The link between Evidence-based policy making and Big Data Analytics: Policy analytical capacity and Digital-Era Governance as enablers of big data assimilation

Konstantinos Zeimpekis

Faculty of Governance and Global Affairs
Leiden University, The Hague
June 2018

Supervisor: Dr. S. N. Giest
Assistant Professor, Faculty of Governance and Global Affairs, Institute of Public Administration, Leiden University, The Hague

First Reader: Dr. V.E. Pattyn
Second Reader: Dr. J. Schalk

A thesis submitted in partial fulfilment of the requirements for the award of MSc in Public Administration, in the International and European Governance Track

Student number: 1915134

To Polyxene and Thalia, for being supportive and without whom I would not have made this journey of knowledge

K.Z.

Abstract

The scope of this study is to examine the impact of data analytics inside a modern public administration, in order to find out how a public organization can reach their unmitigated assimilation. In the endeavor to unveil this concern, the study synthesizes the relationship between the concepts of policy analytical capacity, evidence-based policymaking and Digital-era Governance in order to locate the possible causal links leading to big data assimilation. Eventually, it is established that policy analytical capacity is an essential quality for utilizing data-driven methods inside a public organization. However, it cannot be conclusively confirmed that policy analytical capacity is a necessary condition for big data assimilation to occur. Alternatively, Digital-era Governance is found to be a catalytic enabler of big data assimilation, as it inserts the most appropriate public-sector settings for the advent of the big data analytics technology.

Keywords

Policy Analytical Capacity, Digital-Era Governance, Evidence-based policymaking, Big Data Analytics, Innovative technology Assimilation

Acknowledgements

I would like at first to declare my gratitude to my supervisor, Dr. Sarah Giest for her assistance during this thesis, and also for providing me with the necessary guidelines for undertaking this project. Her profound knowledge, scientific integrity and teaching method, provided me with valuable knowledge and assisted me in the outcome of my thesis. Without attending her course "Data-Driven Policymaking, during the second block of my studies and without her proper guidance, this project would have been considerably less inspiring.

It would be terrible omission of mine not to mention the people that were close to me during this time. I should thank my mother, Polyxene for tirelessly supporting me during my studies and for all of my choices up till now, and my father Evangelos for conveying to me the ambition for studying in a high-quality international environment. Without them, my studies would not have made possible. I would like to thank Thalia, likewise, for providing me with all the support during the demanding and difficult periods of this master's.

Finally, I would really like to thank all my teachers that contributed to cultivating my personality both inside and outside of the classroom. I am doing my best to prove worthy of their efforts and concerns.

Konstantinos Zeimpekis

The Hague, June 2018.

Contents

1. Introduction

Big data can be characterized as a sweeping torrent, as it gradually penetrates the structure of every organization, regardless of whether it is a private company or a public administration. Organizations globally begin to sense the promising potential of big data analytics, as trillions of bytes of data are being produced every day; data that contain valuable information, that if analyzed, they offer extraordinary insights that otherwise they would have remained hidden. Big data technology is deemed important as it combines fragments of heterogeneous information, in order to identify information flows that are yet undetected (Höchtl et al., 2016). In a Promethean spirit, big data analytics, make use of aggregated information in order to offer us new tools for confronting our intractable problems and improving our standards. This promise cannot be left untapped by public administrations, as the potential for producing public value from it, is extensive.

In that respect, the employment of big data applications by public organizations and the integration of such technologies into their day-to-day operations, are usually recognized as catalysts, that provide unprecedented benefits to the public affairs. In detail, big data analytics present the possibility for enhanced governmental efficiency, effectiveness and transparency (Klievink et al., 2017), as well as, promises for ameliorated evidence-based policy making and transformation of public service delivery (Giest, 2017). However, opposite voices claim otherwise, suggesting, not only that nothing has changed, and that big data is just an ephemeral technological trend, but that big data is accompanied by challenges, that can be detrimental to the public administration (Margetts & Sutcliffe, 2013). Nevertheless, in order to leverage the effects of big data, public organizations have to display certain characteristics, such as organizational capabilities, structural alignment, maturity (Klievink et al., 2017), or organizational speed and openness (Höchtl et al., 2016). Generally, the argument supports that big data analytics come with an array of requirements, that a public organization must satisfy in order to render the assimilation of big data applications real.

Influenced from the work of Purvis and colleagues (2001), big data assimilation can be defined, as the eventual outcome that a public organization experiences when, it has selected, implemented and completely incorporated a big data application, as a part of its standard operating procedures. As the use of big data technology in the public sector is a recent issue, and the academic literature that examines the assimilation of big data analytics by public

organizations is almost scarce, the objective of the study, is to identify enablers that can facilitate the assimilation of big data applications inside the public sector. Specifically, the study aims on uncovering possible causal factors drawn from relevant literature, which can function as necessary requirements that a public organization has to have in order to accomplish the assimilation of big data technologies.

Inspired by the research on e-government (Bertot et al., 2014), the study synthesizes the literature in order to distill practices and variables that can facilitate the assimilation big data technology. Accordingly, by looking on literature relevant to evidence-based policy-making (Sanderson, 2002; Head, 2008; Howlett & Newman, 2010; Giest, 2017), it is established that the former, is practically related to big data use in the public sector, as evidence-based policy analysis requires the presence of a large pool of information that functions as a resource for better evidence-based policy outcomes; Big data can offer this possibility, but individuals, inside the public administration, are still required to have the capacity to comprehend this process (Giest, 2017, p. 368). This is where the concept of policy analytical capacity, that is "*the knowledge of policy substance and analytical techniques at the individual level*" (Howlett, 2015), comes into view. The study introduces the concept of policy analytical capacity, as it is pointed by Howlett (2015), inside the research niche, concerning big data use in the public sector, and respectively attempts to examine the causal connection between policy analytical capacity and big data assimilation. In other words, the former is assumed to be an enabler of the latter. In addition, the study aims on uncovering possible alternative or omitted causal factors that might influence innovative technology assimilation. Similarly, digital literacy is introduced inside the research, as it is assumed to prescribe individual qualities, that can have the same effect on big data assimilation. In more detail, digital literacy is theorized as a prerequisite for having the ability to utilizing digital formats (Bawden, 2008). Additionally the relevant IT knowledge is assumed as a precondition for IT assimilation (Ranganathan et al., 2004); Accordingly, digital literacy is examined as a surrogate to policy analytical capacity.

Driven by a reasoning for controlling for equifinality, the study tries to find different causal sequences that can lead to the same the end state of big data assimilation. Respectively, the study expands the search for causal factors besides the level of individual skills, and searches for structural and systemic influences on big data assimilation. Respectively, the effect of the macro-theory for public sector development is introduced into the research, as the Digital-Era Governance paradigm is tested as a potential enabler of big data assimilation. The literature on DEG, as it was first introduced by Dunleavy, Margetts and colleagues (2006), perceives this

governance paradigm, as an administrational approach, that has the ambition to reestablish the structure and processes of the public sector, in order to achieve reintegration, needs-based holism and digitalization within the public services (Margetts & Dunleavy, 2013). Thereby, the DEG approach, relying heavily on the utilization of digital technologies, calls for the introduction of big data analytics within the public sector, as a part of its reintegration component (Margetts & Dunleavy, 2013). Subsequently, it is hypothesized that the DEG approach introduces the most appropriate settings for the advent of big data and therefore, stands a systemic causal factor, that can affect public organizations positively towards attaining big data assimilation. In sum, the aim of the study is therefore twofold, as it first seeks to conclude on whether certain conditions exist in the empirical world, for example the assimilation of big data applications and the employment of the DEG paradigm; and second it aims to explore causal mechanisms, placed at a different level of observation, through which big data assimilation can be achieved.

In order to do so, the case of the UK public sector is selected, as the country stands as an appropriate case for inspecting the relationship between policy analytical capacity and big data assimilation. In more detail, even though the country endorses digital transformation, is receptive to innovative technology and displays an outstanding eagerness for digital governance, it is a country that in the same time, experiences considerable gaps and staff shortages related to its digital capabilities. This argument imposes a theoretical anomaly, as a country that is underequipped in digital competencies within its civil service, it is expected to be unable to overperform in digital governance and IT adoption. Thereby, this paradox is a puzzle hat troubles the theory, taking again into account, that the assumption that IT knowledge is positively related with IT assimilation (Ranganathan et al., 2004). In the supposed absence of digital literacy, policy analytical capacity can be tested as a self-sufficient causal factor of big data assimilation, increasing the confidence that the expected outcome of interest, is not the effect of alternate civil service skills sets that are semantically relevant to policy analytical skills. In addition, the independent variables drawn from the individual level, are juxtaposed with the systemic factor that the DEG introduces. Being under a digital transformation, the UK stands as the perfect environment for searching for the presence of DEG features, that might influence the assimilation of innovative technologies by the government. Thereby, the study aims on discovering causal sequences that relate policy analytical capacity with big data assimilation inside the UK public sector, as well as, on questioning the fact pertaining that the UK displays low levels of digital literacy within the civil service. Eventually, the main research question that

emerges is "*How does policy analytical capacity affect big data assimilation in the public sector?"* Except for these variables, the study sets to provide an answer on whether the DEG paradigm is prominent within the UK public sector and sets to reveal any possible causal links between the DEG's features and the utilization of big data applications. Subsequently, the sub-question that is posed, is "How does the Digital-Era Governance paradigm can function catalytically for big data assimilation?".

As the research on big data use in the public sector is something recent to the public administration science, yet there is no established literature that explains how public organizations can assimilate big data technologies. The answer to the above questions is expected to provide theoretical foundations and verifiable empirical arguments that can be tested scientifically. By combining insights from the already established theoretical apparatus on e-government, evidence-based policymaking, as well as, from the more generic and expanded literature on public administration, the study aims on uncovering potential causal links and constructing causal mechanisms that can signify how big data assimilation can be achieved. In this way, the study contributes back to the literature, as it provides insights to this novel research branch that can be useful for future scientific research. Additionally, by searching for empirical evidence and testing the literature-extracted causal mechanisms, the study acquires a theory-testing nature, as it evaluates the premises and the predictions, emerging from the previously-established theories. Finally, it is expected that new academic research will emerge, as an answer or a contribution to this endeavor.

Except for the academic relevance, the answer to the study's question aims at providing advice that can later be used by public administrations operationally. In other words, the study's ambition is to inform public officials and governments, about the ways of how big data assimilation can be achieved inside an actual public administration setting. As the knowledge-based society, living in the digital-era, craves for more information, the big data technology is becoming more and more penetrative, making its way into different sectors of the global economy. Accordingly, public administrations are asked to keep up with the latest innovative technological demands, and respectively great resources are being put in that endeavor. By becoming aware, governments can relocate their resources and get themselves prepared for utilizing the potential that accompanies the adoption of innovative technology; Eventually, that is the societal contribution that this study wants to achieve.

Regarding the thesis' structure:

This thesis begins in *Chapter 2,* where a literature review of the older and more current theoretical foundations that have already laid the groundwork for research relevant to big data use is the public sector can be found. The literature starts by specifying the models of digital change and governance that laid the basis for the advent of big data applications in the public sector. I begin by describing the impact of e-government in digital change and then I proceed to identifying the theories that stand behind the norms of public sector development. For establishing the concepts around policy analytical capacity, the related concept of evidence-based policy making is used and their connection with the adoption of big data in the public sector is signified. The literature review continues by specifying the arguments related to innovation diffusion and adoption. After having clarified the conceptual difference between the innovation assimilation and adoption, the possible barriers to innovation, which subsequently inhibit the application of big data technologies, are explained as well. The chapter concludes by displaying the hypotheses of the study and deconstructing the causal mechanisms that stand behind the hypotheses.

Chapter 3 describes the methodology that was followed regarding the single-case design of the study. The chapter begins by stating the research goal of the study, as well as, the rationale behind the case selection. Next, the case, the units of observation and the units of analysis are defined accordingly. Later, the data collection phase is described, and the nonreactive research of existing sources is placed as the study's method for making observations. Later, I proceed to providing conceptual and operational definitions that will ensure the construct and measurement validity of the study. Concluding, the overall single-case research design of the study is explained, and then the theoretical guidelines for putting process-tracing method of analysis into empirical use, are sufficiently presented.

Chapter 4 presents the findings from the data collection procedures, along with examples of big data application usage in the UK public sector, as well as findings on each of the independent and dependent variables. Additionally, figures present the composition of the UK public sector and indicate evidence on policy analytical capacity within the civil service.

Chapter 5 presents the empirical analysis of the study. The chapter begins by assessing evidence on how policy analytical capacity affects big data assimilation in the UK public sector. The causal mechanism is elaborated and then the hypothesis is tested against the empirical evidence, as the process-tracing probability test method prescribes. Eventually, the confidence towards accepting that policy analytical capacity is an enabler of big data assimilation can be increased,

however, the evidence is deemed circumstantial and the hypothesis is not confirmed conclusively. Respectively, the same sequence of actions follows for the evaluation of digital literacy. However, the causal mechanism suggesting that digital literacy can lead to big data assimilation is rejected, as digital literacy is accepted as an unnecessary condition for big data assimilation to occur. Later, it is presented that the UK does indeed operate under the reign of the DEG paradigm, and the causal mechanism suggesting that the paradigm facilitates big data assimilation is considered valid. Finally, the academic implications, and limitations of the study, as well as, the suggestions for future research are expressed thoroughly.

Finally, in the *Conclusion* of the study, it is established that policy analytical capacity can be theorized as an enabler of big data assimilation. However, as the evidence is not critically decisive, this hypothesis is not conclusively validated. On the contrary, the DEG is accepted as being an important facilitator of big data assimilation in the UK, as it introduces the most convenient settings for the utilization of big data analytics within the public administration environment.

2. Literature review

The inventiveness to incorporate big data applications into the public sector, as well as, the public administration's resourcefulness towards becoming more data-driven future is not something that happened in a vacuum. For examining the impact around the advent of big data in the public affairs, one needs first to begin by interpreting the theoretical foundations that proposed the inclusion of Information and Communication Technologies and new competences into the government's everyday operations in the first place. In other words, the concepts around E-government and the trajectory towards Digital Era Governance are aspects that are going to be described in detail. Additionally, as the employment of big data in the public sector arrives with an increase in evidence availability, the ideological and theoretical apparatus of evidence-based policy making will be made explicit. Last, as big data is semantically relevant to innovation, the process of innovation diffusion is going to be explained and the potential barriers to innovation are going to be classified. The purpose of this structure is to distill valuable causal mechanisms from the theory, that have the potential to explain the possible causal sequences, that connect the concept of policy analytical capacity with that of big data assimilation, as well as, to interpret the effect of other possible confounding factors[1].

2.1 Theories, Policies and Practices behind the advent of big data in the public sector

Concerning the use of big data analytics by the public sector, there are numerous policies that govern the way of how, information and data are managed, used, repurposed and accessed by public sector entities (Bertot et al., 2014, p. 9). The policy scheme, that set up the public-sector environment for the arrival of new technologies and the exchange of digital information into the public administration, is the E-government strategy. E-government can be defined as the "*utilization of the internet and the word-wide-web for delivering government information and services to citizens*" (UN & ASPA, 2001, p.1). Specifically, E-government refers to the utilization of technologies by the government, with the purpose of improving the accessibility and the circulation of government information to citizens, as well as, to public and private entities.

[1] The first two sections of the literature review were inspired on November 2017 during the final assignment of the "Data-Driven Policymaking" course, that was offered by Dr. S. N. Giest.

(Layne & Lee, 2001, p. 123). The employment of Information Technologies in the public sector, that allowed the distribution of government information to multiple receptors, was established as an essential element for achieving improved internal managerial efficiency and upgraded quality of public service delivery to citizens (Moon, 2002, p. 424). Enthusiasts of the e-government scheme include to the list of benefits, possibilities for increased citizen participation, democratic value diffusion, as well as, "*enhanced government accountability and transparency"* (Ahn & Bretschneider, 2011, p.414). Thereby, except for seeking after premeditated goals related to efficiency and productivity, e-government additionally establishes political and social goals, as the development of social inclusion and public trust (Grimsley & Meehan, 2007, p.134). In other words, the e-government's asset spectrum, ranges from improved intra-organizational changes to public value production through social participation.

Nevertheless, critical studies imply otherwise, pertaining that, instead of being an instrument that grants to citizens accessibility to information, by allowing them to enjoy public service delivery in an effective and approachable way, e-government is only an inadequate policy effort that underperforms in reconfiguring the public administration (Ahn & Bretschneider, 2011, p.414). In more detail, fault-finding scholars such as Bekkers & Homburg (2007), posit, that the E-government policies are only myths, promising an inevitable technologically driven public-sector innovation, that is responsive towards the citizen needs, as well as, more democratic and efficient through rational information planning (p. 373). Yet, the same scholars suggest that these promises have a small possibility of materializing. However, even the tales about e-government sum public value, as they present "*mobilizing capacities",* by incentivizing public officials to overcome bureaucratic stagnation and consider the prospects of technological progress (Bekkers and Homburg, 2007, p. 380). In other words, by reflecting on e-government, public officials and citizens alike, weary of the traditional administrative practice's banality, began thinking about a desired institutional innovation (Bekkers & Homburg, 2007). Thereby, it can be theorized, that the diffusion of the e-government services, established, for the first time, the rationale for the future arrival of technological innovations in the public administreation.

Innovative or customary for adding value to the public administration, one premise that the majority of scholars agrees upon, is that e-government established open information channels in the public sector, where data were travelling back and forth in an unprecedented manner. Compared to the time-consuming traditional information and service delivery, e-government increased the volume of government data and promoted a new era of policy, through the utilization of technology. As Bertot and colleagues (2014) argue, the emergence of e-government

in the early part of this century, introduced the demand for policies that address the role of public organizations on how they handle, allocate and store digital government information (p. 10). Respectively, agencies were required to circulate information rapidly, equitably, efficiently and in a practical manner (Bertot et al., 2014). Governments began to consider disparities of access to information and came up with alternative strategies for information dissemination (Bertot et al., 2014). Therefore, the never seen before, utilization of information technology trade began with the implementation of e-government services. In other words, governments around the globe, compelled by the e-government potential for enhanced public policymaking, decided to embrace such practices and conform their public organizations in accordance to the e-government standards. The effort towards endorsing this transformative form of government, forced public organizations to take into account the contingencies and setbacks that e-government services may bare.

A few of the complications related to the employment of ICTs in the name of e-government and open data, that public administrations confronted throughout the years, involve, low-grade quality of data and interoperability issues, as well as, lack of required capabilities and infrastructural resources (Jetzek, 2016). In order to resolve comparable problems, the public sector is required to acquire the necessary infrastructure, and in turn focus on equipping the civil service with the demanded information and data analysis-based skills (Jetzek 2016). As it will be made apparent later on, comparable sets of skills are assumed as being a requirement for the proper implementation of technological innovations; skills that if are absent, the technological potential remains unexploited. However, it is important to note that most countries are facing shortage of skills in the area of data management, as Jetzek (2016) argues.

From all the above, it can be deducted that a decade earlier to the installation of big data analytics into the public sector, the strategy towards achieving an effective e-government, was the trigger that incentivized public officials to acknowledge the significance of analytical skills, as the urgency for data analysis and management became apparent. Scholars like Reffat (2003), began to consider that public officials and civil servants need to obtain a required set of skills for rendering the e-government's advantages real. The model skill set is related with analytical and technical expertise that enables the civil servant to manipulate digital information and consequently, makes him or her competent in implementing the e-government policies effectively (Reffat, 2003). The demanded set of competences is identified within an array of particular skills, including analytical capacity, information and management expertise and technical skills. The presence of analytical skills is the initial precondition for the effective

implementation of E-government initiatives. Analytical skills are associated with the ability to interpret and identify processes and problems that may emerge concurrently to the practice of organizational operations. Analytical skills enable the interpretation and identification of problems and the corresponding organization processes that stand behind the complications (Reffat, 2003, p. 5). Therefore, by having analytical skills, the civil service is capable of sensing the symptoms of a complication and locating the problem that disturbs the organization. In turn, civil servants are capable of uncovering processes, imitating other organizations, and designing a system or solution for overcoming the problem (Reffat, 2003). Consequently, it is assumed that the civil service needs to be equipped with respective credentials in order to be capable of resorting to analytical tools, such as performance reviews, statistical trending and similar techniques when the need appears (Reffat, 2003, p. 5).

Although it is difficult to separate them in practice, except for the analytical skills, there are the Information Management skills, that are essential to collecting and treating the information as "*valuable organizational resource"* (Reffat, 2003, p.5). When the personnel of public agencies, is competent in information management skills, it is also acquainted with the "*content, quality, format, storage, transmission, accessibility, usability, security and the preservation of data*" (Reffat, 2003, p.5). Respectively, information management skills are part of the competency arsenal of various professions, such as program managers, IT professionals, archivists and researchers. Lastly, technical skills are an essential requirement for the use of e-government, as they make the organization capable of applying a chosen solution for counteracting an imminent problem, that is threatening the successful implementation of e-government (Reffat, 2003). Nevertheless, the exact description of the demanded technical skills varies in relation to the nature of the corresponding e-government challenge. To sum up, analysis and interpretation skills are essential for an e-government project to materialize, as they contribute to the evolution of the project, from its prototype design, to its evaluation, to its implementation. On the other hand, information management and technical skills help the organization on how the information is handled and utilized.

Thereby, the above skill set is posed as a necessary requirement for comprehending and implementing E-government. In other words, being outfitted with these skills equals being able to effectively analyze and utilizing information, as well as, being capable of employing problem-solving techniques for addressing an information-related complexity. Thus, it can be theorized, that the implementation of information and technology driven innovations demands a set of requirements, ranging from organizational reform to skillful personnel. In other words,

government transformation through the use of IT and innovative technologies, necessitates from the public organizations to be able to utilize the content of the IT projects. The e-government, as well as big data applications, are technological projects that rely on such premises, necessitating that requirements as such, need to be met, if the public-sector desires to accomplish its digital transformation. However, the implementation of e-government is not realized only by the presence of required skillsets within the public sector. Systemic factors can also influence the advancement and implementation of innovative technologies.

The Digital-Era Governance is a data-based theory that was built upon the footprint of e-government, and along with other data-driven doctrines such as Big and Open Linked Data (BOLD) can be considered as significant driver of government innovation (Giest, 2017, p. 368). Thereby, except for requirements that the civil service should satisfy at the individual skills level, related public sector development approaches, such as the DEG, are posed as systemic requirements for the transformation and digitization of the public sector. The Digital-Era Governance approach is the theory that reorganizes the post-New Public Management environment, by "*reintegrating neglected functions back into the governmental sphere, adopting holistic and needs-oriented structures, and progressing digitalization of the administrative processes*" (Dunleavy et al., 2006, p. 467). In contrast to the NPM, that was established upon a vision for an entrepreneurial public sector, that is lean, cost-effective, disaggregated, and competitively incentivized, due to the demolition of the public-private sector dichotomy (Denhardt & Denhardt, 2000), the DEG is a mode of governance that pays special attention on the importance of digital transformation, without neglecting the significance of the technological, organizational, social and cultural features for the revitalization of the public management (Dunleavy et al., 2006). In other words, the DEG promotes the effort to reconfigure the organizational relationships and processes, that were disintegrated during the NPM's reign. In this way, the DEG aims on establishing isocratic and robust digitalized public services, which are formulated on the basis of the citizen and social wishes (Margetts & Dunleavy, 2013, p. 6). However, in order to comprehend the impact of the possible advent of the DEG paradigm in a post-NPM environment, the theory behind the NPM should first be explained.

