmiş ülkelerle karşılaştırıldığında yetersiz kalmaktadır, büyük veriye ilişkin sistematik bir politika belgesi oluşturulmamaktadır. Çin, bilhassa rejimin güvenliğini sağlamak maksadıyla bu alanda oldukça gelişmiş büyük veri uygulamaları gerçekleştirmektedir, ancak bunun dışında vatandaşlar için çıktılar yeterli bulunamamıştır.

Türkiye'de kamu yönetiminde büyük veri uygulamalarının iyileştirilmesi için öncelikli olarak bir Büyük Veri Stratejisi ve Eylem Planı hazırlanmalıdır. Bunun yanı sıra veri yönetiminin iyileştirilmesi için kamu kurumlarının bütünleşmiş edildiği ulusal bir veri merkezinin kurulması gereklidir. Büyük veri analizi için eğitim programları bulunsa da, bunlar yüksek lisans seviyesindedir bilişim eğitimini erken yaşlarda başlatılarak bu konudaki çıktılar iyileştirilebilir. Bunun yanı sıra açık veri uygulamalarına tekrardan başlanması ve bir açık veri portalı kurulması gerekmektedir.

Bu noktada araştırmanın sınırlılıklarından birinin farklı dillerdeki politika belgelerinin çevirilerine ulaşamamak olduğu söylenebilir. Brezilya, Rusya ve Çin gibi ülkelerin belgeleri ikincil kaynaklarda çeşitli maddeleri İngilizce 'ye çevrilmiş olsa da araştırma esnasında bu kaynakların metinlerini okuyabilmek mümkün olmamıştır. Tüm ülkeleri içermesi mümkün olmayan bu tezde literatürde sık anılan gelişmiş ülke örnekleri ve Türkiye ile karşılaştırılabilirliği olan Brezilya, Rusya, Hindistan ve Çin örneklerine sınırlandırılmıştır. Bunun yanı sıra kamu kurumlarında çalışanlarla genel bir alan araştırmasının yapılması ve Türkiye'deki kamu kurumlarında da uygulanabilecek bir büyük veri olgunluk modelinin oluşturulması ya da adapte edilmesi genel çerçevenin daha iyi anlaşılabilmesi ve bir harita çıkarılması açısından faydalı olacaktır. Gelecekte yapılabilecek araştırmalar açısından kamu yönetiminde artan veri güdümlü karar vermenin, davranışsal bilimler ışığında işlerliğinin arttırılmasına ilişkin bir araştırma gerçekleştirilebilir.

D: TEZ İZİN FORMU

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ■

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : Okuyucu.....
Adı / Name : Aras.....
Bölümü / Department : Siyaset Bilimi ve Kamu Yönetimi.....

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English) : AN ASSESSMENT OF BIG DATA POLICIES AND BIG DATA MATURITY IN PUBLIC ADMINISTRATION IN TURKEY
.....
.....
.....

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ■ **Doktora** / PhD ☐

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ■
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two year. *** ☐
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. ***** ☐

***** *Enstitü Yönetim Kurulu Kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.*
A copy of the Decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature ..... **Tarih** / Date .....