The New Public Management refers to a doctrine for public sector development that was associated with the reorganization of the public-sector bodies, with the aim of adopting the optimum business-driven methods (Dunleavy & Hood, 1994). Subjects of the NPM's influence were mostly the Anglo-Saxon countries, such as the UK, Australia and New Zealand, which are

referred as having had a considerable exposure to influence of the NPM doctrine (Hood, 1991). Components, such as "*hands-on professional management",* performance evaluation, output control, unit disaggregation, public sector competition, businesslike managerialism, and frugality (Hood, 1991), were some of the typical doctrinal characteristics of the NPM, that influenced the respective countries and brought an administrative shift within their public administrations. Moreover, the NPM stressed the importance of accountability and high-performance, by restructuring bureaucratic organizations and rethinking the organizational missions (Denhardt & Denhardt, 2000). In other words, under the NPM's reign, public bureaucracies sought to increase their productivity and sought to discover new alternative service-delivery mechanisms that were driven by public-choice assumptions (Denhardt & Denhardt, 2000). Concerning, technologically-driven government aspirations, it is important to note that even though the NPM supported the trend that favored the adoption of automation through the employment of information technology (Hood, 1991, p.3), its rest predominant aspects favored the managerial and business-like approach at the expense of the adoption of technological changes (Margetts & Dunleavy, 2013, p. 2). In other words, the introduction of ICT technology inside the public service had been relegated to the second place at the time, as the governments focused more on the rest of the NPM's aspects.

More recently, even though the DEG is not the completely adverse paradigm to the NPM, Dunleavy and colleagues (2006) theorize that the former has come to replace the latter as the dominant theory for public sector development. The DEG's distinctive feature, which advances the use of big data analytics, is its eagerness to establish patterns that are aligned with the contemporary technological trends transpiring inside the social, civil and corporate world. In more detail, the DEG considers that the public sector has to be fine-tuned on the basis of developments which are driven by the internet and the online social processes (Margetts & Dunleavy, 2013). Thereby, public entities which are placed under the DEG rule, are compelled by the paradigm to become parallel with the latest contemporary technological advancements. For instance, certain public organizations sync and update their operations in accordance with the most recent technological developments, such real-time data analysis systems, because they are interested in redesigning their public service provision towards a more agile and client-based perspective (Margetts & Dunleavy, 2013, p. 6). Considering the rationale behind the DEG, it can be assumed that the paradigm advocates the utilization of e-government and big-data analytics for the development of the public sector. That is because the DEG paradigm perceives such technological-driven innovations, as an integral component of the operations of a

modernized public sector. After all, the typical characteristics of the paradigm, and especially the needs-based holism and digitalization component, cannot be implemented, without the public sector being in line with the latest technological trends. Moreover, Margetts & Dunleavy (2013) argue, that the DEG's reintegration component presents big data analysis as a key element for public sector development and denotes that such technologies are already being used inside the public sector. Having inserted the big data concept, the next question that arises at the moment is, "what is big data?"

Big Data constitute the most recent stage of the Information Revolution (Richards & King, 2014, p. 393). Richards & King (2014) continue by arguing, that "*Big Data Revolution*" is comparable to the industrial revolution, as its impact will radically change the pre-big data society (p. 393). At the beginning of this revolution's timeline, the traditional government was converted into e-government by instituting information channels and employing the internet as an organizational tool. Nowadays, the public sector is getting ready to revolutionize once again, by employing Big Data Analytics. But what are the distinguishing characteristics of big data analytics, and why are they deemed so important for the society in general and for public administration in particular? To answer that question, Big data in public affairs is first defined as *"a combination of administrative data collected through traditional means and large-sale data sets created by sensors, computer networks or individuals as they use the internet"* (Mergel et al., 2016, p. 928). More generally, the term Big data refers, to the size of the newly emerging data sets, that contain manifold observations and variables, which take many forms, and are mined by continuous and automatic data collection processes from disparate sources (Mergel et al., 2016). An additional distinct feature of big data is that of its granularity (Mergel et al., 2016). More specifically, granularity refers to the fine-grained structure and the particularity of the data-sets. For example, when a big data set is comprised by specific, detailed or thoroughly subdivided information, it is considered more granular and subsequently enables the recording of even the most discrete attributes of the collected sample.

The previous attributes are also illustrated by the three Vs describing big data. Big data *volume* is the primary feature, which refers to the enormous size of the data sets, containing myriad bytes of information (Kim et al., 2014). Later on, *velocity* refers to speed at which data-sets are being mined and processed, while *variety,* at last, refers to diversified forms on which data sets are structured (Kim et al., 2014, p. 79). However, as the computing power capabilities of big data are extensive, the set of Vs extends beyond the initial three. The aspect of *veracity* is added to the assortment, as it indicates the liability of uncertainty, meaning that the data-sets' quality and

validity as a resource can be questioned (Giest, 2017, p. 368). Thereby, veracity raises concerns about the information's trustworthiness, as data can be unreliable and less controllable to manage (European Commission, 2016, p. 16). Additionally, the European Commission's report introduces the dimension of data *viability*, which describes how appropriate are the collected variables for a particular purpose, and how feasible is to analyze the data and extract insightful information (European Commission, 2016, p. 16). Finally, *value* is the last of the Big data's characteristics, implying that big data presents a considerable potential for providing invaluable insights for improving nearly every step of the policy lifecycle (European Commission, 2016, p. 16). After all, the value deriving from big data analytics, is what makes governments worldwide to consider adopting data-driven technologies in the first place. Consequently, it can be theorized that *value* is the driving force that incentivizes governments to reestablish their organizational processes and prepare their civil service workforce for utilizing big data, so to leverage the value that the technology bears.

For a camp of scholars, the arrival of big data analytics, signals an era of rapid and optimal policy decisions; a time period where governments no longer rely on collecting samples for analyzing information, but process data in real time, as big data augments the public organizations' methodological tool case (Höchtl et al., 2016, p. 152). The benefits that big data applications bare, make governments to perceive them as the desired norm for processing information in the public sector (Kim et al., 2014, p. 84). Respectively public organizations consider them as an essential instrument for enhancing their output delivery. In more detail, the advantages derived from the appropriate utilization of big data applications range from promises for increased effectiveness and government responsiveness, as public organizations exploit the big data's velocity, to opportunities for refined transparency, collaboration and public participation (Chen & Hsieh, 2014). In other words, big data utilization allows public organizations to become more agile, enabling them to respond dynamically to potential contingencies, while at the same time, it also supports objectives of different nature, by encouraging governments to become more transparent by undertaking open data initiatives (Chen & Hsieh, 2014).

Accordingly, Open data initiatives are associated with big data and improvement in government transparency, as data are being published openly. By accessing open data, citizens can exploit, scrutinize, and repurpose this data into new products (Bertot et al, 2014, p.6). Additionally, big data present a promising opportunity for improving the delivery of public services, where the citizen is at the center of concern (Chen & Hsieh, 2014). This can be justified by the fact that, big

data allow the rapid collection and dissemination of data. In turn these aspects indicate a shift, where public organizations are not anymore "*input-centric*" receivers of information, but instead they utilize big data to measure government performance and make the public sector more accessible for everyone (O' Malley, 2014, p. 555). In this way, governments are focusing on becoming more performance-driven, more diaphanous and open to scrutiny and lastly more cooperative and interactive, as they descend deeper into the digital age of governing (O'Malley, 2014). Simultaneously, governments distance themselves from the ideological, hierarchical and traditional bureaucratic paradigms (O'Malley, 2014). Nevertheless, the benefits are accompanied by caveats and challenges, which may emerge when public organizations are not prepared for approaching such immense big data sets.

The call for the digitalization of the public sector introduces complications, which can impose detrimental effects on the public administration and act as an obstacle, preventing the smooth utilization of big data analytics in the public sector environment. Margetts and Sutcliffe (2013) point that even though big data analytics offer extraordinary possibilities, they also set a lengthy list of potential challenges. The extended list stretches from ethical considerations about possible privacy violation, to suboptimal policymaking that impairs civil equity. Regarding operational considerations, the list continues from technical challenges about infrastructural and computational deficiencies, to doubts about how data can actually be utilized by analysts within the civil service (Margetts and Sutcliffe, 2013). Except for these considerations, Lavertu (2016) adds that technological advances, such as big data applications, can create the problem of goal displacement inside a public organization, as failed measurements impel certain goals at the expense of others. Public administrations globally are requested to address these challenges, as the demand for big data implementation intensifies. However, as the implementation need increases, the complications and challenges increase proportionally. In more detail, Chen and Hsieh (2014), claim that governments do not only have to confront internal contingencies related to the urgency for establishing a data-driven policymaking culture, but they also have to encounter challenges, that result from the underdeveloped and immature big data applications; Additionally, public administrations need to address concerns for limited resource availability and shortages of required skills in order to become capable of utilizing big data analytics (Chen & Hsieh, 2014).

At this point, it is apparent that the unfamiliarity with big data analytics, due to deficiencies in competences and skills demanded for utilizing big data sets, can impose serious consequences upon the public sector. As it was stated previously, the lack of individual skills within the civil

service can impair a public organization's operations and lead to suboptimal policymaking outputs or partial policy implementation. Subsequently, governments intimidated by the challenges that accompany big data analytics, may reconsider their transformation towards a more digitalized future. As Klievink argues, uncertainty, produced by the big data challenges and threats, compels public organizations to rethink whether they possess the capacity and the know-how for utilizing big data applications (Klievink et al., 2017, p. 268). Thereby, big data analytics do not only provide predictive insights and precise measurements, but they also generate a sense of ambiguity and paralysis, as public organizations are unsure if they are equipped with the essential tools for taking advantage of the big data opportunity. This intimidating impression that can exist inside the public sector, created by the possible caveats and barriers that accompany big data technologies, may supposedly affect adversely the implementation of big data applications.

In sum, by inspecting the literature on e-government, public-sector development and big data use in the public sector, a common point that is drawn, is that a civil servant who is working inside a public organization, needs to be equipped with the needed analytical skills and competences in order to be able to utilize data and eventually avoid getting overwhelmed by the technology that is going to be implemented. As Mergel and colleagues (2016) argue, civil service personnel needs to the have the capacity to "*manage and process large accumulation of unstructured, semi- structured and structured data; analyze the data into meaningful insights for public operations; and interpret that data in ways that support evidence-based policymaking" (p. 934).* By looking at this statement, the term of policy analytical capacity can be assumed as the link that connects the literature between big data use in the public sector and evidence-based policymaking, as it is going to be presented in the next section. However, besides the concept of policy analytical capacity, as a skill requirement at the individual level, it is safe to assume that the DEG framework, at the systemic level, is also linked directly Evidence-based policy making and the use of new technologies and big data streams (Giest, 2017, 378). Next, in order to comprehend the connotations between evidence-based policymaking and big data use, I should first provide a detailed description of what the former is.

2.2 Policy analytical capacity as the link between evidence-based policy making and big data analytics

Evidence-based policymaking shares expectations with the DEG approach, as the paradigm allows policymaking practices that are based on rigorously identified objective evidence, while both are directly associated with the use big data applications in the public sector. This argument is expressly specified by Giest (2017), supporting that the theoretical concepts of DEG and Evidence Based-policymaking "*link directly to public use of new technologies and big data streams" (p. 378),* in other words, the concepts are interdependent, yet each of them displays distinct opportunities and contingencies for the employment of big data technologies in the public sector. The reason that binds this conceptual relationship, is that big data analytics insert novel types of data-formats and evidence into the policy cycle. The accumulated data-sets are welcomed as possibly meaningful input for the evidence-based policymaking cycle. This input gets translated into concrete evidence and has the potential of providing essential insights for the policymaking output's development. Thereby, evidence-based policy making is theorized as a practice for improving the policymaking process by relying on tangible data. In more detail, Howlett (2009) defines evidence-based policy-making as the "*attempt to enlarge the possibility of policy success by trying to improve the volume and type of information processed in public policy decision-making, as well as, the methods used in its assessment*" (p. 157). However, does the literature support that the evidence-based policy-making can contribute to better policy outputs?

The rationale behind evidence-based policy making, supports that the best policymaking practice, is the one supported on unlimited data availability. In other words, evidence-based policymaking entails a positive relation between large data availability in the public sector and the achievement of optimal policy outcomes. This relationship can be described by the fact that evidence-based policymaking, expects that, the more the information available, the more it can be integrated into the policy cycle and eventually more sharpened decisions will turn up as a result (Howlett, 2009, p. 157). Therefore, accumulated information, provides greater insights and in turn allows the employment of more sophisticated methods that produce more precise results. The large pool of information that is required by evidence-based policymaking, can be acquired by utilizing big data analytics. Big data are characterized by a large volume of information, so they serve as a warehouse of information and evidence. According to Head (2008), governments invest in searching for more productive and cost-effective means for mining and processing large-data sets. Thereby, big data presents potential for closing the governments' data gaps.

Nevertheless, the quality and reliability of the collected evidence, as well as, performance of evidence-based policy making in closing these information gaps, is occasionally questioned by some scholars. As described, critical academics claim that the obtainment of more data for filling the information void is futile. For example, Head (2008) argues that more data does not necessarily equal to optimal policy solutions, as a policy puzzle demands the reconcilement of perspectives of different value. Additionally, scholars advocate that the policy lens through which evidence is inspected can distort the data's quality and in turn can affect the output of evidence-based policymaking. More specifically, political know-how and judgement can influence the way of how data is interpreted, and in turn they might contaminate the evidence's objectivity, with partisan, tactical, causal or opportunistic intents (Head, 2008, p. 5). Alternatively to the ideal purpose of evidence-based policymaking, one can claim that the use of big data inside the public sector, can be rendered inconsequential, as politics are capable of exploiting the technologically generated information, as a resource for persuasion rather as an instrument for objective decisions. For example, Sanderson (2002, p.5) argues that politicians often disregard technical evidence and undermine evidence-based policymaking, because they prefer information that is essential for them to get reelected. Instances like these, compel us to consider that objective scientific information that is derived from big data, "*does not automatically get translated into better policies*", as evidence eventually become the byproduct of political and ideological-driven priorities (Höchtl et al., 2016, p. 157). To sum up, Höchtl and colleagues (2016) conclude that the process of policymaking can still be affected by the political factor, yet the utilization of technological advancements reduces the time frame and enlarges the evidence based for formulating policy decisions (p.148). Similarly, Head (2008, p. 3) admits that, although the evidence-based policymaking does automatically foster optimal policy outcomes, governments need to invest in establishing information banks for scientists and decision-makers, as data availability can be helpful for confronting complex social problems.

Ideally, the evidence-based policymaking practice fits within the rational decision-making model, while, the ideal form of knowledge used to provide a firm basis for policymaking, "*is seen as derived through quantitative methodologies that are empirically tested and validated*" (Sanderson, 2002, p. 6). Additionally, Sanderson (2002) implies that evidence-based policymaking, emphasizes the need for more and better data, and familiarity with research designs and especially acquittance with the quantitative techniques; a theory that was operationally put into practice by the UK Cabinet office during the previous decade (p. 6). From this observation, the connection between evidence-based policymaking and policy analytical

capacity can be justified, as the later refers to the government's ability of applying such methodologies and techniques (Howlett, 2009). Thereby, it is assumed that public administrations, require a considerable level of policy analytical capacity in order to perform tasks that are essential to implementing evidence-based policymaking (Howlett, 2009, p. 153). In other words, the presence of substantial policy analytical capacity is considered as "*an essential precondition for the adoption of evidence-based policymaking and the improvement of policy outcomes through its application" (*Howlett, 2009, p. 161*). By* taking into account these premises, it can be assumed that policy analytical capacity is not only an enabler of evidence-based policymaking, but as big data in the public sector stands as a foundation that supports the evidence-base, it can also be hypothesized that policy analytical capacity can be an enabler of big data utilization. Alternatively, it can be assumed, that the lack of policy analytical capacity can generate serious implications for the public sector and cause possible hinderances to big data adoption. In order to make this assumption more comprehensible, the concept of policy analytical capacity should first be described thoroughly.

According to Howlett (2009), policy capacity is an umbrella concept that covers a broad range of components that are related with the government's ability "*to review, formulate and implement policies within its jurisdiction"* (p. 161). From this definition, it is evident that policy capacity is related with the government's operations in each stage of the policymaking cycle. Thereby, policy capacity can be placed within an extensive area of policymaking activities, as it related with the public administration's ability of executing its day-to-day operations. However, the concept of *policy analytical capacity, "is a more focused concept, related to knowledge acquisition and utilization of the policy processes"* (Howlett, 2009, p. 162). In more detail, the concept refers to government's ability of applying statistical or applied research methods, advanced modeling techniques, and analytical techniques, such as trend analysis, and predictive methods, by using gathered data with the purpose of gauging public opinion (Howlett, 2009). Thereby, policy analytical capacity, is essential when a government aims on anticipating future policy impacts, avoiding policy failures and improving its policymaking outcomes. The skills and capabilities that individuals who are working in public organizations possess, are the foundation for building policy analytical capacity. The level of the civil service's skills determines the extent to which the corresponding administrations can employ applied analytical techniques. The more equipped the civil servants are, with the required skills relevant to policy analytical capacity, the more refined the methods public organizations can use.

It is important to note, that although the study's scope is to examine policy analytical capacity at the level of the civil servant, it must be noted that the concept is dispersed throughout the system of government. Specifically, scholars identify different levels upon which policy analytical capacity is considered as vital determinant for attaining a good policy outcome (Wu, Ramesh & Howlett, 2015). As it was acclaimed previously, policy analytical capacity at the individual level, refers to the individual's ability to analyze problems, implement policies and evaluate their impact. Nevertheless, the concept is not only restrained at the micro-level but extends to the meso-level; that of the organization. For policy analytical capacity to exist at the organizational level, it requires the presence of an established enabling institutional, economic and informational context that functions as a foundation for helping the situated individual to perform the necessary tasks that are relevant to policy analysis (Hsu, 2015). One step higher, to the systemic macro-level, the country's state on educational and research institutions and facilities and the government's freedom in enjoying unrestricted access and direct availability to high quality data, are considered as variables that affect a government's policy analytical capacity dynamic (Hsu, 2015). In other words, it is not only up to the civil servant to present adequate policy analytical skills in order to produce a good policy outcome; the civil servant's policy analytical capacity is also influenced by the organizational environment, inside which he or she is operating and the state's higher-education and research standards. Finally, it can be assumed that the propagation and effectiveness of policy analytical capacity is contingent on a bidirectional relationship between the individual, the organizational and the systemic level; where the quality of policy analytical capacity that a civil servant has, depends on whether policy analytical capacity exists on the two other levels, while when the policy analytical capacity is already present at the individual level, the particular condition positively affects the government's general capacity on policy analysis.

Summing up, the presence or absence of analytical capacity is a variable that influences the overall governmental capacity and in turn affects the government's performance on delivering policy outputs (Howlett, 2015, p. 174). A few of the concept's fundamental components, which can be utilized when public officials are equipped with considerable levels of policy analytical capacity, are tool evaluation, forecasting methods, trend analysis, statistics and applied research (Riddell 2007 as cited in Howlett, 2009, p. 164). Accordingly, it can be assumed that the individuals who are working inside the public sector need to be outfitted with analytical capacities and technical expertise, in order to be capable of collecting, managing and analyzing data derived from the employment of the big data technology. Altogether, the concept of policy

analytical capacity is intrinsically related with data-utilization capabilities, as having a considerable degree of policy analytical capacity, is equal to having the ability of making use of large-data sets and in turn being able to repurpose the information by employing it for applying complex technologically-driven techniques. In other words, policy analytical capacity, depending on its degree and quality, can be considered as an enabler to the way of how big data analytics are being assimilated inside the public sector.

However, except for the impact that policy analytical capacity can have upon big data assimilation, other factors, such as the mode of the intra-organizational governance within an organization, or the data science expertise that a public organization displays at its workforce level, can also be considered as facilitators to the employment of data-driven technologies. According to Klievink et al., (2017), the presence of data and IT governance and data science expertise within an organization, is an indicator that implies that a public organization has the required organizational capabilities and the capacity to exploit the big data's complete, being able to derive meaningful content from it, while avoiding getting overwhelmed (Klievink et al., 2017, p. 273). Moreover, IT knowledge placed at the managerial level, is considered to be positively related with the assimilation of technology related to IT technology and data manipulation (Ranganathan et al., 2004). Thereby, except for policy analytical capacity, digital literacy can be hypothesized as an additional enabler of data-driven technology assimilation. Similar to policy analytical capacity, the construct of digital literacy prescribes skills that are related with an array of competences that support the utilization of data-driven technologies. Namely, the ability of accumulating reliable information, critical thinking skills for evaluating the validity of the information sources, analytical skills for evaluating the credibility of digital information, computer-based skills and general awareness for exploiting digital tools to the fullest are some of the competences of an individual that displays digital literacy (Bawden, 2008). In sum, throughout this section, the possible factors that can affect the assimilation of big data technologies were mentioned in detail; it is now time to examine the concept of big data assimilation.

2.3 From Diffusion of innovation to innovation adoption and assimilation

In order to further understand the process of big data assimilation beyond the effects of policy analytical capacity and surrogative skills, in this section I will analyze the conceptual

differentiation between innovation adoption and assimilation, as well as, the steps that are taken for an innovation to be implemented. As, it was noted earlier in the literature review, the public sector has transited through different waves of change throughout its existence in order to improve its output and fulfil its mission. The New Public Management reforms were innovative at the time, as they supported a business-like form of government and at the same time introduced the benefits of automatization. As Windrum & Koch (2008) suggest, the NPM doctrine has been a major force for ushering organizational reforms inside the public-sector services. Afterwards, the emergence of the E-government transformed the public administration to a more efficient, effective and citizen friendly environment, where public service delivery was radically altered and improved on the basis of the advantages that the information technology had to offer. Subsequently, public organizations reestablished their institutional structures, procedures and technologies in order to support such innovations. Then, Digital-Era governance arrived to stand as a pillar to this innovative endeavor. Each of these changes constitute an example, where the public sector welcomed technological and organizational innovation. Currently, the advent of big data technologies is no different. Public administrations are asked again to incorporate the big data technology in order to improve their output and address the contemporary wicked problems that challenge the today's society. However, it is first crucial for public officials to develop a complete understanding of how the process of innovation takes place inside the governments, in order to be prepared for welcoming the big data technology, exploiting its full advantage and avoiding any potential obstacles to its application.

Innovation in the public sector is not an automated process but a sequence of actions. Innovation in governments takes place in particular favoring organizational settings and can be influenced positively or negatively by various internal and external factors to the public organization. For the time, lets first define what innovation is. According to Rogers, innovation is "*an idea, practice or object",* that is conceived as new by an individual or another unit of adoption (Rogers, 1995, p. 11). A similar definition is offered by Walker, who sees innovation as a process through which new ideas and practices are constituted, developed or reinvented and are new to the recipients (Walker, 2008, p. 592). However, he insists that innovation cannot be confined as a perception of an idea, but it requires its implementation to occur (Walker, 2008). More specifically, Windrum & Koch (2008), taxonomize public sector innovation in various types, and insist that service, organizational and administrative innovation, are about the introduction of a new service or product, and the improvement or alteration of the already existing structures in the public sector. Although, the OECD defines innovation from the private

perspective, yet it similarly introduces a semantically relevant definition, as it describes innovation, as the *implementation* of new or significantly improved product (good or service) or a new organizational orientation (OECD, Eurostat & EU, 1997).

By reading these definitions, it is apparent that there are generally two common elements. First, except for the definition of Windrum & Koch (2008), an innovation is required to be implemented in order to be characterized as such. A valuable idea itself is not enough to be considered an innovation, but it needs to be operationally put into action (Daglio et al., 2015). Secondly, scholars such as, Rogers (1995), Walker (2008) and Windrum & Koch (2008), as well as the corporate insights, stress the importance of the distinguishing characteristic of innovation, that of *novelty*. Innovation must be new to the organization. The *newness* element of innovation may be expressed in terms of knowledge, persuasion or a decision to adopt (Rogers, 1995, p. 11). Additionally, the OECD (1997) adds and another one element to innovation, that of its *impact*. When referring to innovation in the public sector, the innovation's impact is an important aspect, as something cannot be considered as an innovation, if it does not create public value; for example, by addressing a public policy challenge or achieving a better outcome through its improved efficiency and effectiveness (Daglio et al., 2015, p.4). After identifying these emerging propositions and rethinking about the object of the study, can one consider big data as a public-sector innovation?

Resting on the aspects that were also noted in the beginning of the literature review, we can infer that big data technologies can be characterized as a public-sector innovation, as they satisfy all of the three above principles. Big data can be considered as an innovation because it introduces novel and reinventing approaches into the public organizations. Big data inserts new organizational changes to the public sector and pushes organizations and agencies to reestablish their structure and technological apparatus, in order to able to collect and distribute data with the aim of improving both organizational performance and public service delivery. Big data, therefore, is a novel technology as it signatures a change from traditional bureaucratic government, towards a more cognizant, interactive and performance-based style of governing (O'Malley, 2014), while it reinvents and enhances the opportunities that e-government offered beforehand. Secondly, big data is not just an idea that remained inside the minds of computer engineers and public officials, but a reality, as governments around the world have initiated big data applications projects in order to address real life challenges in economy, welfare, health care or safety and security (Kim et al., 2014). Lastly, big data can be conceived as an innovation because it creates public value and has a beneficial impact upon the public administration and

the society, at least according to the voices of its proponents. Moore (1995) introduced public value as the value created by the activities of public managers, while other scholars have defined public value as the value that originates by the government through its services, regulations and other actions (O'Flynn, 2007). Public value is synonymous to enhancing the government itself as an asset, and to offering benefits to the society. Under this assumption, big data can be considered as a producer of public value, as it promises greater government efficiency and effectiveness. Simultaneously, since public value has an impact on public affairs, it can be justified as an innovation because it aims on attaining better public policy results and supporting both the government and the society.

Having already clarified the innovative aspect of big data technology, now it is time to review the process of how big data applications are dispersed, and later adopted across the public organizations. According to the diffusion of innovation theory, before an innovation becomes the norm it is initially distributed through certain communication channels over time and among the members of a social system (Rogers, 1995, p.4). In order to understand the process of innovation adoption, it must first be noted that innovation does not happen in a vacuum, but it is a product of evolution rather than radical change (Walker, 2008, p. 592). Innovations usually originate from the external environment to the organization and the new practices and ideas are disseminated across public organizations through the process of imitation (Walker, 2008, p.592). Actual innovation adoption of IT technologies is the result of a response to a change in the environment, in which the organization operates, or when the innovation becomes a requirement for the organization's operation or by the belief that it will enhance its output performance (Hameed et al., 2012, p. 218). Eventually, IT innovations are adopted because they enable the organization to upgrade their productivity and quality (Lee & Xia, 2006)

However, it is important to make a distinction between the innovation adoption and assimilation. The adoption of innovation is the prerequisite for the assimilation; or else a stepping stone towards it; "*it is the process that results in the assimilation of a product, process or practice that is new to the adopting organization"* (Damanpour & Schneider, 2009, p. 497). Adoption, therefore, is a multistep process, where a decision-making unit moves onwards from first knowledge about the innovation, to developing a stance towards the innovation, to a decision to adopt or to reject and finally to the implementation stage that is dependent on the previous decision (Rogers, 1995; Frambach & Schillewaert, 2002). The innovation adoption has two distinguished phases, according to other scholars, that of initiation and that of implementation (Damanpour & Schneider, 2009). The implementation stage is a crucial

component of the adoption, as an innovation is not truly adopted when it has been initiated, but instead it is adopted when it has been actually put into practice by the organization (Damanpour & Schneider, 2009).

On the other hand, assimilation is qualitatively different, as it refers to an intra-organizational process which expands from the initial alertness about the innovation, to the potentially formal adoption and most importantly to the *full-scale deployment* of the innovation (Pudjianto & Hangjung, 2009). According to Purvis*, "assimilation is displayed by the extent to which the use of the technology diffuses across the organizational projects or work process and becomes routinized in the activation of those processes"* (Purvis et al., 2001, p. 121). Meyer, in his definition of the organizational assimilation of technological innovations, except for the already defined characteristic, that the assimilation process initiates when organizational members become aware of it, he adds that assimilation leads to the acquisition of the innovation and becomes reality when the organization has fully accepted, utilized and institutionalized the innovation (Meyer & Goes, 1988, p. 897). In sum, the common element in all the definitions, is that assimilation is achieved when the technology is not only selected and accepted, as in the adoption, but it is fully embraced and totally ratified within the public organization, becoming a part of its standard procedures. Assimilation does not suffice to just the implementation stage, as the adoption does. Assimilation and adoption thereby are semantically relevant but not identical, as assimilation takes the process of technological approval one step further and demands the complete routinization and institutionalization of the innovative technology.

2.4 Barriers to innovation equals barriers to big data assimilation

By this time, it is established that big data analytics is an innovation that is willing to transform both the governance and the public sector. Public organizations are therefore required to develop a welcoming attitude towards big data, while they are "forced" to reestablish their structures, enhance their IT expertise and re-cultivate their culture towards the new technologies once again. Therefore, just as any other innovation, organizations in their attempt to achieve an exhaustive assimilation of the big data technologies, are required to reach an array of new standards and satisfy the new demands for change. However, in their endeavor to transform, public organizations may stumble upon certain barriers to adopting the new technologies. Throughout the years, the literature about e-government, IT innovation and

transformational governance has recognized various barriers to technological adoption by the public sector; barriers that correspondingly, have the potential to stand as impediments to big data assimilation.

In their research about the impediments to e-government adoption, Ebrahim and Irani (2005) display some of the key obstacles when the public sector adopts a new technology. The researchers identify multiple barriers and classify them along the dimensions of IT infrastructure and IT skills, security and privacy and that of organizational aspects. In the dimension of IT infrastructure, the lack of resources standards and common policies, in combination with the absence of integration across the government systems and the inadequate knowledge and the high complexity of the new technologies, are factors that form a setting that restricts the embracement of new technological innovations (Ebrahim & Irani, 2005). In addition to the security and privacy threats, an organization may also face difficulties in the dimension of IT skills, as factors such as the existence of unskilled staff, the lack of IT training, the shortage of expert IT staff both inside the job market and the organization are making matters worse for the successful exploitation of an IT technology. Additionally, organizational barriers cannot be neglected, as the lack of coordination between departments and the influence of politics also have an impact on innovation adoption (Ebrahim & Irani, 2005).

Similar findings but in a different classification are noted in the research of Van Veenstra, Klievink and Janssen, who identify impediments to transformational government; that is the government that realizes profound changes in the public-sector practices and structures that are usually so influential that are irreversible (Van Veenstra et al., 2011). The impediments are divided along three categories. In the governance category, some of the barriers to transformation are the insufficient IT governance, the structure of the public sector itself, the incompatibility of the public-sector culture towards innovation and negative political pressure (Van Veenstra et al., 2011). In the organizational and managerial category, the lack of IT skills and personnel, the lack of coordination and collaboration, the lack of organizational readiness and fragmented organizational structure are the key obstacles. Lastly, in the technical category, the complexity of incompatibility, security threats and the lack of architecture standards for interoperability are the crucial factors that hinder the assimilation of IT technologies.

Most of the authors who examined impediments to the adoption of new technologies by the public sector, adapt their classification in accordance to Ebrahim & Irani, and support their claims, by dividing the hindering factors along the same three categories; the *Institutional and political, the Organizational and managerial* and third *the Technological* (Van Veenstra et al.,

2011; Jetzek et al., 2014; Savoldelli et al., 2014; Schedler et al., 2017). The main barriers across the literature are in agreement. Based on the synthesized literature the following table presents the main obstacles to adoption of an innovative IT technology by the public sector.

Table 1: Barriers to the adoption of IT innovations

Type of Barrier	Barrier
Institutional and Political	Politics and political impact Insufficient IT governance Structure of the Public Sector Public sector culture Lack of trust and transparency Lack of citizen participation Digital divide
Managerial and Organizational	Lack of skills Shortage of qualified staff Lack of IT training programs Lack of organizational readiness Lack of coordination between departments
Technical	Lack of knowledge regarding technological interoperability Lack of infrastructural readiness High complexity Security threats Lack of open sources software

Sources: *(Ebrahim & Irani, 2005, Van Veenstra et al, 2011; Jetzek et al, 2014; Savoldelli et al., 2014; Schedler et al., 2017).*

Public organizations in their endeavor to incorporate and leverage the advantages of big data technologies may encounter such barriers from each of the respective categories. Regarding the institutional and political barriers, it can be assumed that institutional arrangements and political mandate can influence the everyday behavior of the public officials and affect negatively their dispositions towards the adoption of big data technology (Schedler et al., 2017*)*. However, it must be noted that this relationship can work the other way around. Additionally, concerning the factor of transparency, the innovations that are applied in the public sector are subject to public scrutiny (Albury, 2005), in contrast to the private sector. Therefore, transparency is a required value when adopting new technologies, as citizens tend to assess the necessity of a new public-sector innovation. Under conditions in which the citizens do not approve the adoption of new technology, institutions and organizations, who are accountable to the citizens, are obliged to rethink their stance towards adopting innovation; a situation that can inhibit the process of assimilation and can also be relatable to the case of big data assimilation. Finally, the structure and the culture that is developed within the public sector can be rigid and uncompromising towards the new innovational changes.

Regarding the managerial and organizational impediments (*see Table 1)*, given the requirements that big data imposes, organizations are in a high demand for data scientists, and IT technology experts. Organizations whose workforce is not equipped with the necessary data processing skills and is not familiarized with handling state of the art technologies, may be incapable of fully exploiting the public value that big data has to offer; and subsequently low performance may create a disfavoring attitude towards big data assimilation. Additionally, these shortages have an impact on the overall organizational structure and result to an organization that may be unprepared for technological changes. Similarly, in the technical category, factors such as inadequate digital and hardware infrastructures, may impose practical impediments to the organizations. The complexity of the big data sets can overwhelm the capacities of vulnerable data processing systems, creating computational difficulties and leading to ineffective outputs for the organizations. Additionally, the organization has to consider the interoperability issue, as different digital platforms between the organizations or internally between the departments may cause problems that can make the new technology unproductive and unwanted. Lastly, security and privacy considerations may create challenges for protecting vital information, and the fear of the organization becoming exposed can revert the organization to the use of conventional methods of information processing; an operation that is inconsistent with the adoption of innovation.

2.5 Hypotheses and causal mechanisms

To recapitulate, the literature review by far, has established the theoretical foundations advocating for the inclusion of technological innovations inside the public sector. In more detail, the study indicates that by constructing information channels, the e-government strategy compelled the public sector to reconstitute its operations and services and raised focus on the importance of analytical skills within the public sector. Thereby, the e-government strategy, was the spark that incentivized public officials to acknowledge the importance of analytical capacity, as the imperativeness for data management and analysis became sensible. Respectively, the DEG approach, a data-driven theory for public sector development, that is built upon the same premises as e-government, considers the integration of technological trends as an integral part of the modern public administration. Subsequently, it can be assumed that the DEG might have an impact upon the assimilation of innovative technologies.

Nevertheless, the use big data analytics inside the public sector is additionally mobilized by the evidence-based policymaking practice. Thereby, in terms that the evidence-based policymaking introduces, big data analytics is perceived as a rational quantitative methodology for improving the government's operation, as the output of big data is based on hard facts. The main argument of the study claims that evidence-based policymaking and big data use inside the public sector are associated. The connective link between them, is the concept of policy analytical capacity. Based on Howlett (2009), policy analytical capacity is a crucial requirement for the establishment of evidence-based policymaking, and the improvement of policy outcomes through its application cannot be done otherwise. Respectively, policy analytical capacity, is in turn, assumed to be positively associated with big data assimilation. Thereby, policy analytical capacity stands as the main independent variable of the study, and big data assimilation is the outcome of interest.

Except for policy analytical capacity, the study examines the impact of alternative possible factors to big data use inside the public sector, as digital literacy is examined as a surrogate to the main independent variable. Accordingly, the possible influence that the DEG might have on big data assimilation at the systemic level cannot be omitted. Having distilled the literature and having explored the potential causal mechanisms, that are backed up by theoretical justifications, it is now time to state the study’s research hypotheses.

The main hypothesis of the study is the following:

Hypothesis 1: *The individual policy analytical capacity, that civil servants inside the UK public sector have, is an enabler of big data assimilation*.

The hypothesis is based on the assumption prescribing that individuals who own the skill of policy analytical capacity, also present the ability to utilize quantitative and data-driven methods (Howlett, 2015, p. 178), and subsequently are able of handling elaborately big data sets and helping their organization to successfully assimilate big data applications. The causal mechanism that can explain this narrative suggests, that civil servants who display analytical capacity, are also capable of applying advanced computational methods and using analytical techniques. In consequence, it is assumed that these highly competent civil servants can rigorously employ complex big data applications with the aim of exploiting their promising value to the fullest. Therefore, a public organization that can enjoy the extensive potential of its big data applications, due to its advantage of presenting policy analytical capacities, is also more prone to assimilating the big data technologies. In other words, the presence of policy analytical skills within a public organization's workforce, leads to the exhaustive utilization of the big data applications' value, and in turn it provokes the organization to incorporate big data applications as an integral part of its day-to-day operations.

However, as the aim of the study is also to uncover alternative causal paths that can lead to the assimilation of big data applications, except for the impact of policy analytical capacity, alternative independent variables are going to be examined as well. Thereby, the variable of the digital literacy is introduced into the research design, as digital literacy can be considered, as the basic prerequisite that allows the individual to get familiarized with digital tools and be able to utilize digital resources. According to Martin (2006), digital literacy is defined as the "*awareness, attitude and ability of individuals to use digital tools and facilities, in order to identify, access, manage, integrate, evaluate, analyze and synthesize digital resources, construct new knowledge, create media expressions and communicate with others [..] in order to enable constructive social action" (p. 155);* By reading this definition, one can recognize that the concepts of *policy analytical capacity* and *digital literacy* are semantically related. However, unlike the concept of policy analytical capacity, that can be implied that it is only about "technical and analytical skills", as it deals with the individuals ability to employ advanced statistical and information-based methods, the concept of digital literacy is not limited to the ability of employing complex tools and understanding policy, but it is a more broader term, as it refers to the ability of an individual to master ideas driven by the digital world (Bawden, 2008).

Digital literacy, therefore, is about the mentality that an individual develops towards the usage of digital resources in general.

Consequently, it can be assumed that digital literacy can be an alternative factor that might influence big data assimilation, as it can also be hypothesized that for the successful exploitation of digital and data-driven applications, a basic digital-friendly disposition and knowledge is needed as a prerequisite. This assumption can be supported by several scholars, such as Purvis, Sambamurthy and Zmud (2001), that found IT knowledge, at least at the managerial level, to be an effective enabler of IT assimilation (Ranganathan et al., 2004). Thereby, the assumption suggests that a civil servant should not only be able of employing complex techniques as, trend analysis, statistics or applied forecasting and evaluation methods (as the concept of policy analytical capacity suggests), but, he or she should first be equipped with literacy per se, ICT literacy, and central digital competencies for utilizing digital formats, and creating, communicating and evaluating digital information (Bawden, 2008, p. 29). In other words, a civil servant that already has excelled his/her digital skills and has developed a friendly mentality towards utilizing digital applications, might be also able to utilize the potential of the big data applications. In order to strengthen the rigor of the research methodology, the alternative hypothesis implies that policy analytical capacity might even be redundant towards achieving big data assimilation. Respectively, the alternative hypothesis is the following:

Hypothesis 2: *the digital literacy, that civil servants in the UK public-sector display, is an enabler of big data assimilation*.

Except for the policy analytical capacity and the degree of digital literacy, the dominant paradigm for public sector development, can be considered as a systemic factor to the public sector that may also influence whether big data applications are assimilated completely. For instance, in contrast to the traditional process-oriented and hierarchical bureaucratic paradigm, or the results-oriented and entrepreneurial NPM approach, the DEG wave, aims at shaping more flexible and innovative bureaucracies. More specifically, the digital-era civil service is open to collaboration internally, across siloed units of government and externally, by establishing information channels across various levels of government, as well as, between non-governmental actors (Clarke et al., 2017, p. 458). Under the conditions that the DEG introduces, the civil service re-cultivates its values and utilizes practices that are associated with the era of information and technology. Openness, collaboration, peer-production and citizen involvement through the establishment of information channels, as well as, receptiveness towards the use of

technological innovations, such as big data applications, are some of the key characteristics of the DEG approach (Margetts & Dunleavy, 2013).

By observing the paradigm's key features, one can assume that the DEG approach creates a friendlier environment for the advent of big data applications inside the public administration. In other words, the DEG might contribute to the assimilation of big data applications, by restructuring the organizational processes and reinventing the purpose of the public sector. More specifically, the transformation that the DEG endorses, aims on inserting more open, joined-up and digital features inside the government (Margetts & Dunleavy, 2013). Under the reign of the DEG, the public sector will be required to adapt to the technological demands and revamp its information technology procurement with the aim of upgrading the government workforce. Thereby, it can be hypothesized that the DEG introduces a technology-oriented attitude to the public sector; inside which the utilization of big data applications can be made effortlessly. Subsequently, the DEG approach contributes to the assimilation of big data applications, as it introduces the most appropriate settings for the operation of a data-oriented public sector that aims at incorporating technologies, to use them in new ways for improving the policy outputs.

Regarding the above premises, the following alternative hypothesis is developed.

Hypothesis 3: *Digital-Era Governance has been established inside the UK public sector and functions as a driver for big data assimilation*

By introducing the variable of the paradigm for public sector development, an otherwise overlooked systemic variable that might have the same effect with policy analytical capacity is inserted into the research design. The aim is to, account for this probable omitting causal relationship, as the DEG might have a positive effect upon big data assimilation, as the policy analytical capacity is assumed to have.

Except for the impact that the dominant paradigm for public sector development might have upon big data assimilation, an additional influencing variable is going to be prescribed; However, the following variable is not going to be included into the research design due to the restricted time frames and writing margins. Nevertheless, the introduction of the additional variable can stand as a suggestion for an another single-case study, as the study's nature demands the existence of multiple variables and competing hypotheses in order to produce inferences about multiple points that are related to the study of the big data use inside the public

sector. Accordingly, the one additional independent variable that might influence big data assimilation, is the presence of potential barriers to innovation.

As it is already established throughout the literature review, public entities, in their attempt to assimilate big data applications, that are considered as an innovative technology for the public sector, might stumble upon certain barriers. The types of the barriers to innovation range from Institutional and Political barriers, to Managerial, Organizational and Technical ones. Barriers to innovation affect the process of big data assimilation in a public-sector environment, as the following assumption suggests. Consider, an imaginary public-sector entity that is equipped with the sufficient policy analytical capacities, is staffed with skillful and competent civil servants and it is fully predisposed towards adopting a new big data application. By now it is assumed that the public organization is armed and ready, in order to make the most out of the available big data applications and subsequently will assimilate the particular technology by embracing it and making it as a part of its standard procedures. However, this sequence might get interrupted when barriers to innovation appear.

In reality it is likely for a public organization to experience multiple barriers to innovation. Some of them are shortage of staff, limited digital expertise, restricted organizational resources, absence of political support or even politicized polemic against the adoption of new innovations. Subsequently, a public organization, which encounters any of these types of barriers, and is unable to overpass them, due to its own institutional complexities, organizational features and possible disabilities, might get unmotivated to complete the innovation adoption process. In other words, barriers to innovation influence the adoption of new technologies by public organizations, as they stand as objective impediments that obstruct the process of adoption, disincentivize the organization towards incorporating the technology and finally do not allow the complete assimilation of the technology. Based on this narrative, it is assumed such barriers might affect the assimilation of big data applications.

According to the above narrative, I propose the following suggested hypothesis: “When a public organization experiences barriers to technological adoption, it is less likely to assimilate big data applications”. The hypothesis suggests that a public organization might not only experience one single barrier, but it might encounter several barriers to adopting an innovative technology simultaneously. When barriers are stacked upon each other, the organization faces a greater burden and then strays further away from assimilating big data technologies. For example, an organization that presents low ICT expertise in its staff, probably will not succeed in realizing

big data assimilation, unless it might overpass this organizational barrier, by recruiting qualified staff or by introducing compulsory training programs for ICT skill development. However, an organization, whose staff presents insufficient ICT expertise, but at the same time, it also experiences negative political pressure and is unable to implement training programs due to resource shortages, it is critically less likely to assimilate big data applications, as it is overpowered by the weight of the barriers. The confrontation of barriers is not an easy task for a public organization. Actually, the existence of barriers to utilizing and adopting a new technology, such as the big data applications, can create a vicious cycle were more implications are generated. Although, the scope of the study cannot be extended to examining the potential impact of possible barriers to big data assimilation, the above section aims on providing the incentive for future academic research.

For the time being, in the next part of the study, that of the research design, the key concepts are going to be defined conceptually and operationally and indicators will be assigned to each of the concepts, that are included in the hypotheses, in order to be identified during the empirical analysis.

3. Research Design

By now it is apparent that big data technology offers the possibility for improving the public organizations' day-to-day operations and general policy outputs. In order to achieve such outcomes, public organizations should have a workforce that is equipped with certain qualities, as well as, be able to overpass various barriers to innovation adoption. The overwhelming characteristics of the big data technology in combination with lack of necessary skills may demotivate governments to assimilate the big data technology in order to eventually capitalize on its value. Conversely, policy analytical capacity is assumed to be an enabler of big data assimilation. In other words, the main assumption of this study is that policy analytical capacity provokes the effective assimilation of big data applications. Correspondingly, the variable of digital literacy can be considered as an alternate possible causal factor to policy analytical capacity, presumably presenting the same causal relationship with the variable of big data assimilation. Overall, it is assumed, that the presence of well-equipped and skillful civil servants within public organizations, can affect the assimilation of the big data applications by the public-sector entities. This assumption leads to the following question: "*How does policy analytical capacity affect big data assimilation in the public sector*?" In addition to answering that question, one more variable that is going to be examined because it is considered to have an effect on big data assimilation, is the dominant theory for public sector development. Except for the impact of individual skills, the effect of a systemic factor is going to be put under the test. Thereby, an additional goal of the study is to deconstruct the potential causal sequence that begins with the Digital-era Governance and ends in big data assimilation.

3.1 Research Goal

The main objective of the research is to recognize the causal links that associate the concept of policy analytical capacity, with that of the assimilation of big data technologies in the public sector. This aim can be accomplished by examining the possible causal strings that can lead to the outcome variable, by evaluating a series of alternative hypotheses that account for the potential intervening factors that may affect the relationship between high analytical capacity and the outcome variable of big data assimilation. Thereby, the methodology of the research acquires an exploratory character, as it aims on identifying potential causal factors that can be tested further in future studies. In other words, by being exploratory, the study examines the contemporary issue of big data use in the public sector, and develops preliminary causal

connections and assumptions, in order to suggest refined research questions for future studies (Neuman, 2013, p. 38). However, the study's purpose is not limited to providing a pure description of the theoretical assumptions for future research, but it also aims on providing an "*enhanced analytical description"* of the suggested explanatory hypotheses (Toshkov, 2016, p. 8). For understanding the connection between the independent and the dependent variables, as well as, exposing the relevant causal mechanisms, the study focuses upon the case of the UK public sector.

Subsequently, the intention of the study is to retrospectively examine the causal mechanisms and identify the potential effect that high analytical skills may have on big data assimilation in the UK public sector. For doing so, possible confounding factors, such as the dominant theory for public sector development and the degree of digital skills that civil servants in the UK possess, are some of the independent variables that are going to be put to the test. All things considered and regarding the direction of causality of the study, the research is grounded on the hypothesis that the independent variable of policy analytical capacity affects the way of how big data applications can be assimilated in the public sector.

3.2 Case selection

As it is already implied in the previous section, the study adopts the design of a within-case research, where multiple observations are made for a broad number of variables, that are drawn from a single case. Respectively, for examining the hypothesized impact that policy analytical capacity might have on big data assimilation, the case of the UK public sector is selected. The case of the UK public sector refers to the set of Government Departments that comprise the Central Government of the United Kingdom. The Government Departments are Ministerial and Non-Ministerial administrative organizations that are responsible for the provision of all the public services.

The UK public sector stands as an appropriate example for inspecting the relationship between analytical capacity and big data assimilation, because, even though the country endorses digital change, is receptive to new technologies and displays an outstanding initiative towards digital governance, yet it experiences significant skill gaps among its civil servants (NAO, 2017). This paradox stands as a puzzle that troubles the theory, given the assumption that higher levels of skills and expertise lead to more effective policy outcomes and higher production of public

value. Thereby, the case of the UK public sector can stand as a deviant case. This is because, by being inside a different context, a country that lacks the expected digital skills is subsequently not expected to exceed in digital governance and technology adoption. By imposing a theoretical anomaly, the UK supposedly overperforms in digital governance and technological innovations, although it is not equipped with the expected competencies that serve as requirements for being a digital-world leader.

Conversely, the literature suggests that the UK exceeds in evidence-based policy making. As Sanderson (2002) argues, evidence-based policy making has gained renewed currency in UK since the previous decades. Specifically, previous governments were committed to modernize the public sector, by making it more joined-up, strategic and responsive, based on the slogan "what matters is what works" (Sanderson, 2002, p. 4). Thereby, it is suggested that evidence-based policymaking is an established practice in the UK public sector. As the study presumes that evidence based policymaking and big data usage are intrinsically associated, the UK stands as the appropriate case for investigating this relationship. Moreover, by selecting this case, the study will deliver evidence which can strengthen or disconfirm the suggested emerging scenarios for the future of the DEG in the UK, as they were introduced by Margetts & Dunleavy (2013). Finally, the case of the UK public sector is also selected for the purpose of convenience, as there are no language barriers in collecting data, as well as, because the UK government has established online portals, where the relevant evidence and data can be found freely.

Concerning the dependent variable of interest and because it is nearly impossible to include the whole set of public organizations that comprise the UK public sector in this single-case study, four respective governmental departments are selected as a manageable subset of the entire UK public sector. In other words, it can be said that these four organizations function as a sample that represents the population of organizations that comprise the whole UK public sector. The four public organizations are the Department of Work and Pensions, the HM Revenue and Customs, the Department of Health and the Department for Environment, Food & Rural Affairs. The executive agencies that are in arm's length to each of the corresponding core departments, cannot be neglected during the phases of data collection and analysis, as most of them are the responsible bodies for conducting the departments operations. This spectrum of organizations that are part of the UK's central government was selected, because it constitutes a diverse set of cases, as every organization presents a different purpose, and presumably a different organizational structure and workforce.

To begin with, the DWP is one of the largest public organizations in the UK; the DWP is overseen by the Secretary of State for Work and Pensions, and the total number of staff that work inside the organization, is 75.150 individuals (DWP, 2017). Additionally, the objective of the organization is to ensure the effective operation of the country's welfare system. Continuing to the second largest organization inside this subset, the HMRC is a non-ministerial department, that is accountable to the Chancellor of the Exchequer, who has delegated the responsibility for oversight to the Financial Secretary to the treasury (HMRC, 2017). Thereby, the HMRC must comply with the directions given by the Treasury ministers. The purpose of the HMRC, is to collect tax revenues and provide targeted financial support (HMRC, 2017). Its annual report for 2016-2017, displays that 62.800 full time equivalent civil servants are working for the core department and the Valuation Office Agency (HMRC, 2017). Next, the third selected organization is the Department of Health. The DH presents nearly 6700 individuals working as full-time civil servants for the core department and Public Health England, which is the department's executive agency, additionally supporting the operations of the NHS (DH, 2017). Concerning, NHS England, it is the accountable arm's length body for carrying out some of the department's responsibilities, related to the health system (DH, 2017). Last, Defra employs more than 5515 individuals in the core department, as well as, inside its executive agencies[2] (Defra, 2017). Defra is responsible for implementing policy related to environmental protection, to food industry and to rural economy.

By forming the sample consisted by the particular public organizations, the aim is to introduce potential variation in the dependent variable of big data assimilation inside the research design. In addition, as each of the departments requires different professional specialisms in order to conduct distinct operations and deliver differentiated public services, it can be assumed that the organizations will also differ in the variable of policy analytical capacity. Last, the different structure, size and purpose of the organization are terms that might introduce additional variation to the sample.

3.3 Case, units of observation and units of analysis

Regarding the unit of analysis and unit of observation, the level at which the data is collected, and measurements are made, is at the same level that the analysis is made (Toshkov, 2016). The constructs of policy analytical capacity and digital literacy apply to individuals who are working

[2] Rural Payments Agency, Animal and Plant Health Agency

within the UK public sector. Respectively the study's unit of analysis and unit of observation is the British civil servant. Accordingly, the independent variable of policy analytical capacity is examined at the individual level. Similarly, the unit of analysis is the same around the context of the alternative hypothesis, related to digital literacy.

However, the construct of the DEG applies to the systemic level of the UK public sector. Accordingly, for the hypothesis on DEG, the study does not draw observations or analyzes data at the level of the individual, but rather from the structural context of the UK public sector. In more detail, regarding the hypothesis that examines the assumed influence that the possible DEG approach has on the dependent variable of interest, the level of observation and the level of analysis coincide, as the unit of analysis is the entire public sector and the proposed inferences and data are also drawn from UK public sector at the aggregate level.

3.4 Data collection

The purpose of the data collection phase is to gather qualitative and quantitative data through the nonreactive research of existing sources. Public documents stand as the source for collecting data related to the study's research question. By adopting a form of document analysis, through the data collection phase of the research, various forms of text are collected, reviewed, interrogated and examined, as the prime source of research data (O' Leary, 2014). Concerning the general type of the examined documents, the study examines internet transmitted public records, which are the product of specific activities or routine administrative procedures of organizations, private or public. Respectively, data is drawn from government documents, public organization annual reports, audit office reports, non-departmental public body reports, existing statistics and indexes of IOs and private company publications.

In more detail, documents are classified in the following categories. Inside the "UK Government Documents" category, there are documents such as, Annual Reports of Governmental Departments, Cabinet Office Government Strategies, Civil Service Profession Competency Frameworks, Government Digital Service publications and reports from the National Audit Office and the Office for National Statistics. The second data source category is the "EU Documents", including reports from the European Commission. Respectively, inside the "International Institutions category, are reports published by the OECD. Last, the "Independent Documents" category, includes reports that are drawn from private companies and think tanks,

containing documents from Deloitte, Sopra Steria, Tech UK, Capgemini UK, SAS, and the Institute for Government. Online press sources are utilized as well.

In sum, depending on the context of the UK public sector, my aim is to distil any useful observations from sources that are pertinent to the research hypotheses. Data collected through the existing documents, will function as evidence, that can support or discredit the hypotheses. I have to note that, the main characteristic of the data collection phase in a single-case study is that the collected observations are not quantifiable and comparable (Toshkov, 2016). This is because, the many aspects of the case and the corresponding data are of different nature, thus they cannot be organized and standardized in data matrixes. However, the relevant data are categorized according to the corresponding hypotheses and variables (*see Table 2)*. Thus, the data collection method can be characterized as qualitative in nature (Toshkov, 2016, p. 295), and the next paragraph explains how it was done operationally.

By following the guidelines of O'Leary (2004) for document collection analysis, I framed a list of documents, that were going to be possibly related to my study's research question. The objective was to search for data that indicate the presence or absence of policy analytical capacity, and digital literacy of the British civil servants at the individual level. The same method applies for searching data that indicate the contextual and structural setting of the UK public sector, for identifying the DEG paradigm. Accordingly, by developing the scheme for categorizing the documents, as it is mentioned above, I searched for related documents that include qualitative or quantitative evidence. The next step was to assess the authenticity of the documents. As most of the evidence are contained in public records of government agencies or renowned private companies, the authenticity and credibility of the text can be ensured. In order to further ensure the validity of the evidence, I gathered and reviewed additional documents, with the aim of triangulating the data, so to establish that similar evidence can be found on different sources. The final action is the step where the data are analyzed, in order to be qualitatively interpreted, so to become usable for the qualitative method of process-tracing.

The advantage of the document collection method is that there is a lack of obtrusiveness and reactivity, thereby the stability of data is ensured as the data source is not affected by the research process (Bowen, 2009). However, this data collection method always entails the danger of biased selectivity, as the collection of documents can never be exhaustively complete (Bowen, 2009). Given the efficiency and cost-effectiveness of this method, as well as, the study's time span, this is the major potential flaw of the data collection method.

3.5 Conceptual definitions and operationalization

By now, it is time to continue to the process of measurement. The measurement's function is to link the abstract concepts and subconstructs with matching observational indicators. Measurement helps the researcher to identify the corresponding concepts in the empirical world. However, this process first begins with the conceptualization phase. Conceptualization can be referred as the process of developing "*clear, rigorous, systematic, conceptual definitions for abstract concepts*" (Neuman, 2013, p. 205). Thereby, under this section, I will try to define all of the concepts, that are parts of the relevant hypotheses and explain their attributes as elaborately as possible. This is an important step for the integrity of the study, concerning both the internal and external validity of the research, as poor conceptualization can lead to false conclusions. Thereby it is important to set coherent definitions, which contain the relevant sets of various attributes of the concept, by clarifying both the intrinsic attributes, that specify the internal properties of the concept, and the extrinsic ones, which indicate how the concepts is interrelated with the other ones (Toshkov, 2016, p. 92). In sum, throughout this section, the relevant concepts will be deconstructed, and their attributes will be clearly specified. The aim of this section is not only to describe what the concept is, but to crystallize its distinguishing attributes and present how it resembles or differs with other semantically related concepts.

After assigning a conceptual definition to a certain concept, I will proceed to the operationalization process, by assigning indicators to each of the concepts respectively. Operationalization refers to the "*translation of abstract concepts and their attributes, into less abstract ones, that can be detected, classified and measured in the empirical world"* (Toshkov, 2016, p. 100). Thereby, during the process of operationalization, I will link the conceptual definition to specific indicators and detectors, that will allow me to identify the relevant concept inside the case of the UK public sector. In this way, concepts will be interpreted as measurable variables. During that process, I will attempt to designate the rules of correspondence, by providing statements that will display, how a construct can be identified and how the respective indicator corresponds to the specific construct (Neuman, 2013, p. 208). The process of operationalization is important, because when the variables are properly operationalized, they reflect the concepts accurately, and they provide valid representations of the empirical world (Toshkov, 2016). In this way, operationalization, allows us to strengthen the validity of the study and avoid any random measurement errors. So far, we discussed the methodology of the

measurement proceeds; it is now time to proceed to providing conceptual and operational definitions to each of the concepts.

Big data applications: The first main concept of the study is the concept of Big data applications, as big data is the general theme that governs the study. As it is already noted in the literature review, Big Data Applications are characterized by having the capability to process, with high velocity, data-sets that are large in *volume* and diverse in *variety* (Giest, 2017; Höchtl et al., 2016; Kim et al., 2014). Thereby, an additional distinguishing attribute of the big data applications is that they offer the possibility of processing data-sets that cannot be processed by conventional ICT infrastructures and standard analytical instruments (Höchtl et al., 2016). Big data applications, therefore can combine heterogeneous forms of data in vast sets, so to include more variables into the data analysis and provide insights, which cannot be obtained with traditional analytic techniques.

When big data applications are employed specifically inside the public sector, their definitive attribute is that they serve the purpose of increasing the public value provided by a public service, or they aim at improving the operations of public organizations. Subsequently, big data applications are concerned with data which are created by citizens in their everyday habits, data that are automatically generated by digital sensors, and data that are the byproduct between the interactions of public entities (Mergel et al., 2016). A semantically relevant term to big data applications is that of big data analytics. However, big data analytics refers to the *process* of utilizing large and varied data sets, that is driven by the specialized big data software applications. Concerning the previous attributes, big data applications that are used inside the public sector can be defined as software applications, which have the ability to analyze large and diverse data-sets, which are obtained by both continuous and automatic collection process in real time, with the aim of producing public value.

Because big data applications can be utilized in different levels of government, ranging from local and central government, to public bodies, while they can also be found in diverse areas of public service, ranging from healthcare, education and security to fiscal and social policy, it can be assumed that the concept refers to a large extension of cases. Additionally, the set of cases is further extended, if we consider that big data applications can be detected globally, inside the public sectors of a large number of countries.

In order to identify the presence of a big data application in real life settings, the operational definition suggests that one should look for an innovative software application that aims at producing public value or making the public organizations' operations more efficient and effective, by analyzing large and diverse data-sets. In other words, some of the variables for identifying big data applications are, their aim to produce public value, aim to improve public service operations and by the ability to analyze large and diverse data. However, the type, format and function of the big data applications can also serve as indicators for identifying and categorizing the concept.

Data applications can be detected as having the function of, *data warehouses,* which are vast, structured data stores that save information for several uses, depending on the respective policy area. Additionally, one more big data application type that can stand as an indicator to identifying it, is the *data lake* format, that is similar to the data warehouse type, but it stores unstructured raw data from a wide range of internal and external sources to the organization. *Data mining* is also considered as a big data application software and aims at recognizing patterns and insights in data stores. One more type is that of *predictive analytics,* which serves the purpose of forecasting likely future events. Last, a big data application can be identified as *a prescriptive analytics* application, that suggest optimal courses of action, aiding decision-making processes (Harvey, 2018).

Big data assimilation: After having provided the relevant set of indicators for identifying a software system as a big data application, it is now time to describe the corresponding conceptual and operational definitions for the dependent variable of interest of the study, that of big data assimilation. The defining attribute innovative technology assimilation, is that it requires from the organizational staff to be aware of the innovation and from the organization to fully accept, utilize and institutionalize the innovation (Meyer & Goes, 1988). Additionally, assimilation is defined in terms of six stages (Fichman & Keremer, 1997), where each of them can considered as a necessary condition for assimilation to occur. An organization initially becomes aware of the technology, and then it indicates its interest towards adopting the technology. The next phase is that of trial and evaluation, where if the results are positive, then the organization commits in adopting the technology. Last, the organization establishes the technology as a part of its procedures, but its deployment can be either limited or general (Fichman & Keremer, 1997). Respectively, big data assimilation can be defined as the eventual outcome that a public organization experiences, when it has selected, accepted, implemented

and totally incorporated a big data application, as a part of its standard operating procedures. The term of big data assimilation should not be confused with the relevant term of big data adoption, which refers to process of accepting and implementing a big data application, but without diffusing it across the organizational structure, and without routinizing it as an available mean that serves the organizational purpose.

Accordingly, big data assimilation in the empirical world can be observed by following the sets of conditions of the six-stage model. *Awareness, interest, evaluation, commitment, and limited and general deployment,* are some of the detectors that function as conditions for understanding that the assimilation might be taking place. For the purpose of the study, the stages of awareness, interest and general deployment are considered as sufficient conditions for the process of big data assimilation to occur. Respectively, the main indicator of the variable, is the general deployment of the big data application, meaning that the application is available and has been used inside the public organization. As expected, subsequent indicators, are the *absence of rejection, stalling or obstruction* towards the use big data technology. Lastly, it can be assumed that *improved organizational performance*, *enhanced public service delivery* and *added public value* are some of the additional indicators, which imply that big data assimilation has taken place.

Policy analytical capacity: Having established the guidelines for measuring the presence of big data applications and assimilation, it is now time to assign conceptual and operational definitions to the independent variables. The main independent variable of the study is that of policy analytical capacity. By policy analytical capacity, we refer to "the amount *of basic research a government can conduct or access, and its ability to apply statistical methods, applied research methods, and advanced modelling and statistical techniques, such as trend analysis, forecasting methods and scanning in order to gauge broad public opinions and attitudes [..] and to anticipate future policy impacts"* (Howlett, 2009). However, policy analytical capacity can be perceived differentially, depending on the level of analysis that is examined. At the individual level, policy analytical capacity refers to the ability of individuals within the public to analyze problems by using advanced statistical techniques, as well as, their ability to apply policies and contribute to the overall design and evaluation of the policies themselves. However, policy analytical capacity at the organizational level, refers to the ability of the public institutions of ensuring the presence of policy analytical capacity by providing the enabling context for the civil servants in order to perform the above functions that are related to

policy analysis (Hsu, 2015, p. 199). Lastly, at the systemic level, the availability and access to high quality information and research that higher-learning institutions provide to the government, is considered as the systemic requirement to a government's policy analytical capacity (Hsu, 2015).

Thereby, the concept of policy analytical capacity can be applicable to every level of analysis that is relevant to the research; from the individual level of the civil servant, to that of the public organizations and bodies, to the systemic level that examines attributes of the UK government. Accordingly, the variable of policy analytical capacity can be observed within a public sector, by detecting that the civil servants are capable of utilizing advanced computational and statistical techniques, while their function is supported by the institutional context of their organization and the general research which the government facilitates. Respectively, some of the indicators of individual policy analytical capacity are, *knowledge of policy substance*, *analytical techniques* and *skills at the individual level* (Howlett, 2015, p. 178). More specifically, individuals can be considered to have policy analytical capacity, if they present the *ability to utilize statistics, trend analysis, forecasting methods, theoretical research, and to undertake evaluation of the means of meeting goals* (Howlett, 2015, p. 178). At the organizational level, policy analytical capacity can be detected by the existing, *institutional context and opportunity structures* that the public sector offers. Last, policy analytical capacity is demonstrative at the systemic level, if variables such as, *high quality of information, penetration of information and communication technologies, and ready availability of data,* are found within the systemic milieu of the government.

However, the study aims at observing the quality of the policy analytical capacity only at the individual level. Respectively, for the purpose of transforming the variable to a more detectable and tangible concept, a proxy variable is introduced into the research design. A proxy is a variable that is used in order to measure an unobservable variable of interest (Lewis-Beck et al., 2004). By observing a proxy variable, the researcher is able to obtain a valid measure, as the proxy is strongly correlated to the unobservable variable of interest. In the case of this study, the proxy used for locating the existence of policy analytical capacity within the public-sector workforce, is that of the civil servant's recognized *profession*. The civil service in the UK is categorized on the basis of 28 professions, and each profession has developed its own competency framework. Respectively, professions that require competencies that are pertinent to possessing policy knowledge, analytical and statistical skills theoretical and applied research

proficiency, can be used as a measure for detecting the variable of policy analytical capacity. Accordingly, public organizations, which are staffed by such professionals, are subsequently assumed having policy analytical capacity within their premises.

Digital literacy skills: The independent variable of digital literacy is used as an alternative variable that can condition for the effects of policy analytical capacity in the research design. Digital literacy is an umbrella concept, as various skill clusters are placed under it and many relevant terms can be perceived as synonyms to each other. Digital literacy refers to the individual's ability to perform tasks operationally inside a digital environment, while he/she possesses a digital-friendly mentality and deep understanding of the digital technology that is driven by the ownership of relevant technical and computing skills.

Under this definition, the relevant sets of skills can be utilized as indicators to the study. Thereby, the indicators that can imply that an individual possesses digital literacy skills, are *digital capability, digital expertise and knowledge of technology*.

Digital era governance: Continuing to the conceptualization and later the operationalization of the independent variable of Digital-Era governance, the DEG can be defined as an approach for public sector development, that aims at re-establishing the structure and function of the government, by using digital technology as a driver for innovative and competitive government (Margetts and Dunleavy, 2013). The DEG approach utilizes digital technologies in order to accomplish some of its defining attributes, that are, agile, institutionally simple, administratively simplified, automated, citizen-centered and social-problem solving government (Tassabehji et al., 2016, p. 225).

Some of the key components of Digital-Era governance that can function as indicators for identifying the presence of this specific approach of public sector development, are *features of Joined-up Governance, Procurement concertation and specialization, One-stop public-service provision, Data warehousing usage, Agile government processes, Electronic Service delivery, and Open-Book Government* (Dunleavy et al., 2006, p. 481). These indicators are placed under the themes of reintegration, needs-based holism and digitization processes that govern the DEG approach.

For the table including each of the concepts and the corresponding indicators, see Table 2.

3.6 Research Approach and Design

The selected research approach is a single-case study that examines various evidence drawn from the UK public sector. In the simplest terms, a single-case-study research examines many features of a single case (Neuman, 2013). Similarly, all case-study research is initiated by the desire to derive in depth comprehension of a single case that is set in the empirical world (Yin, 2012). Accordingly, the case of the UK public sector will be examined intensively, as the above described variables, that are part of competing hypotheses will be put to the test. Thereby, I preferred the single-case format, because it is appropriate for holistically explaining a specific outcome of an individual case (Toshkov, 2016). Additionally, the single-case format is selected as it is appropriate for supporting the study's objectives. According to Yin (2012), single-case studies are suitable for both exploratory and explanatory purposes (p.5). Thereby, the single-case format of this study is deemed apt for developing questions and hypotheses for subsequent research, as the research purpose suggests, while it is also appropriate for explaining how an effect can lead to a certain outcome. Accordingly, the single-case format of the study also allows the application of the process-tracing method in the research, as this method aims on reconstructing causality within a single case (Gerring, 2006).

In this way, this research design allows me to follow an abductive reasoning, where various hypotheses are tested consecutively until I can find the one that cannot be rejected (Toshkov, 2016). The study's research design helps in exposing the causal mechanisms that stand behind big data assimilation in the case of the UK public sector; subsequently, this specific research design allows to distinguish the competing hypotheses, that can be met with approval, from those that will be rejected. In other words, the research approach draws hypotheses from the relevant theory, and then evaluates the competing hypotheses in order to make inferences, by uncovering the related causal mechanisms, that lead to the outcome of interest. Eventually, the hypotheses that are validated by being inferentially strong, can contribute back to the literature by being foundation for future academic research.

In sum, the distinguishing characteristics of the case-study are the following. First, the single-case study ensures the existence of conceptual validity within the study. Case studies are deemed suitable for elucidating the concepts/variables that are of greatest interest and proceed to specifying their essential meaning from abstract theory (Neuman, 2013, p. 42). Second, the heuristic format of the study is pertinent for exploring the boundaries between the related

problems and identifying the causal mechanisms in detail (Neuman, 2013). Last, the single-case format is the most appropriate design for supporting the use of process tracing as a method of analysis, because of its ability to capture the complexity and trace the relevant process of interest over time and space (Neuman, 2013).

3.7 Method of Analysis

In order to evaluate the evidence that will surface from the phase of data collection, the technique of process tracing is adopted as the applicable method of analysis. Bennet and Checkel define process tracing as the technique of "*using evidence from within a case in order to make inferences about causal explanation of that case*" (Toshkov, 2016, p. 298). In more elaborate definition, Collier (2011) suggests that process tracing is a fundamental tool of qualitative analysis, which can be referred as the "*systematic examination of diagnostic evidence selected and analyzed in light of research questions and hypotheses posed by the investigator"* (p. 823). Process tracing can be considered as an appropriate method for this study, as its function aligns with the research's aim, namely uncovering the causal mechanisms behind big data assimilation. This is because, process tracing follows a generative perspective of causality, that aims at providing a detailed description of causal mechanisms that lead to a certain effect (Punton & Welle, 2015). In other words, process tracing demonstrates a causal relationship by testing the relevant evidence and then offering a meticulous explanation of how a cause can lead to a specific outcome.

When applying process tracing, an inference is not the result of a one-shot observation. Instead, multiple pieces of heterogeneous evidence, that are drawn from different units of analysis are utilized for the verification of a single inference (Gerring, 2006). In this fashion, multiple bits of evidence, are used to display a long causal chain, where X1 → X2 → X3 → Y, instead of demonstrating that an independent variable can lead directly to the outcome of interest (Gerring, 2006). By applying process tracing in this case, I will attempt to assess the heterogeneous pieces of evidence against alternative hypotheses, with the aim of disentangling the causal mechanisms and establishing a causal chain of discrete steps that relates the independent variables to the outcome of big data assimilation.

Regarding the operational use of this method, the relevant pieces of evidence will be evaluated under the light of probability tests that assess the strength of the specified causal relationships

(Punton & Welle, 2015), in order to point the likelihood under which certain evidence can lead to the identified outcome. The probability tests, used in process tracing, organize the evidence along the dimensions of certainty and uniqueness; where the main purpose, is to assess the necessary and sufficient conditions for inferring that evidence. According to Toshkov (2016), for a condition to be deemed *"causally necessary for an outcome, the outcome must not occur in the absence of the condition"*. Respectively, a condition is "*causally sufficient for an outcome, when the outcome always occurs in the presence of the condition"* (Toshkov, 2016, p. 270). But how, process tracing can be applied functionally?

On the basis of these dimensions, four probability tests were developed in order to allow the investigator to examine the available evidence and evaluate the evidence's inferential significance for proving or rejecting the causal hypotheses. Namely, the tests are the "*Straw-in the Wind*" test, the "*Hoop*" test, the "*Smoking-Gun*" test and the "*Doubly decisive*" test (Bennet, Collier, Van Evera, in Toshkov, 2016). More specifically, tests with high uniqueness, as the Smoking Gun and the Doubly Decisive test, are employed to strengthen the confirmatory evidence for a hypothesis, by proving that a certain piece of evidence was sufficient to validate it (Punton & Welle, 2015). Respectively, tests that with high certainty, just like the Hoop and again the Doubly Decisive test, are utilized to disconfirm alternative explanations by presenting that a piece of evidence is necessary for the hypothesis to hold (Punton & Welle, 2015). These arguments leave out the Straw in the Wind test, which handles evidence that are both low on uniqueness and certainty, meaning that it presents neither necessary nor sufficient conditions to confirm the hypothesis. In other words, if evidence is found and can pass the hypothesis, the confidence of the investigator that the hypothesis is true can be merely increased, yet the evidence is not sufficient to prove the hypothesis conclusively. On the other hand, if the available evidence does not pass the test, the confidence in the hypothesis is decreased, but the evidence is not high in certainty in order to rule out the hypothesis once and for all.

Concerning the application of the process tracing method in this case, the following sequence will take place through the analysis. After having collected relevant evidence through the data collection process and taking into account that the method of process tracing is compatible with identifying evidence that result from different levels of analysis and collection methods (Toshkov, 2016, p. 299), certain hypothesized causal assumptions for each of the evidence will be constructed. Under each of the probability tests, a corresponding hypothesis will imply the existence of a causal mechanism, that describes how a hypothesized causal chain leads to the

effect of big data assimilation. Respectively, an alternative hypothesis that implies otherwise will be established. At first, the available evidence will be reviewed under the main hypothesis of the probability tests, with the aim to prove that the causal sequence is true in the theorized way. Later, the same evidence will be tested under the alternative hypothesis. The probability tests' hypotheses will resemble the three main hypotheses of the research design, as they will function as sub-assumptions that describe how each of the independent variables, such as policy analytical capacity, digital literacy and the DEG, can lead or not to the outcome of interest. In the end, hypotheses will be confirmed depending on the availability and the inferential weight of the evidence that are found. Explanations and theorized causal mechanisms that are incompatible under the light of the evidence will be discarded. By using the insights offered after testing the hypothesis, corresponding causal models are developed, describing the sequence of the accepted causal mechanisms.

Table 2: Combined table of conceptualized and operationalized variables

Dependent variable	Definition (Literature Review)	Indicators	Data source
Big data assimilation	The eventual outcome that a public organization experiences, when it has selected, accepted, implemented and totally incorporated a big data application, as a part of its standard operating procedures.	General Deployment of a big data application Absence of rejection *stalling or obstruction* for employing a big data application Improved organizational performance Added Public Value Enhanced	**Department of Health/NHS England:** *The NHS is beginning to use […] predictive models of analysis by using historical data to effectively and reliably predict the future risk of complications*; (Capgemini UK, 2017). **HMRC:** The HMRC uses *"data mining tools and methods to tackle the hidden economy in the UK".* "The HMRC employs *"analytic tools (Connect), analytical methods (such as dynamic benchmarking and predictive analytics)"* with the aim to reduce tax gap, improving tax compliance and reducing evasion *[…] "Recent data published by HMRC shows a long-term reduction in the tax gap from 8.3% in 2005-06 to the latest figure of 6.5% in 2014-15".* (European Commission, 2017). *"HM Revenue & Customs (HMRC) is driving the use of open source technology with a Hadoop NoSQL big data engine to analyze corporate tax".* (Saran, 2014) *HMRC has replaced conventional debt processing systems with an innovative IT solution capable of mass customizing debt collection interventions based on insights into customer behavior.* (Capgemini,

		public service delivery	2014). *"Personal tax accounts from HMRC now take a real-time digital approach" (Manzoni, 2017).* **DWP: "***The DWP's Strategy Directorate used SAS to develop Genesis, a generic architecture for implementing dynamic micro-simulation models" (SAS, 2014). "The DWP also makes extensive use of SAS predictive analytics for regression analysis, to evaluate historical data and develop models to predict the likelihood of a particular event happening to an individual based on their characteristics at a particular point in time". (SAS, 2014).* *"In DWP, for example, providing job seekers with more targeted advice, and opportunities that closely match their personal profiles. The department is also working on data-informed tools, such as interactive visualizations of benefit claimant trends". (Manzoni, 2017).* **DEFRA: "***The Common Agricultural Policy (CAP) policy team and the Rural Payments Agency (RPA) are already exploring how using Earth Observation data will enable us to reduce the CAP disallowance and the number of farm inspection visits"* (Coley & Newman, 2016).

Main Hypothesis: ***The individual policy analytical capacity, that civil servants inside the UK public sector have, is an enabler of big data assimilation***				
Independent Variable	**Definition (Literature Review)**	**Indicators**	**Proxy Variable**	**Data source**
Policy Analytical capacity	"*the amount of basic research a government can conduct or access, and its ability to apply statistical methods, applied research methods, and advanced modelling and statistical techniques, such as trend analysis, forecasting methods and scanning in order to gauge broad public opinions and attitudes [..] and to anticipate future policy impacts*" (Howlett, 2009).	Knowledge of policy substance Analytical techniques and skills Ability to utilize statistics Ability to utilize trend analysis/forecasting methods Ability to conduct theoretical research Ability to undertake evaluation	Profession requiring the corresponding policy analytical skills	National Audit Office [NAO], 2017; Office for National Statistics [ONS], 2016; Institute for Government [IFG], 2017; Government Economic Service [GES], 2013; Government Knowledge & Information Management [GKIM], 2009; Digital-Data and Technology [DDaT], 2017; Government Operational Research Service [GORS], 2014; Government Social Research [GSR], n.d.; Government Statistical Service [GSS], n.d.; Intelligence Analysis [IA], n.d.; Policy Profession [PP], 2018;

Second hypothesis: *the digital literacy that civil servants in the UK public-sector display, is an enabler for the enabler of big data assimilation*.			
Independent Variable	**Definition (Literature Review)**	**Indicators**	**Data source**
Digital Literacy	Digital literacy refers to the individual's ability to perform tasks operationally inside a digital environment, while he/she possesses a digital-friendly mentality and deep understanding of the digital technology that is driven by the ownership of relevant technical and computing skills.	Knowledge of technology Digital Expertise Digital capability	*"The 43% of civil servants told us that they […] had not received enough digital training to do their job well" (Sopra Steria, 2017,).* *"Skills are consistently identified by civil servants as a significant (often the most significant) barrier to digital transformation". (Sopra Steria, 2017).* Although the 44% of the civil servant's state that they are excited about the advent of data analytics in the public sector, perceiving innovative technology as an enabler for public sector development, nearly 1 out of 4 Civil servants admit that skill shortages are a barrier to transforming the public services *(TechUK, 2017).* 24% Disagree or strongly disagree that their organization has the digital skills it needs in Data management and analytics. *(TechUK, 2017).* *"93% find workforce and skills to be a challenging area to manage in their organization's transition to digital" (Deloitte Digital, 2015).*

Third Hypothesis: *Digital-Era Governance has been established inside the UK public sector and functions as a driver for big data assimilation.*			
Independent Variable	**Definition (Literature Review)**	**Indicators**	**Data source**
Digital-era Governance	The approach for public sector development, that aims at re-establishing the structure and function of the government, by using digital technology as a driver for innovative and competitive government (Margetts and Dunleavy, 2013); it utilizes digital technologies in order to accomplish some of its defining attributes, that are, agile, institutionally simple, administratively simplified, automated, citizen-centered and social-problem solving government (Tassabehji et	Features of Joined-up Governance Organizational reintegration Procurement concertation and specialization One-stop public-service provision Agile government processes Electronic service delivery Open-data	*"Services driven by open data are already giving people more choice in where they get their healthcare, where they live and where their children go to school. […] By doing so, we can actually change the way citizens interact with us - making the relationship we have with them more transparent, more responsive, and based on increasing levels of trust"*. (Manzoni, 2017). *"According to the OECD OURdata index, the United Kingdom has one the broadest standards on open government data among OECD countries, with the fourth-highest score in data availability, accessibility and support for private sector re-use of government data"*. (OECD, 2017). *"Bringing the tools, techniques and culture of the internet to public procurement has already begun. GDS operates the Digital Marketplace, which has embedded user-centered, design-led, data-driven and open approaches across digital, data and technology procurement and contracting. Our priority for 2020 is to embed these more widely across the whole marketplace for public sector*

	al., 2016, p. 225).	government	*procurement"* (Cabinet Office & GDS, 2017). *"Departments are also adopting new ways of working with the private sector, and others, to increase the effectiveness of their spend. We have reported on the challenges and some of the early failures in adopting these, including: adopting more of a commissioning and whole-systems approach, which involves more collaborative working between organizations, using more complex contracting methods such as payment by results, contracts for difference, and management insertion contracts, agile methods and more in-house delivery;"* (NAO, 2017, p. 15).

4. Data and Results

This section includes a detailed description of the research findings pertinent to the study's hypotheses. The presentation of the results begins with evidence indicating the presence of big data assimilation in the UK public sector. Then, evidence suggesting the presence or absence of the study's independent variables are displayed respectively. Along these lines, by identifying the independent and dependent variables of the study, the following results were found. However, before proceeding to the results, here are some relevant information about the country and its public sector

The UK was home to 65.6 million people in 2017 (Deloitte, 2017). In the same year, the UK experienced economic growth of 1.8%, while long term unemployment remained low (European Commission, 2018). Regarding governance, the UK's form of government is the unitary parliamentary constitutional monarchy. At the administrative level, the central government administration is referred as the "Whitehall". In more detail, Whitehall governs on the right of some of the following conventions. The first one, is about the PM's power to hire and fire ministers; this rule emerges from the royal prerogative of the monarch and it is also a formal power of the premiership. Additionally, the ministers' accountability to the parliament, and the civil service's impartiality and constrains on ministerial control, are the rest of the administrative conventions (IFG, 2018). Respectively, as the line of accountability reaches the electorate, public organizations in the UK are accountable to all the citizens, from which 45.7 million individuals have the right to vote (Deloitte, 2017). Moreover, the UK public sector employs the 17 per cent of the country's total workforce (Deloitte, 2017). Finally, concerning the main administrative structure, the UK public sector is comprised by 25 ministerial governmental departments and 20 non-ministerial ones, while it also includes a wide array of agencies and other public bodies.

After providing a general overview of the UK, the next sections will present the results and findings relevant to the study's research question. However, it is essential to notice, that there is a different granularity of data for each of the different variables. Concerning the dependent variable of big data assimilation, the collected data apply to the four specific governmental departments that were selected during the research design phase. The same condition applies for the independent variable of policy analytical capacity. It is important to note that, even though data are indicating the presence of policy analytical capacity UK-wide, statistics, showing

the profession composition by governmental department, offer the opportunity of concentrating findings related to each of the four respective public organizations. However, results and findings on digital literacy and DEG refer to the whole UK public sector, without displaying any distinctions by governmental department.

4.1 Examples of big data application usage in the UK public sector

Initially, before proceeding to testing the hypotheses with the aim of finding out what are the causal mechanisms that can actually lead to big data assimilation, the question that arises is "whether big data applications are actually being employed inside the UK public sector". Throughout this part, examples of big data applications that are being used by the selected British public organizations are going to be reported. These findings serve as the starting point for analyzing the capacities and skills that the public organizations' workforce has, and how these variables are interrelated with the organizations' ability of employing big data applications to the fullest. By using the indicators for identifying a big data application, as they were introduced previously in the research design, the following applications were identified inside the UK public sector.

DH/NHS England: The first organization that is recognized to be using big data analytics technology is the National Health Service. NHS is an executive non-departmental public body, that is under arm's length with the government and it is financed by the Department of Health with the aim of improving health and ensuring high quality care through sustainability (NHS England, 2017). NHS is using predictive data analytics, by combining historical data, in order to prognosticate the future risk of chronic disease complications (Capgemini UK, 2017). The purpose of this big data application allows the NHS to improve its healthcare services and minimize any potential costs that result from further appointments. Additionally, the NHS Business Services Authority, has employed a data warehousing, computing and cloud system, with the name Oracle Exadata, since 2015 (Post, 2018). This application allows the NHS to aggregate billions of data from different sources from all the country and across the organization. Due to its extensive computing capabilities, this database machine combines and analyzes information from prescriptions, medical records and doctor interactions in order to uncover new efficiencies and help the NHS to improve its service delivery while being cost-

effective (Post, 2018). Nevertheless, big data applications are not being used only inside the health services sector.

HMRC: The HMRC is another public organization that employs big data applications, but for another reason; that is to curb tax avoidance. The HM Revenue & Customs is a non-ministerial department, that is the UK's tax, payments and customs authority (HMRC, 2017). Since 2010, the HMRC uses Connect, as one of its main analytical tools for data mining in order to detect hidden economic activities (European Commission, 2017) Additionally, in 2014 the HMRC, purchased the Analytics for Debtor Profiling and Targeting (ADEPT) closed loop big data analytics system, in order to deal with potential tax evasion and fraud (Capgemini, 2014). ADEPT has the capability of automatically combining and processing debt information with integrated socio-demographic data, in order to create valuable intelligence that will be used for debt management. As the conventional debt processing systems are a thing of the past, additionally the HMRC initiated "*the use of open source technology with Hadoop NoSQL big data engine to analyze corporate tax"* (Saran, 2014). Specifically, the HMRC, uses Hadoop, which is an open source software for storage and processing that can handle different types of data, in order to increase its operational efficiency and gain insights that otherwise would have remained hidden. To this effect, Hadoop is spreading throughout the UK public sector since 2015 (Mathieson, 2016). Nevertheless, the usage of big data analytics is not a privilege of only the two last governmental departments.

DWP: The spectrum of public organizations in the UK using big data applications also extends to the welfare and pension policy area. The Department for Work and Pensions is the UK's largest public service department and is responsible for welfare, pensions and child maintenance (DWP, 2017). Its purpose is to provide services and welfare to people who are seeking employment, people with disabilities or health conditions, pensioners and children. For improving its services, the DWP's Strategy Directorate in collaboration with SAS, developed the Genesis application (SAS, 2014.). This big data system allows the organization to gain insights from dynamic micro-simulation models that result from large spreadsheet processing. By using this system, the DWP gains access to long-term simulations of pensioner income distributions and forecasting benefit expenditure (SAS, 2014.). In other words, SAS Analytics, has provided the DWP with predictive modelling and forecasting applications that are capable of doing large regression analyses and creating "what if" scenarios, in order to improve decision-making by making highly complex problems clearer (SAS, 2014.). It is important to note that SAS, claims that the application of this big data platform requires no specialized analytical skills, as the ones

that the SAS is equipped with. Thereby, without being dependent on external partners, the DWP makes use of its large data repositories in order to improve its public service delivery.

Defra: The Department for Environment, Food & Rural Affairs, uses big data analytics likewise, yet it displays a different mission from the previous organizations. Defra is the government department that is responsible for protecting the natural environment, supporting the food and farm industry and sustaining the rural economy (Defra, 2017). Regarding big data application usage, Defra's Common Agricultural Policy (CAP) team and the Rural Payments Agency (RPA), are using Earth Observation data sets, that are coming from satellite imaging in order to curb CAP disallowance and the number of farm inspection visits (Coley & Newman, 2016), making the agencies more efficient. As, Satellite imaging generates immensely large and complex raw data sets, Defra in collaboration with the Earth Observations Centre of Excellence, is trying to transform the satellite-generated data into data that is analysis-ready, by implementing automated processing systems (Dowdeswell & Boyd, 2017). After having the data ready to be exploited, Defra along with the EOCoE are establishing a cloud-based online platform, that allows the central processing of data, without having to download the large-data files, and supports web-mapping services, that provide insights to Defra's task force and field teams (Dowdeswell & Boyd, 2017). Except for using web-mapping services, Defra's Animal and Plant Health Agency, in collaboration with a start-up called Mastodon C, is using big data analytics, and especially new open source technology, to refine its large amounts of unstructured data through text-mining and data visualization techniques, so to eventually gain access to new insights of their data repositories (Open source data analysis, n.d.). However, this was only a prototype project, that was not used in a large scale across the organizational structures, but only demonstrated the high potential that large-data sets have to offer for using innovative analytic techniques in the future.

By converging the evidence, the prior cases can be characterized as instances of big data application usage, because the applications being employed, present capabilities in processing large and diverse data-sets, display data warehousing, data-mining or predictive and perspective function, and present potentiality in improving organizations operations and producing public value in the public affairs environment. As it is already implicit, big data applications are indeed being employed by British public organizations

4.2 Research findings on big data assimilation

Regarding evidence relevant to the outcome of interest, it is evident that the four selected public organizations have indeed assimilated big data applications. Beginning with the NHS, evidence suggest that since the 2015, that is when the NHS consolidated its strategy for using big data analytics, the organization saved 581 million pounds, by supporting its operations on information backed up by the Oracle Exadata Database machine, that allowed the public organization to adjust the costs and effectiveness of the healthcare treatments that it delivers (Post, 2018).

In a same manner, in 2014, the HMRC entrenched the ADEPT predictive analytics platform, that became a part of the debt administration process. Big data assimilation is suggested by the fact that, the public organization's previous systems were *"integrated with "Adept",* utilizing predictive modelling techniques to inform a comprehensive risk and behavior-driven collection tactics (Morabito, 2015, p. 131). Thereby, the Adept system was diffused across the organizational structure and became a part of the standard operating procedures. Additionally, in 2014 the HMRC estimated that until the March of 2015, an extra 3 billion pounds of debt was gathered, as the overall debt collection capacity of the organization was increased; a condition that is highly likely to have occurred because of Adept systems assimilation (Morabito, 2015, p. 131). Additionally, evidence suggest that the comparable predictive analytics application "Connect", that is one of the HMRC's main analytical tools since 2010, has also contributed to collecting the additional 3 billion pounds in tax revenues, as it made the organization able to reduce the tax gap by 1.8% in 2014-15 (European Commission, 2017; HMRC, 2014, p. 16). Later in 2015, the HMRC spend something more than 7 million pounds for collaborating with Cloudera in order to design and implement a Hadoop capability, as part of the HMRC's Enterprise and Data Hub Project (HMRC, 2015). In detail, the 2015 project had the aim of replacing and augmenting the existing Data Warehouse functionality of the organization, to establish a new data hub "*that will support the HMRC Data analytics and Digital ambitions"* to improve yield and productivity across the organization (HMRC, 2015). Evidence suggests that by using the computational capabilities that big data analytics have to offer, the HMRC is moving towards establishing a single repository of customer data, in order to personalize its service delivery and enhance its compliance work (Mathieson, 2016). Thereby, by observing the above data, it is apparent that the HMRC is a public organization that invests and relies heavily

in its data capabilities, as it managed to routinize diverse big data analytics applications as a part of its standard operating procedures.

Respectively, the DWP managed as well to incorporate big data applications within its organizational structures. By relying on the Genesis data engine, the DWP gained insights from reliable forecasts on pensioner incomes and shortfalls (SAS, 2012, p. 4). On the basis of these forecasts, the organization began planning its benefit expenditure correspondingly. Additionally, the DWP is using data mining applications for identifying new types of fraud (SAS, 2012). The SAS report (2012) also suggests that, due to big data analytics, the department managed to identify over than 200 million pounds in overpayments, making the organization more effective and cost-efficient. The incorporation of these innovative technologies has an impact on both the intra-organizational operations and public service delivery. For example, M. Prakash, the Director General for Digital, Data, Technology and Security of the DWP, stated, in the first months of 2017 *"over three million pensions forecasts viewed online by people all over the UK* (Sumner, 2017). Thereby, due its beneficial potential, this data analytics application, supported the DWP's operating procedures, and in turn allowed to an immense number of citizens to get access to improved public services.

Regarding big data technology usage in Defra, by analyzing the data that is available from Earth Observation, the organization has access to dynamic risk monitoring of rural areas, making the organization able of detecting landscape change. In this way, Defra uses the potential of big data application, in order to monitor environmental change and later take action for protecting the environment, while making its teams more informed, operationally capable and cost-efficient. For example, large raw data gathered from earth observation, is being used operationally to streamline the Rural Payment Agency's land change detection with the automated monitoring of greening requirements under the CAP, reducing the number of field inspections (Dowdeswell & Boyd, 2017). Additionally, Defra's Rural Payments Agency managed to produce a crop map for England from radar data (Dowdeswell & Boyd, 2017), and as this service has the potential to be applied to a broad range of policy operations, it can subsequently be diffused easily across the organization's structures and agencies.

Summing up, evidence suggest that all of the four public organizations in the UK public sector have assimilated at least one big data application. To be more specific, regarding the NHS, the employment of predictive analytics and the establishment of the Oracle Exadata big data analytics machine, clearly rendered benefits for the executive body, as it yielded a large number

of saved pounds. Thereby, by institutionalizing a big data application the organization managed to both improve its healthcare service and cut down its operational costs. Similarly, evidence presents that the HMRC managed to assimilate big data applications, as it routinely employs the Connect system since 2010, uses Hadoop, and lately, it integrated the Adept system within its standard organizational systems. This observation can be also supported by the fact, that the HMRC managed to collect additional tax revenues in billions of pounds, achieving its organizational aim of reducing the tax-gap. Respectively, by employing the Genesis application, the DWP managed to identify overpayments in order to cut down excessive costs. However, forecasting data analytics did not only contributed to the organization's frugality, but it decisively improved public service delivery, as citizens gained access to three million pension forecasts. Lastly, regarding Defra, it is apparent that the organization has clearly made use of big data applications to gather and analyze satellite data, but it is still undecisive whether it employs these applications limitedly or holistically within the organization. However, it must be noted that the organization has managed to improve its services and save resources by establishing the use of web-mapping services, that are a product of big data collection.

4.3 Research findings on policy analytical capacity

Having already illustrated the observed instances of big data application usage and assimilation inside the UK public sector, it is now time to display the evidence found, suggesting the presence of policy analytical capacity at the individual level in the civil service. As it is already prescribed in the research design, the measure for providing empirical evidence on policy analytical capacity at the individual level, is the civil servants' recognized *profession*. Accordingly, from the 28 currently recognized professions inside the UK civil service (Civil Service, n.d.), 9 of them correspond to posts that require the existence of policy analytical skills (*see Table 3)*. More specifically, civil servants that are occupying posts in the *Government Economic Service, Government Knowledge & Information Management, Digital-Data and Technology*, *Government Operational Research Service, Government Social Research, Government Science and Engineering, Government Statistical Service, Intelligence Analysis* and *Policy profession,* are assumed as being qualified with substantial policy knowledge, analytical and statistical skills and abilities for undertaking complex policy and research tasks. The next section presents the competencies and operational context of the corresponding professions in more detail.

Table 3: Policy Analytical capacity professions in the UK civil service

1. Government Economic Service (GES)
2. Government Knowledge & Information Management (GKIM)
3. Digital-Data and Technology (DDaT)
4. Government Operational Research Service (GORS)
5. Government Social Research (GSR)
6. Government Science and Engineering (GSE)
7. Government Statistical Service (GSS)
8. Intelligence Analysis (IA)
9. Policy Profession (PP)

To begin with, by inspecting the competency framework of each of the 9 civil service professions, it can be observed that there is an analogy between their competency frameworks and the prescribed skills related to policy analytical capacity. Specifically, the competency framework of the GES expects from the civil servants to be able to conduct impact assessment, to manage analytical projects and be familiar with data utilization and analysis (GES, 2013). Moreover, the GKIM profession, inside its professional skills framework, prescribes that the civil servants, even at the practitioner level, should be able to understand critical decisions and provide corresponding analytical services (GKIM, 2009). Regarding data scientists, who are assigned as DDaT professionals; they are required to employ data analysis and synthesis, by collecting, collating, cleansing and interpreting data in order to derive meaningful insights (DDaT Profession, 2017a). Respectively, DDaT performance analysts are expected to produce performance measurement frameworks, while at the senior levels, civil servants are requested to pursue ways of improving policy implementation (DDaT Profession, 2017b). Moreover, professionals inside the GORS present equivalent policy analytical capacity, as they are able to conduct advanced spreadsheet analysis, data analysis, and regression, so to provide valuable insights for certain policy areas (GORS, 2014).

Similarly, but by employing different types of research, GSR professionals support the implementation and evaluation of policy and delivery, by conducting in depth-data and objective analysis. This task is supported by operations, such as gauging for social trends, formulating government policy and producing statistics (GSR, n.d.). In a comparable fashion, civil servants of the GSE profession are requested to be aware of the most contemporary analytical approaches, in order to be able to conduct pro-active research in the field of science

and engineering. (GSE, 2013, p. 4). Additionally, the GSE officials acknowledge the policy substance, as they are requested to always take into account the political, policy and operational impact of the issue that they examine (GSE, 2013, p.15). Accordingly, the GSS professional group, presents indistinguishable skills and abilities from the previous policy analytical professions, as the civil servants who work as statisticians are capable of generating and employing repeatable statistical applications (GSS, n.d.). It is interesting to note, that GSS professionals, under the data analysis competency, are expected to "*actively seek out example of statistical innovations in other internal and external products,* in order to consider their appropriateness, applicability and redeployment within their own work area (GSS, n.d. p, 9). By utilizing data analytical capabilities, the statistician's duty is to maximize the statistical potential of the existing data sources, so to share insights and best practices with their counterparts at the policy profession (GSS, n.d.). In the same manner, professional intelligence analysts can combine various bits of information with the aim of translating evidence into valuable insights for decision-making (IA, n.d.). Finally, the Policy professionals present the analogous policy analytical skills, as a policy public official is able to understand and analyze the information provided by colleagues in order to support the making of the policy design (PP, 2018). In more detail, a capable and skilled policy professional is able of comprehending the holistic context of his/hers policy area and has complete awareness of the policy-making substance.

To sum up, the proxy variable of the recognized policy profession allowed the inspection of indicators such as, the *ability of applying analytical techniques, the ability to utilize statistics, analytical methods, research and evaluation, as well as,* the *awareness of the policy substance*. Thereby, by examining the competency framework and operational activities of the 9 respective professions (*see Table 3)*, it is made explicit, that these certain professionals form the base of the civil service workforce, that can display considerable policy analytical capacity. In turn, the corresponding indicators suggest that a noticeable part of the UK civil service is equipped with policy analytical capacity. Having already established that the competency framework of these 9 recognized civil service profession indeed corresponds to the skills set that Howlett (2009) describes for policy analytical capacity, it is now time to illustrate the civil service's composition, so to distil evidence suggesting the extent to which policy analytical capacity is present inside the four selected public organizations separately, as well as, across the public-sector UK-wide. The evidence for identifying the presence of policy analytical capacity have been found in account and pattern evidence that originate from official government documents and statistics and they are the following.

To begin with, by inspecting evidence regarding the independent variable of policy analytical capacity, that resulted by utilizing the civil service recognized *profession* as a measure, it can be observed that the civil servants who occupy the professions, that are identified as the ones that require individual policy analytical capacity, make up the 13.5% of the total civil service workforce in the UK (NAO, 2017, p. 62). Specifically, by looking at data published for each of the selected departments (ONS, 2016), Defra's civil servants occupying professions that require policy analytical capacity, make up the 33.1% of the public organization's total workforce. The same number for the Department of Health fluctuates at 19.2%, while in the HMRC the percentage of the total workforce, that is equipped with policy analytical capacity is at 4%. Accordingly, the DWP presents the number of 4.2% for civil servants that are assigned in the corresponding policy analytical professions (ONS, 2016).

Table 4: Professionals with Policy Analytical Capacity by department in 2016[3]

Department	Number of Professionals with policy analytical capacity	Total Workforce[4]	Percentage of Civil Servants with policy analytical Capacity
Department of Health[5]	1420	7390	19.2%
HM Revenue and Customs[6]	2890	70510	4%
Department for Work and Pensions	3620	84920	4.2%
Department for Environment, Food and Rural Affairs[7]	2010	6060	33.1%

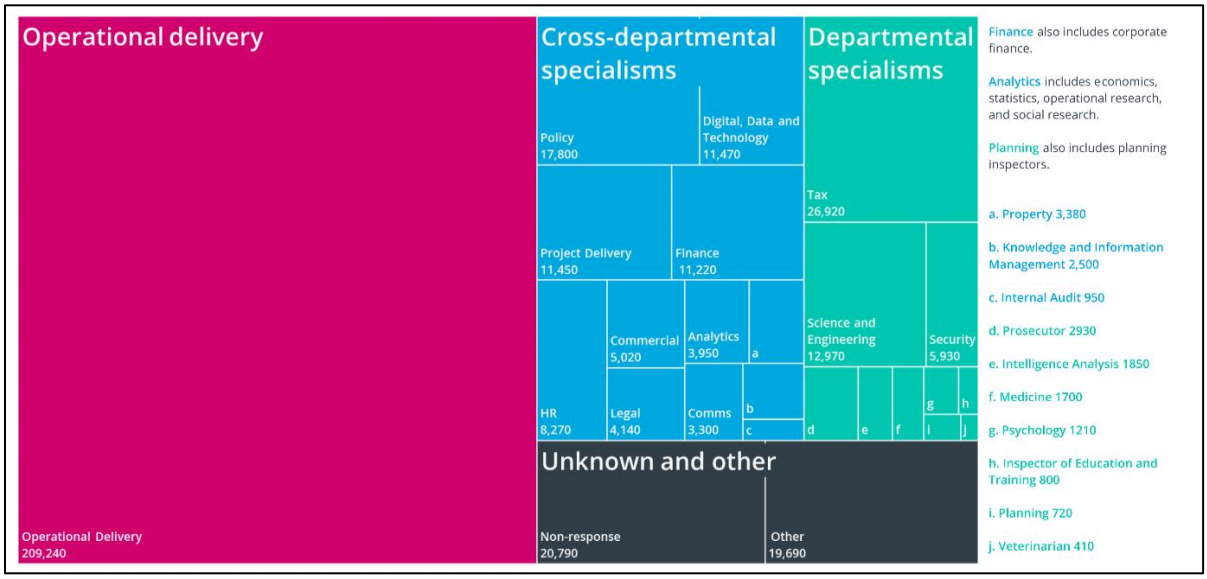
Source: Office for National Statistics, 2016

[3] The term "Professionals with Policy Analytical Capacity" refers to civil servants who are assigned in the nine professions that present policy analytical capacity within their competency framework
[4] The Office for National Statistics declares that data is considered to have lower level of precision than other HR data; this is the reason why there are some fluctuations between the total workforce presented here, and the total workforce presented in the annual reports cited during the case selection section.
[5] Includes Public Health England
[6] Includes the Valuation Office Agency
[7] Includes the Plant Health Agency and the Rural Payments Agency

However, one must take into account that the percentage of civil servants that are qualified with policy analytical capacity skills, fluctuates in each of the departments depending on the purpose of the organization, and subsequently on its operational and staff needs. For example, Defra, being involved with environmental protection and the food and farm industry, presents a high number of (620) civil servants in the Science and Engineering profession, in relation to its total workforce, that is 2010 individuals (ONS, 2016). On the contrary, the percentages of the policy analytical capacity equipped civil servants, that are presented in the HMRC, DWP, are presented as fairly low because the overwhelming majority of the civil servants in these two largest departments, is assigned to the Operational Delivery profession, as the needs in this service are greater (*see Figure 2)*.

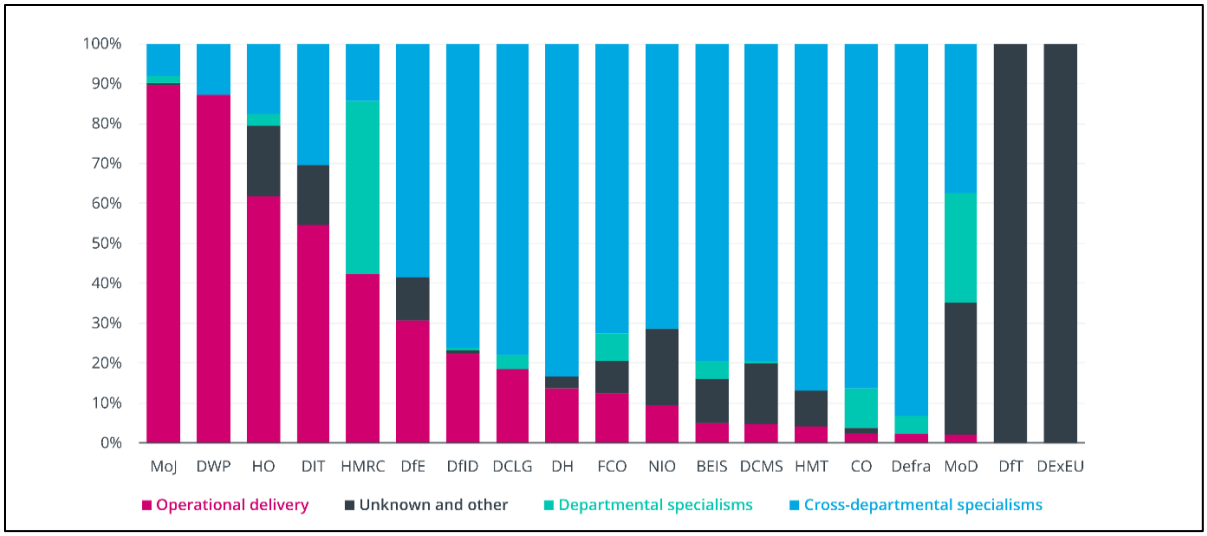
In sum, in 2017, the 54% of the UK government's total 389.000 civil servants work inside the operational delivery profession, while the 36% of the civil servants are acquainted with Cross-department and departmental specialisms working in the rest 27 professions (IFG, 2017). From the total 28 professions, the remaining 9 professions that are characterized as specialisms that require policy analytical capacity, make up the 13.5% of the total civil service workforce. It is important to note, that if we control for the 209.000 civil servants who are working in operational delivery and, the total number of civil service workforce is 180.000 individuals, from which the 50,540 are working in the 9 policy analytical professions (IFG, 2017), making up the 28.0% of the civil servant workforce; excluding the civil servants engaged in offering public services immediately to the public. Additionally, NAO suggests, that the main capability gaps for the civil service, are related to three priority skills of digital, commercial and project delivery (NAO, 2017, p. 7). Most importantly, there are no identified civil service capability (skills shortages) and capacity (staff shortages) that are pertinent to the policy analytical functions, as they are presented for the rest policy analytical capacity recognized professions. Thereby, policy analytical capacity is present in every selected governmental department, although its degree varies depending on the public organization's purpose and workforce composition (*see Figures 1,2,3)*. Subsequently, evidence suggests that policy analytical capacity is dispersed throughout the UK public sector, and subsequently there are no skill or staff shortages concerning policy analytical capacity.

Figure 1: Specialism of Civil servants (full time equivalent), March 2017



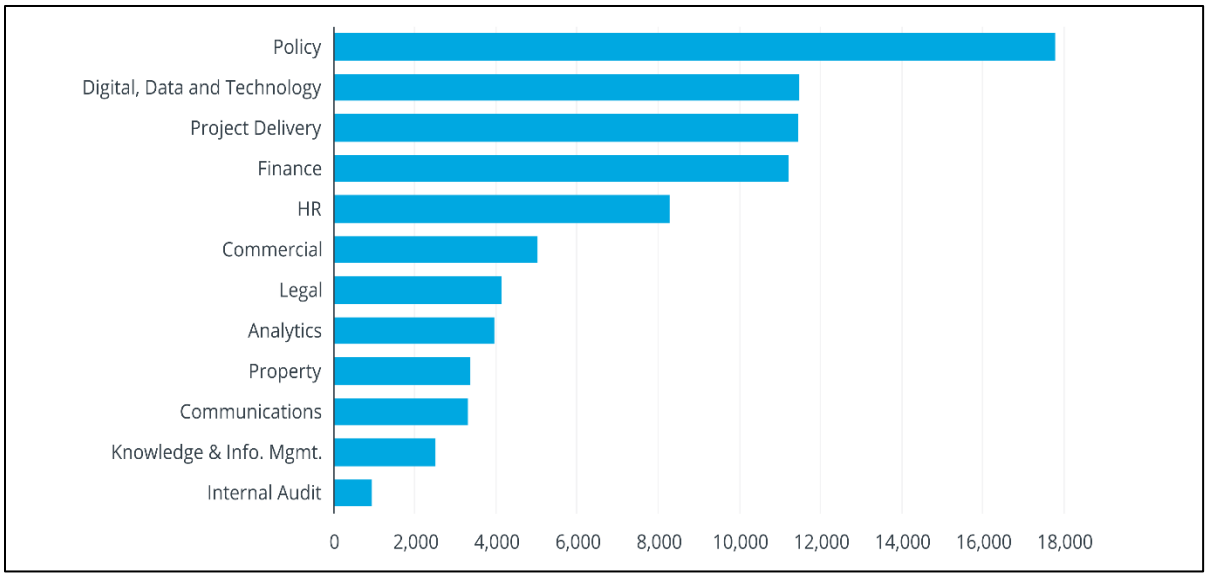
1:Note. Reprinted from Specialisms in the Civil service, IFG, 2017.

Figure 2: Specialisms of civil servants (full time equivalent) by department, March 2017



2:Note. Reprinted from Specialisms in the Civil service, IFG, 2017.

Figure 3: Cross-departmental specialisms in the civil service (full time equivalent), March 2017



3:Note. Reprinted from Specialisms in the Civil service, IFG, 2017.

4.4 Research findings on Digital Literacy

Concerning the independent variable of digital literacy, by supporting the research on the indicators prescribed on the research methodology and by searching for evidence in governmental and private company reports, the evidence found, suggest that there is digital skills gap within the UK civil service that cannot pass unnoticed. The relevant evidence is going to be presented.

Beginning with inspecting a Deloitte survey concerning digital skills and expertise within the UK public sector, findings signify that the 93% of the respondents "*find workforce and skills to be a challenging area to manage in their organization's transition to digital transformation*" (Deloitte Digital, 2015). This answer implies the condition that civil servants in the UK face difficulties in working with digital environments due to skill shortages. A more recent survey of Deloitte indicates that the same problem remains, as it suggests that public organizations lack skills which are required for utilizing the digital transformation, "*either because they could not afford the high salaries that digital professionals can attract or because their senior teams were not sufficiently aware of the possibilities of new technologies*" (Deloitte, 2017, p. 30). Thereby, it can be observed that the mentality of digital literacy and the awareness towards

mastering digital-driven ideas is noticeably absent inside the UK public sector. Similarly, inside the most recent three-year survey of Sopra Steria (2017), which examines the civil servants' disposition towards digital transformation, findings suggest that the 43% of the civil servants do not have adequate digital skills training to do their job (Sopra Steria, 2017). This statement within the civil service suggests that digital literacy is not a competency that civil servants of the UK feel comfortable with. This can be inferred from the fact that *"digital skills are consistently identified by civil servants as a significant (often the most significant) barrier to digital transformation"* (Sopra Steria, 2017, p. 8). In other words, the report boils it down to the point that conveys that digital skills and capabilities are lacking within the UK public sector and the civil service does not exploit its own data to support its transformation.

Additionally, a report by the company TechUK suggests comparable digital skill shortages. In more detail, more than 41% of the civil servants working in a public organization in the UK, tend to agree or strongly agree that their organization does not have the critical digital skills required in the area of data management and data analytics (TechUK, 2017, p. 14). It is important to note, from the same report, that although the 44% of the civil servants in 2017, state that they are excited about the advent of Big data analytics in the public sector, perceiving innovative technology as an enabler for public sector development, nearly 1 out of 4 Civil servants admit that skill shortages are a barrier to transforming the public services.

Finally, the National Audit Office, strengthens the evidence for the UK civil service's digital skills gap, as it reports that the government currently experiences a "*large capacity gap for people with digital skills"* (NAO, 2017, p. 38). Likewise, the central government faces a growing capability gap, "*that is the difference between specialist staff currently employed by government in a given specialism and the staff that departments think they will need in five years"* (NAO, 2017, p. 38) Particularly, public organizations and departments of the UK government in five years, will be restrained by a "*total headcount gap of 2000 staff´* with digital, data and technology-specific capabilities (NAO, 2017, p. 38). Consequently, by investigating the relevant evidence it can be observed that, currently, civil servants in the UK public sector do not have the adequate degree of digital literacy, as they face difficulties when they are called to work with digital applications. Additionally, the UK civil service does not only experience digital skills gap but might face staff shortages that can aggravate the former problem. To sum up, digital literacy and ICT expertise are conditions that are noticeably absent across the UK public sector.

4.5 Research findings on Digital-Era Governance

Once again in order to locate observable manifestations of the DEG paradigm inside the UK public sector and explore the causal chain of evidence that can lead to big data assimilation, certain indicators are being employed. Features of joined-up governance, such as collaboration between governmental departments across the public sector, stand as indicators that the UK is applying the Digital-era Governance paradigm. Moreover, organizational reintegration, as well as, procurement concentration and specialization are additional signs of the DEG. Additionally, the paradigm cannot be established without the presence of one-stop public service and electronic service provision, as well as, open-data government attitude. Thereby, all the above traits stand as indicators as well. Based upon the previous guidelines and by searching official government documents and interpreting international organization's indexes on digital government, the indicators signify the presence of confirmatory evidence towards the establishment of the DEG inside the UK public sector.

In more detail and beginning with examining the Government Digital Strategy of 2012, it is observed that the government at the time posited the aim of transforming and rebuilding the high-volume services, in order to make them *"digital by default"* (Cabinet Office, 2012). Some of the main components of this transformation, were the re-designment of the transactional services in order to adapt to the new digital standards, the establishment of a single online government digital platform in order to improve the departmental publishing of open data, as well as, the employment of common technology platforms that will support the new digital services. It is evident that the UK government has indeed involved in restructuring its processes and reinventing the public sector's purpose. This can be supported, by inspecting the more recent Digital Transformation Strategy published in 2017. Inside the 2017 policy paper, it can be observed that the UK public sector was indeed reestablished, as departments restated their processes by sharing platforms, components and the best digital practices in order to improve their interaction (Cabinet Office & GDS, 2017). Additionally, as it was expected since 2012, the UK government managed to develop and establish the "*award winning and internationally renowned GOV.UK platform"* (Cabinet Office & GDS, 2017, p. 5), concentrating all the government-citizen interactions under a single website and establishing a digital platform that has the same architecture across the whole government.

By looking thoroughly into policy papers, the government of the UK can be characterized as being in alignment with the DEG premises, as it aims to achieve reintegration and digital change by pursuing to establish the next three broad components until 2020. Specifically, the UK government today, wants to first transform the whole citizen-facing services for improving the service recipient's experience through digitally enabled solutions; second, aims to fully transform the governmental departments in order to deliver policy objectives more flexibly, efficiently and citizen-friendly, and third, desires to achieve an internal government transformation, that is crucial for the government to collaborate better and manage the digitally enabled change more decisively (Cabinet Office & GDS, 2017, p. 5). It is observed that, the government in the UK, based upon the digital-era influenced provisions, is disposed towards securing the dominance of digital public services and completing the internal and external digital transformation of all the governmental departments, in order to overcome organizational boundaries and make them more efficient, effective and technologically-friendly. The aim of the UK government is to keep transforming the public sector digitally, and the purpose of this transformation is to achieve an open, joined-up and digital government; public sector attributes that are similar with the ones that the digital-era governance paradigm fights for. But, can evidence suggesting that purpose, actually be found inside the UK public sector?

To answer that question, concerning openness, J. Manzoni, the Chief Executive of the Civil Service, stated that the UK "*released over 30.000 non-personal data sets in machine readable formats, for not cost, and open for anyone to use or build upon" (Manzoni, 2017).* The validity of this statement is strengthened by the fact that OECD presents the UK as an exemplar country when it comes to open government data. By inspecting OECD statistics, it can be observed that the UK has "*the broadest standards on open government data among OECD countries,"* as it has the fourth highest score globally in data availability, accessibility and support for private sector re-use of government data (OECD, 2017). Regarding the joined-up governance aspect that discerns the DEG approach, evidence suggest that the UK government presents such characteristics. Joined-up governance features remain until today in the government's agenda, as the UK government lists the implementation of joined-up, back-office and end-to-end services as a top priority for its public service transformation strategy until 2020 (Cabinet Office & GDS, 2017). Thereby, the previous and current government strategy is in agreement with the DEG governance.

Concerning, the degree of the digital government implementation, evidence suggests that the UK once again stands as a world-class deliverer of digital services. By examining the practices

that the UK government follows during the last years, it appears that it employs the digital technology to the fullest in order to provide high quality digital services to the largest possible audience. Respectively, the wide-spread provision of digital services by the government and their usage by the British citizens is apparent from the next reported instances. Specifically, from 2015 to 2016, "*over 33 million people taxed their vehicle online, 93% vehicle tax transactions were done online, and over 5.7 million people used the voter registration digital service, with a peak of 469.000 online applications in one day"* (Cabinet Office & GDS, 2017). Moreover, since 2011 that is when the Government Digital Service kickstarted the GOV.UK and began implementing its digital by default service revamp, "*3.56 billion pounds were saved for taxpayers for the three financial years from 2012 to 2015" (*Cabinet Office & GDS, 2017, p. 18*),* as the introduction of consistent digital services, helped the public sector to work more efficiently and repurpose its budget. More recently, additional digital services were introduced by the government, with the purpose of providing holistic services, just as the DEG paradigm suggests. Particularly, the UK government developed the GOV.UK Verify digital service, that is an identity assurance service that allows the citizen to prove his/her identify online and provides single access to all government services, ranging from checking tax to applying for social security benefits (GDS, 2018). Similar examples of holistic service provision across the government are the GOV.UK Pay and Notify systems (Cabinet Office & GDS, 2017, p. 13).

Lastly, by examining the indicators of procurement concentration and collaborating with the private sector for specialization, it can be observed that the UK government has indeed focused on such practices. Under the Digital Transformation Strategy, an objective for the government is to apply its tools, techniques and digital culture in order to improve its procurement strategies. For example, the GDS is currently operating the Digital Marketplace, "*which has embedded user-centered, design-led, data-driven and open approaches across digital, data and technology procurement and contracting"* (Cabinet Office & GDS, 2017, p. 44). Moreover, governmental departments are endorsing new ways of interacting with the private sector, by commissioning more collaboration between organizations, by using more complex contracting methods, and by attempting to improve the success rate of digital projects, by reducing reliance on external IT providers and utilizing more agile methods and more in-house delivery (NAO, 2017, p. 15). Consequently, it can be said, that the UK government follows the DEG paradigm, and concentrates on procurement concentration in a new technologically-influenced way, while it seeks private partnerships prudently, in contrast to what the NPM doctrine supported. From all the above, it can be observed that the UK government indeed relies on practices driven from

the DEG paradigm, as the indicators for signifying the exercise of Digital-era Governance, demonstrate the presence of corresponding evidence that support the former argument.

As the relevant data and the subsequent results have been presented throughout this section, it is now time to proceed to the study's analysis, where the corresponding evidence will be evaluated against the probability tests of the process tracing method. The aim is to assess the inferential weight of the evidence reported here, in order to affirm or disconfirm the respective hypotheses.

5. Analysis

The intention of the study is to answer the question of "*How does policy analytical capacity affect big data assimilation in the public sector?".* Additionally, the study aims on answering the relevant sub-question, prescribing that the DEG paradigm can have a positive effect of big data assimilation. Thereby, the scope of the study is extended towards providing insights about the possible alternative causal sequences that can result in the dependent variable of interest. Respectively, throughout this section, the relevant evidences that can help in answering these questions and confirming the main or alternative hypotheses of the study, are going to be evaluated under the light of the theory and under their own inferential weight.

As it is prescribed during the study's research design section, the method of analysis that is going to be employed is the process-tracing method. The first step of process tracing involves the elaboration of the causal mechanism that is going to be tested (Punton & Welle, 2015). Respectively, during this section the causal sequence between the hypothesized cause of policy analytical capacity and the outcome of big data assimilation is going to be deconstructed in detail.

5.1 Assessing how policy analytical capacity affects big data assimilation in the UK public sector

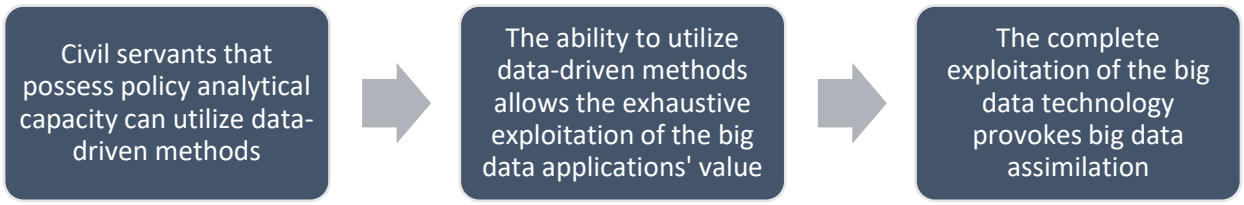
To begin with, policy analytical capacity is examined at the individual level, as the British civil servant stands as the unit of analysis. Accordingly, policy analytical capacity refers to the ability of the civil servants to apply statistical, applied research and modeling methods in order to be aware of future policy impacts (Howlett, 2009). The first part of the causal mechanism, originating from theory, prescribes that civil servants who are equipped with the policy analytical skills, are also competent in utilizing data-driven methods (Howlett, 2015, p.178). By observing the competency frameworks of the nine identified policy analytical capacity civil service professions, it can be inferred that the civil service standards require from the civil servants to able of conducting data analysis, regression or similar scientific methodologies. In addition, the respective civil servants are assumed to be having policy analytical capacity, as at the same time, are able of evaluating policy impacts and are aware of policy making issues. In turn, by observing the civil service composition, the findings seem to suggest that individual

policy analytical capacity is found in every selected governmental department, allowing the evaluation of the next part of the causal mechanism.

The second part of the causal mechanism, hypothesizes that civil servants who are able of utilizing data-driven methods, are familiarized with handling big data sets, are aware of the big data applications' operational capability, and as a result are able to draw more value from big data, as the proper employment of a big data application requires acquaintance with the subject. In turn, by looking at findings on both policy analytical capacity and big data assimilation, it can be observed that productive employment and assimilation of big data applications is actually achieved in every of the four public organizations. Subsequently, the successful employment of at least one big data application within the four public entities can be credited to the organizations' workforce; The proper utilization and employment of big data applications can be deemed as a condition that is likely to have achieved due to the competences of policy analytical capacity equipped civil servants. In other words, civil servants help the organization in facilitating big data applications, by being operationally ready for applying their complex analytical skills that in turn enable the utilization data-driven methods. Subsequently, the successful utilization of the big data potential, can be credited to the presence of policy analytical capacity within the organizations' premises.

Moving to reviewing the last part of the causal mechanism (*see Figure 4)*, it is expected that the extraction of additional value from big data, through the appropriate utilization of the big data technology, helps the organization to institutionalize the big data application, as the organization enjoys the advantages that the application has to offer. Alternatively, by relying on a counterfactual reasoning, it is assumed that if the public organization lacks policy analytical capacity, it will not able of utilizing big data properly, and in turn it will not be incentivized towards institutionalizing a big data technology. The relevant findings support the former argument and increase the confidence towards confirming the causal mechanism's assumptions, as all of the organizations are found enjoying benefits, such as enhanced organizational performance and improved public service delivery; benefits that are yielded through the use of big data. However, the sub-hypothesis derived from the last part of the causal mechanism cannot be conclusively confirmed, as it can alternatively be hypothesized, that the realized value extracted from big data applications is not an enabler of big data assimilation, but big data assimilation is the causal agent, which creates public value in the first place.

Figure 4: How policy analytical capacity can affect big data assimilation



Having analyzed the causal mechanism and having stated the durable and less strong parts of the causal mechanism, I will now proceed to the next part of the process-tracing method, where the inferential weight of the evidence is tested against specific hypotheses. By weighing the inferential impact of the evidence, a reasonable degree of confidence is assigned to parts of the causal mechanisms (Punton & Welle, 2015). To begin with, the hypothesis originating from the causal mechanism construction, is the following:

Hypothesis: Public organizations in the UK can assimilate big data applications, because their personnel is equipped with policy analytical skills, that allows them to utilize data-driven methods and in turn extract value from big data analytics.

Alternative Hypothesis: Public organizations in the UK can assimilate big data applications, not because they present policy analytical capacity, but due to other confounding enablers that contribute to big data assimilation.

By taking into account all of supporting evidence for policy analytical capacity, as well as the evidence supporting the causal mechanism's sequence resulting in the verifiable condition, that these particular organizations do indeed rely heavily on big data application employment, as they converted the applications as an integral part of their organizational strategy, the main hypothesis of the study can be accepted. However, the accepted hypothesis should be taken with a grain of salt, as the last part of the causal mechanism cannot be validated decisively. In other words, the evidence found fall under the straw in the wind test, suggesting that policy analytical capacity is indeed present in the UK public sector; however, as the evidence can be framed as being circumstantial, it cannot decisively prove that big data assimilation is the outcome of policy analytical capacity, nor it can decisively reject the alternative hypothesis.

Finally, the confidence that the main hypothesis of the study is true, can be increased; however, the evidence is not definitive enough to confirm the hypothesis categorically, as alternative causal factors or different causal sequences may have helped public organizations in the UK to achieve big data assimilation. Nevertheless, it can still be supported that a public organization,

that is competent in policy analytical capacities within its workforce, is also able of improving its operational procedures and producing added public value by assimilating big data applications. In other words, policy analytical capacity can still be assumed as an enabler of big data assimilation.

Before proceeding to testing the next hypothesis, I must note that by analyzing the findings on the professions suggesting policy analytical capacity, an additional theory-driven implication, was identified empirically. Specifically, it is apparent that the certain civil service professions support the evidence-based policy practice heavily. For example, the SEG framework stipulates that the profession's function is to "*provide evidence-based advice to support operational delivery and to inform policy decisions" (Government Office for Science, 2016, p. 7).* Additionally, a duty of GSE civil servant, is to "*promote the culture of evidence-based policy making" (GSE, 2013, p. 7).* Respectively, the competency framework of the policy profession expects from the professional to be able to "*support development of well informed, evidence-based policy recommendations" (PP, 2018).* Accordingly, the rest of the analytical professions prescribe similar practices for expanding the government's evidence-base (GES, 2013; GKIM, 2009 GSR, n.d.); through coherent scientific analysis, data utilization, and policy awareness, analytical professions have the function of producing evidence for supporting policy decisions and yielding improved policy outcomes. Thereby, evidence infirm that there is indeed a connection between evidence-based policymaking and policy analytical capacity, where the latter is a requirement for the successful implementation of the former, as Howlett (2009) suggests. After all, it is not serendipitous that most of the analytical professions' competency frameworks signify their goals in accordance to the evidence-based policymaking practice.

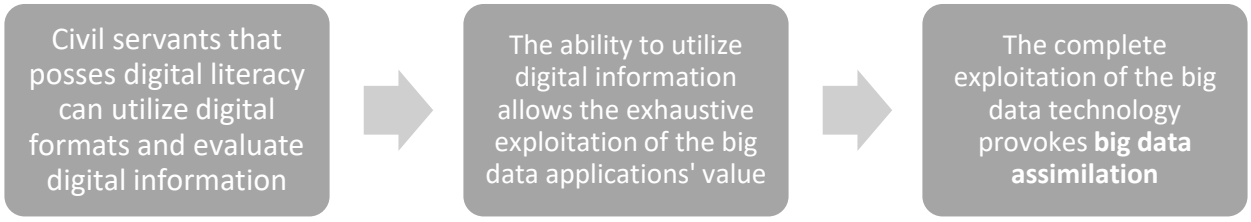
5.2 Assessing how Digital Literacy affects big data assimilation in the UK public sector

Having already specified the causal sequence that connects policy analytical capacity with big data assimilation, it is now time to proceed to examining the alternative hypothesis. Under this section, the possible impact of digital literacy on big data assimilation is going to be put to the test.

Digital literacy, just like policy analytical capacity, is a variable that is going to be examined at the individual level, as the British civil servants stand as the unit of observation. Digital literacy

refers to the civil servant's ability to perform tasks in a digital environment, due to its proficiency in digital technology and computing skills. The causal mechanism begins by assuming that a civil servant who is competent in digital skills, is also equipped with the required literacy per se, and ICT literacy for utilizing digital formats and creating, communicating and evaluating digital information (Bawden, 2008, p. 29). Consequently, it is assumed, that an experienced in digital environments civil servant is also capable of utilizing data-driven digital applications. Being able of handling, creating, communicating and evaluating digital information, big-data sets included, a digitally literate civil servant is competent in exploiting the complete potential of big data applications. In turn, big data assimilation is perceived as the outcome of having digital literate civil servants within the organization. The theory is line with this argument, as IT knowledge is considered to be an important enabler of IT assimilation, at least when it is apparent inside the managerial ranks (Ranganathan et al., 2004). Thereby, by following this causal sequence, digital literacy is considered as a stepping stone towards achieving big data assimilation. *(see Figure 5)*

Figure 5: How digital literacy can affect big data assimilation



Proceeding to the probability test, as findings already suggest that big data assimilation is actually taking place inside the UK public sector, digital literacy is tested as a necessary requirement for big data assimilation to occur; or in other words as a precondition for assimilation, that stands alternately to policy analytical capacity

By taking into account the additional findings on digital literacy, the main hypothesis of the study is tested again by assessing the evidence relevant to digital literacy.

Hypothesis: Public organizations in the UK can assimilate big data applications, because their personnel is equipped with policy analytical skills, that allows them to utilize data-driven methods and in turn extract value from big data analytics.

Alternative Hypothesis: Public organizations in the UK can assimilate big data applications, not because they display policy analytical capacity, but because they are equipped with digital

literacy that enables them to utilize digital formats, and in turn allows them to exploit the potential of the big data applications.

By looking at data drawn for the whole UK public sector, evidence suggesting considerable presence of digital literacy cannot be located. In other words, evidence that support the presence of digital literacy within the ranks of the UK public service are absent. Thereby, the evidence falls under the Hoops test, as the evidence required for the alternative hypothesis to be true is not found, and subsequently the alternative hypothesis is rejected because it fails to display the necessary empirical requirement. In contrast, governmental and private company reports present that the UK public sector is struggling with significant digital skill and staff shortages. However, in spite of the deficiency in digital literacy, evidence show that big data assimilation has been realized within the UK public sector, at least inside the four selected public organizations. Thereby, digital literacy is not a necessary condition for big data assimilation.

Consequently, the alternative hypothesis, prescribing that big data assimilation is the result of digital literacy can be decisively rejected, as big data assimilation in the British public organizations occurs in the presence of significant digital skill gaps within the UK public sector. However, it must be noted that the evidence is not relevant for counteracting the argument of Bawden (2008), as the evidence suggesting the absence of digital literacy are not pertinent for disproving that a digital literate civil servant can utilize data-driven systems. Additionally, IT knowledge can still be theorized as an enabler of IT assimilation (Ranganathan et al., 2004), however it can be inferred that this is not the case in the UK public sector, as big data assimilation is accomplished without the supposed catalytic effect of IT knowledge. Alternatively, evidence suggest policy analytical capacity as a possible enabler of big data assimilation. In a few words, the causal mechanism related to digital literacy (*see Figure 5)* is disconfirmed. The remaining assimilation facilitator that is left to be analyzed, is the dominant theory for public sector development.

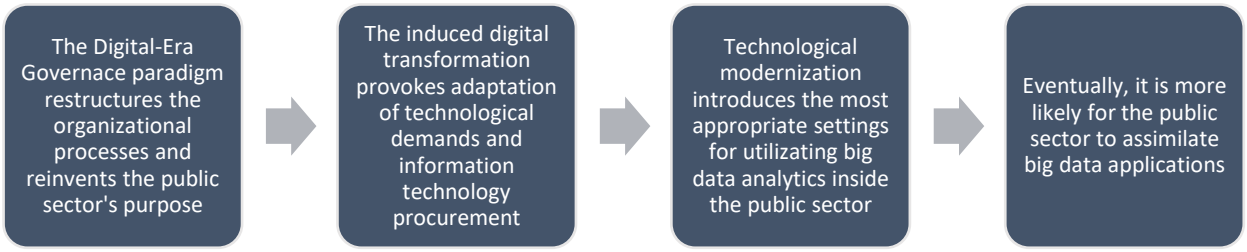
5.3 Assessing how Digital Era Governance affects big data assimilation within the UK public sector.

Except for the catalytic function that intra-organizational factors, such as policy analytical, might have on big data assimilation, an alternative explanation for the outcome of big data assimilation is the impact of the macro-theory for public sector development. Accordingly,

through this section, the possible contribution of the Digital Era Governance approach to big data assimilation in the UK is going to be analyzed.

The assumption, that the DEG has a positive effect upon the adoption of big data applications, begins, by observing the main characteristics of the macro-theory. The first main feature of the DEG, that is reintegration, advocates the establishment of simple, efficient and shared central government processes (Margetts & Dunleavy, 2013). Moreover, agile and resilient government structures, such as one-stop processes and electronic public service delivery, are some of the DEG's additional distinguishing administrative traits. Taking into account the above premises, I hypothesize that, in the attempt to introduce such features into the public administration, the DEG restructures the organizational processes and recalibrates the aim of the public sector. Consequently, the public sector aims on becoming more joined-up, reintegrated, open and transparent, and most importantly digitally driven. However, in order for the respective government to acquire such components, it first adapts to the contemporary technological demands and overhauls its information technology procurement. Subsequently, the public sector is more prone to using big data analytics, in order to modernize the digital processes and gain access to unprecedented public service delivery capabilities. To sum up, it is hypothesized that the DEG inserts a technology-oriented attitude into the public sector, where the utilization of big data applications is made effortlessly, as the public sector has the appropriate settings for welcoming the assimilation of big data applications *(see Figure 6)*.

Figure 6: How the Digital-era Governance can affect Big Data Assimilation



In line with what Margetts & Dunleavy (2013) suggest, evidence show that public sector in the UK is indeed experiencing a shift by focusing on "*reintegrating services, providing holistic services for citizens and implementing thoroughgoing digital changes in administration*" (p. 1). Findings show that the UK government has revamped its services with the purpose of making them "*digital by default*". In addition, the public services have become more "*citizen-friendly*" and the government-citizen interaction is supported by digital platforms. A key observation that certifies that the UK government undergoes a digital revamp and a public-sector remodeling, as

the first part of the causal mechanism suggests, is that the government aims to establish an open, joined-up and digital government in the next two years (Cabinet Office & GDS, 2017); a goal that the British public sector has already managed to accomplish in a considerable degree, as it enjoys the reputation of being a world leader in open data government (OECD, 2017) and it already provides high quality digital services to a large portion of the population. Consequently, as the publishing of public data and the establishment of joined-up, end-to-end digital services (Cabinet Office & GDS, 2017) is a strategy that springs from the digital change component of the DEG approach (Margetts & Dunleavy, 2013), it can be inferred that by implementing such open data initiatives and digital services, the UK government aligns its purpose with the one that the DEG paradigm suggests.

Specifically, the UK government's strategy for public sector development, satisfies all of the DEG components, as they are prescribed by Margetts & Dunleavy (2013). The *reintegration* component is empirically satisfied by the government's initiative to centralize the government processes by the establishment of a single digital platform, inside which, governmental departments and citizens can interact in simple terms. Respectively, the *needs-based holism and the digitalization* components of the DEG approach are fulfilled by the UK government's endeavor on founding end-to-end digital services that are citizen-centered, and by the fact that the UK government has already designed a Digital Service Standard that is responsible for ensuring the continuous employing of innovative citizen-oriented digital services in the long-term (Cabinet Office & GDS, 2017). The DEG demands the redesign of services from the client perspective and complete embracement and embeddedness on electronic delivery (Margetts and Dunleavy, 2013) and this is what the UK government actually does. Overall, the reported findings are, as expected, in agreement with what Margetts & Dunleavy reported in 2013. Particularly, austerity pressure and budget cutbacks compelled the UK government to adopt joined-up governance practices. In this way, the UK enhanced its joined-up local public service delivery and reintegrated separate government delivery chains, by focusing on citizen co-production, by replacing top-down national regulatory controls and by introducing online services (Margetts and Dunleavy, 2013).

Moving to the next part of the causal mechanism, the reported findings are again congruent with the hypothesized causal sequence. Particularly, it is theorized that the digital transformation guarantees that the public sector is keeping up to date with the technological demands and supports ICT technology procurement. Respectively findings illustrate that the GDS has employed the Digital Marketplace, "*which has embedded user-centered, design-led, data-*

driven and open approaches across digital, data and technology procurement and contracting" (Cabinet Office & GDS, 2017, p. 44). Subsequently, it can be inferred that the government relies heavily on the potential of the technologically-driven digital services, as its departments are seeking to establish new ways of communicating with the private sector, by "*adopting more of a commissioning and whole-systems collaborative approach, by using more complex contracting methods and by trying to improve the success rate of digital projects, while reducing dependence on external IT providers, by adopting agile methods and more in-house delivery"* (NAO, 2017, p. 14). Thereby, the digital revamp does indeed provoke adjustment to the current technological demands and ICT procurement.

By the already established evidence, it can be inferred that the transforming organizational governance that the UK experiences, introduces appropriate settings for welcoming big data analytics in the public sector. Accordingly, Margetts & Dunleavy (2013) agree that this transformation is fueled by the contemporary digital trends, that force governments to *"innovate in catch-up mode"* in order to modernize their interactions with the citizens (p. 7). Additionally, Margetts & Dunleavy (2013) suggest that the reintegration component of the DEG approach is the cause that is responsible for the advent of the big data analysis in the public sector. The study's findings suggest likewise. The four examined public organizations of the study have gone through a digital overhaul and they have indeed integrated big data analytics as part of their standard operations; besides keeping up with the technological current, the aim of the big data analytics introduction inside the public sector was the potential for improved public services and the possibility of cutting down costs.

Having analyzed the causal mechanism's sequence against the empirical evidence, it is now time to assess the inferential weight of the corresponding findings, against the following hypotheses.

Hypothesis: The Digital-Era Governance is the established paradigm for public sector development in the UK, that stands as an enabler to big data assimilation by inserting the appropriate settings for using big data in the public sector

Alternative hypothesis: The UK government is not in line with the Digital-Era Governance paradigm and big data assimilation has not been achieved inside the public sector.

The evidence found fall under the Double-decisive test, as they confirm the suggested causal sequence in every step and are both sufficient and necessary to validate the hypothesis and disconfirm the alternative one. Certain evidence, such as the government's Digital

Transformation Strategy and the "digital by default" public services, prove that the UK has indeed engaged in an endeavor to restructure the public sector's processes and reinvent its goal. By assessing the next step of the causal sequence, evidence suggest that the digital transformation that the public sector undergoes, is forcing the public sector to adapt to the new technological demands and turn its focus on information technology procurement. Consequently, this digital and technological modernization allows the utilization of big data analytics inside the UK public sector, a condition deemed present, as the evidence suggest that big data analytics usage has resulted in exceptionally improved public services. Finally, evidence indicates that the public sector is capable of assimilating big data applications unobscured.

Consequently, the hypothesis prescribing that the UK does actually exercise the Digital-era Governance in its public-sector development, by constituting digital transformation, technological adaptation and procurement, can be confirmed. In support of the empirical findings, I argue that the DEG paradigm creates the appropriate settings for big data assimilation to occur successfully. Thereby, the suggested causal mechanism (*see Figure 6)* can be considered as valid under the light of the study's empirical evidence. In contrast, the alternative hypothesis, pertaining that the UK is not in line with the DEG paradigm and it has not yet achieved big data assimilation within the public-sector premises is discarded, as strong evidence points completely otherwise.

5.4 Academic implications

The study attempted to illustrate the connection between policy analytical capacity in the civil service and big data use inside a modern public administration. By examining the concept of policy analytical capacity at the individual level, as it is pointed by Howlett (2015), and by taking into account that capacity within government defines *"how data is dealt with or used at all",* as Giest (2017) claims, the study establishes a causal link, suggesting that the civil servants' policy analytical capacity is likely to be positively related with big data assimilation in the public sector. Except for the theoretical implication, the study contributes to the area of methodology, by introducing the recognized civil servant profession as a proxy variable for measuring policy analytical capacity at the individual level. Thereby, influenced by the work of Howlett (2015), and by relying on the operationalizable components of individual policy analytical capacity, the study managed to produce findings, supporting that policy analytical capacity can be assumed as a significant factor, that enhances the ability of public organizations to utilize technologically-

driven innovations, and in turn allows governments to articulate achievable goals through the assimilation of big data analytics. Respectively, the study highlights and confirms the claim that policy analytical capacity is intrinsically related with the evidence-based policymaking practice, as it can also be empirically observed that the civil service standards demand from the civil servants to have analytical capacity in order to be able to expand the government evidence-base. Thereby, it is identified that policy analytical capacity, stands as a precondition for the application of the evidence-based policymaking, as the theory suggests (Howlett, 2009).
Except for the main independent variable, the study aimed on including possible omitting variables into the research. Respectively, digital literacy was examined side by side to policy analytical capacity. Subsequently, the research's findings on digital literacy are opposed to the argument of Ranganathan (2004), prescribing that IT knowledge is an enabler of IT assimilation, however, the outcome of the study cannot totally subtract that argument, as digital literacy was observed to be fairly low in the UK public sector. Thereby, the assumption pertaining that digital literacy can be a catalytic agent for big data assimilation cannot be fully examined in this case. Nevertheless, the study managed to demonstrate that public organizations can accomplish big data assimilation, without having a considerable degree of digital literacy, and thus digital literacy is established as an unnecessary condition for big data assimilation to occur.

As the public use of innovative technologies and big data analytics is the general theme that governs the study, the impact of the Digital-era Governance was not neglected as a research issue. The study's outcome provides further support to the fact that the DEG paradigm and Evidence-based Policymaking are conjoined concepts, as they culminate in the use of big data streams by the public administration (Giest, 2017). In a few words, Digital-Era governance is affecting the way of how big data applications are being utilized, as it prepares the public sector for the advent of technological trends as big data. Most importantly, the study indicates that the paradigm for public sector development that governs a country can actually have an effect on whether big data applications are going to be assimilated by the public-sector entities. Regarding the case of the UK, it can be implied that the UK is experiencing similar long-run changes driven by the DEG development. This is also what Margetts & Dunleavy predicted back in 2013, as the most possible scenario for the future public-sector development in the UK. Respectively, the UK government, in the wake of the Brexit, is dedicated in enhancing government productivity and continuously tries to set the liberal democratic public administration in alignment with the digital and technological trends that are taking place in the

private and civil society. After all, the DEG-driven alignment and digital modernization are what makes the paradigm an enabler of big data assimilation.

In the end, with the help of empirical evidence, I managed to confirm and discredit the study's hypotheses, Although, this does not mean that the study does not suffer of scientific limitations. This is reasonable, as the validity of all the inferences and the reliability of all the arguments cannot be verified by a single study. In the next section, I discuss the limitations of the research and then I provide recommendations for academic research in the future.

5.5 Limitations and Suggestions

Similar to the rest types of research, the single-case research design presents limitations and flaws within its methodology due to its own nature. This study signified that policy analytical capacity and DEG are positively related with big data assimilation. However, this causal relationship is only examined inside the case of the UK. That is one of the limitations of the study; the single-case research design of the study provides a limited possibility for generalization beyond the case of the UK public sector. Thereby, the external validity of the study cannot be guaranteed, unless the inferences are made within the frames of highly similar, perhaps Anglo-Saxon countries, that present almost identical public-sector characteristics. However, limited generalization is not entirely problematic, as from the beginning, the scope of the study was to analyze the causal relationship between policy analytical capacity and big data assimilation in the UK. Accordingly, by establishing specific causal mechanisms, the study managed to conduct an in-depth analysis on the UK civil service capacity and the digital developments that the UK public sector undertakes.

The suggested causal mechanisms can function as the foundation for future academic research and provide impetus for examining this causal relationship in a differentiated public sector setting. By relying on study's implications, researchers can design a refined future study for establishing a more systematic relationship between policy analytical capacity and big data assimilation. In order to ensure additional external validity, a Large-N observational research study is suggested, with the aim of estimating the significance of the main causal relationship with more certainty. Additionally, a quantitative study also carries the inferential power for controlling the potential subjectivity of mine that this qualitative study may have. Nevertheless, the study's major hypothesized causal relationship can also be examined in a small-n

comparative design. This can be done, by selecting a set of countries that are as similar as possible, but they vary in respect to policy analytical capacity, while other potential causal factors are held constant. In this way, a future researcher can test whether the findings of this study can be replicated by comparing examples of other countries.

Regarding the inferential integrity of the research data, I would like to note that the dating of the data may raise considerations about the temporal order of causality. In more detail, the statistics collected on policy analytical capacity are dated back to 2016. However, big data assimilation is observed to have initiated earlier, since the start of this decade. Thereby, in order to further strengthen the temporal order of causality, stipulating that an outcome cannot occur earlier than the cause, it would be better if the data were referring to former years. However, prior data were not selected, as the same profession classification was not found inside the civil service statistics before 2016. Nevertheless, this does not mean that policy analytical capacity was totally absent prior to 2016. Policy analytical capacity is still assumed to have an effect on big data assimilation, as the civil service composition presumably remains relatively constant during this short time span.

Except for the methodological considerations, it is equally important to present the recommended path for future academic research questions. Although, the examined hypotheses of the within-case analysis were limited to the number of three due to restricted feasibility and the given time frames of the study, an additional possibly causally relevant variable and explanation had been presented inside the literature review section. Respectively, the hypothesis suggesting that a public organization that experiences several barriers to technological adoption, is less likely to assimilate big data applications can be used in future academic research. Some of the potential barriers to big data assimilation, that can trouble research in future studies, are the impact of politics, the absence of sufficient IT governance, and the lack of organizational and infrastructural readiness. This hypothesis is left to be examined outside of its theoretical frames, as in reality is highly likely for a public organization to experience several of these barriers at once. It is therefore important to identify the average causal effect of each of these barriers on big data assimilation, in order to help the public administration to anticipate the impact of the difficulties that it faces. Therefore, by inspecting this hypothesis, the academic field can inform the public administration for factors that are negatively related to big data assimilation, preventing the governments from adopting detrimental strategies and avoiding barriers to public sector development.

Additionally, scholars interested in the impact of policy analytical capacity and the practice of evidence-based policymaking in relation to the use of big data in public administration, can reproduce this study, by drawing inferences not only on the state-centric and national level, but by assessing the policy analytical activities of government systems that operate in multi-level governance frameworks, as introduced by Marks, Hooghe & Blank (1996). Instead of focusing on the effect of the theory for public sector development, novel research can extend to evaluating the impact of multi-level governance on the implementation of policies related to big data use in public administrations. Respectively, the unit of analysis becomes the subnational, provincial or international policymaking authority, that is interested in incorporating big data analytics into its administrative systems. Policy analytical capacity can still be theorized as an enabler of big data assimilation and can be tested to infer whether it has the same effect, not on a national but on a multi-level governance system setting.

Finally, one last suggestion is that the current research design can be made with the inclusion of the whole range of the policy analytical capacity, from the individual, to the organizational to the systematic level. By measuring the policy analytical capacity and by examining three separate units of analysis, the complete capacity of the public sector is unveiled, and therefore the researcher can draw inferences on big data assimilation with more certainty. To sum up, this research itself has managed to provide an answer to the scientific questions, pertaining that individual policy analytical capacity can have a positive impact on big data assimilation, as well as pointing out that the DEG acts as a catalyst for the assimilation of innovative technologies. However, the issue at hand is multifaceted and is nearly impossible to control for all the adjacent research issues. I hope, that these recommendations will truly contribute in future academic research, acting as a stepping stone for discovering new insights on the usage of big data analytics in public administration.

Conclusion

The purpose of this study was to identify the requirements for utilizing big data applications inside a public administration and to articulate the process towards assimilating big data technology as an integral part of a public organizations standard operations. After all, if the facilitators for utilizing data-driven methods are neglected and the process for achieving big data assimilation is omitted, the public sector is highly unlikely to effectively exploit the value that big data analytics have to offer; not to say that public officials might be forced to face the adverse effects of big data analytics. Respectively, the question that needed to be answered, was "*How does policy analytical capacity affect big data assimilation in the public sector*?" Along this question, the study's research interest was to uncover any possible alternative causal factors that can lead to the assimilation of big data applications. Accordingly, digital literacy was examined as a surrogate to policy analytical capacity. In the endeavor of finding additional causal explanations, the sub-question of "How does Digital-Era Governance paradigm can function catalytically for big data assimilation*?" emerged*. The dominant theory for public sector development was hypothesized as a key instrument for achieving the outcome of interest. The selected public-sector environment for examining these variables was the UK public sector, as the country overperforms in digital governance although it does not display the required digital skills within its civil service.

In the attempt to answer the research question, the study indicates that an array of big data applications is being used by the governmental departments of the UK government. Advanced data warehousing systems, simulation tools or predictive analytics are some of the big data applications that governmental departments of different purpose and composition operate across the UK public sector. Findings suggest that big data analytics have not only been adopted by the British public organizations, but they have also replaced the conventional systems of operations, becoming an integral part of their standard organizational processes. From the National Health System, to Revenue & Customs and from the Department of Work & Pensions to the Department for Environment, Food & Rural Affairs, big data analytics are being used extensively and this fact is apparent by the exceptional policy outcomes that they provide. Improved public service delivery, enhanced decision-making for future policy courses, upgraded intra-organizational operations and acquisition of considerable savings of public money are some of the benefits that the UK public sector enjoys by using big data. Accordingly, it is safe to

say that big data is not just a buzzword, but a tool that is already being put into use by modern public administrations.

Thereby, one of the main argument of the study is that big data assimilation has been achieved inside the UK public sector, and respectively, the main findings of the study indicate that policy analytical capacity and the DEG may have contributed to this process. The policy analytical capacity that the civil servants possess, can be theorized as the precondition that allows the civil service to exploit the value effusing from big data analytics. In turn the added public value provokes the public organizations to fully institutionalize the big data application, as this is the beneficial thing for the organization. Nevertheless, this argument should be received with skepticism, as the evidence can be deemed circumstantial and there are likely many more omitting factors that can influence big data assimilation positively. In either case, policy analytical capacity can still be assumed to be positively related to big data assimilation; a considerable condition that a public organization should satisfy in order to achieve big data assimilation. However, the same claim cannot be theorized for the variable of digital literacy, as the study indicates that the UK public sector assimilates big data applications, although at the same time it experiences significant digital skills and staff shortages. Thereby, digital literacy is not a necessary quality for an organization that aims on integrating big data applications.

In contrast to digital-literacy, the current paradigm for public sector development that governs the country is deemed as a highly influential factor in relation to big data assimilation. Specifically, findings suggest that the Digital-era Governance has entrenched its roots within the public sector of the UK, as the country's strategy for organizing the public sector presents the macro-theory's components in all their glory. In more detail, the post-NPM public sector environment in the UK is formulated on the basis of reintegration initiatives, that centralize the public organization's processes by the use of digital platforms. Respectively, the public services' establishment is done in line with the demanded needs-based holism and digitalization characteristics that ensure that the British citizen can have access on high quality, end-to-end digital services. Thereby, the UK public sector undergoes through a call for digital modernization; a transformation that requires from the public organizations to keep in sync with the technological progress and compels them to focus on information technology procurement. Consequently, this is the reason why the DEG, inserts the most convenient settings for the utilization of big data analytics; because the macro-theory has already made the public sector acquainted with the use of information technology and data-driven methods.

Thus, big data applications are being employed without any obstructions. Subsequently, the DEG is accepted as a factor that is causally responsible for big data assimilation.

By exploring for possible enablers and alternative causal sequences that can lead to big data assimilation, the study managed to provide coherent academic implications that can be utilized for filling literature void on big data use in public administration. In more detail, the study provides theoretical foundations and confirmable empirical arguments, that demonstrate the causal sequence leading to big data assimilation in the public sector. The concepts of policy analytical capacity and DEG are theorized as significant factors that influence the implementation of innovative technologies by the government. Moreover, the synthetization of literature for revealing the connection between the concepts drawn from different research branches, the proposed research methodology, introducing a new way of measuring individual policy analytical capacity, as well as, the enhanced analytical description of the empirical inferences, are some of the study's contributions, that can stand as apparatus for future academic research. By relying on this research, a future researcher can be guided towards developing refined research questions on the constructs of policy analytical capacity, evidence-based policymaking, and Digital-era Governance, always in relation to big data use inside a public administration.

In conclusion, by revealing some of the possible enablers that can advance the assimilation of big data, the study provides advice that can used by public organizations operationally. By relying on the study's findings public sector entities can exploit the study's finding, so to assimilate innovative technologies effortlessly. In other words, as immense time and money is being spent in the endeavor to integrate new technologies into the organizational processes of the public sector, this study provides both the academic field and the public officials with insights on where and how resources should be implanted in order for the big data applications to be assimilated effectively and efficiently. Governments across the globe should invest in their people with the aim of augmenting the policy analytical capacities of the civil service personnel. In addition, if countries will start following the developments around the Digital-era Governance, it is more likely for them to gain the desired impetus for materializing the benefits that big data analytics have to offer. The expected advantages of the big data technology should not be omitted by public administrations, as their potential for providing public value to the society is great. It is evident that we are all living inside the Digital Age, where our every action is based on the information we have. Big data functions as an information pool and public organizations should be ready to utilize this valuable resource. After all, "*information is the oil*

of the current century, and analytics is the combustion engine[8]*",* so why would the public sector not take full advantage of it?

[8] Quotation from Peter Sondergaard, EVP of Gartner

References

Ahn, M. J., & Bretschneider, S. (2011). Politics of e-government: e-government and the political control of bureaucracy. *Public Administration Review*, *71*(3), 414-424.

Albury, D. (2005). Fostering innovation in public services. *Public money and management*, *25*(1), 51-56.

Bawden, D. (2008). Origins and concepts of digital literacy. In C. Lankshear & M. Knobel (Eds.), *Digital Literacies: Concepts, Policies and Practices* (Vol. 30, pp. 17-32). New York, NY: Lang Publishing.

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.

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, *19*(1, 2), 5-16.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, *9*(2), 27-40.

Cabinet Office & Government Digital Service. (2017, February 9). *Government Transformation Strategy 2017 to 2020*. Retrieved May 3, 2018, from https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020

Cabinet Office. (2012). *Government Digital Strategy*. Retrieved May 2, 2018 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/296336/Government_Digital_Stratetegy_-_November_2012.pdf

Capgemini UK. (2017, June 09). *Can big data analytics cure the NHS?* Retrieved April 24, 2018, from https://www.capgemini.com/gb-en/2017/06/can-big-data-analytics-cure-the-nhs/

Capgemini. (2014, May 28*). Big Data and Predictive Analytics Help HMRC Process Debt Payments More Quickly*. Retrieved April 25, 2018, from https://www.capgemini.com/resources/big-data-and-predictive-analytics-help-hmrc-process-debt-payments-more-quickly/#

Chen, Y. C., & Hsieh, T. C. (2014). Big data for digital government: opportunities, challenges, and strategies. *International journal of public administration in the digital age*,*1*(1), 1-14.

Civil Service. (n.d.). About us - Civil Service. Retrieved April 26, 2018, from https://www.gov.uk/government/organisations/civil-service/about

Clarke, A., Lindquist, E. A., & Roy, J. (2017). Understanding governance in the digital era: An agenda for public administration research in Canada. *Canadian Public Administration*, *60*(4), 457-475.

Coley, A., & Newman, A. (2016, February 1). Let's make Defra data driven. *Defra Digital Blog*. Retrieved April 26, 2018, from https://defradigital.blog.gov.uk/2016/02/01/lets-make-defra-data-driven/

Collier, D. (2011). Understanding process tracing. *PS: Political Science & Politics*, *44*(4), 823-830.

Daglio, M., Gerson, D., & Kitchen, H. (2015). *Building organisational capacity for public sector innovation*. Background paper prepared for the OECD Conference "Innovating the Public Sector: from Ideas to Impact". Paris: OECD.

Damanpour, F., & Schneider, M. (2009). Characteristics of Innovation and Innovation Adoption in Public Organizations: Assessing the Role of Managers. *Journal of Public Administration Research and Theory, 19*(3), 495-522.

Deloitte Digital. (2015). *Digital Government Transformation: UK survey data analysis*. Retrieved April 30, 2018, from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/public-sector/deloitte-uk-government-digital-transformation-findings.pdf

Deloitte. (2017). *Citizens, government and business: The State of the State 2017-2018*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/public-sector/deloitte-uk-the-state-of-the-state-report-2017.pdf

Denhardt, R. B., & Denhardt, J. V. (2000). The new public service: Serving rather than steering. *Public administration review*, *60*(6), 549-559.

Department for Environment, Food & Rural Affairs. (2017). *Annual Report and Accounts 2016-17*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/630164/defra-annual-report-2016-2017-web.pdf

Department for Work and Pensions. (2017). *Annual Report and Accounts 2016-2017*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/624319/dwp-annual-report-and-accounts-2016-2017.pdf

Department of Health. (2017). *Annual Reports and Accounts 2016-2017.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/629984/DH_annual_accounts_2016_2017_web_version.pdf

Digital, Data and Technology Profession. (2017a). *Collection: Digital, Data and Technology Profession Capability Framework*. Retrieved from https://www.gov.uk/government/collections/digital-data-and-technology-profession-capability-framework

Digital, Data and Technology Profession. (2017b). *Guidance: Head of performance analysis: skills they need.* Retrieved from https://www.gov.uk/government/publications/head-of-performance-analysis-skills-they-need/head-of-performance-analysis-skills-they-need

Dowdeswell, E., & Boyd, I. (2017, April 7). Why Defra has an Earth Observation Centre of Excellence. *Defra Digital Blog*. Retrieved April 26, 2018, from https://defradigital.blog.gov.uk/2017/04/07/why-defra-has-an-earth-observation-centre-of-excellence/

Dunleavy, P., & Hood, C. (1994). From old public administration to new public management. *Public money & management*, *14*(3), 9-16.

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead—long live digital-era governance. *Journal of public administration research and theory*, *16*(3), 467-494.

Ebrahim, Z., & Irani, Z. (2005). E-government adoption: architecture and barriers. *Business process management journal*, *11*(5), 589-611.

European Commission. (2016). *Big data analytics for policy making*. Retrieved from https://joinup.ec.europa.eu/document/big-data-analytics-policy-making-report

European Commission. (2017). *Good Practice Fiche – UK: Data mining tools and methods to tackle the hidden economy in the UK*. Retrieved April 27, 2018 from http://ec.europa.eu/social/BlobServlet?docId=18525&langId=en

European Commission. (2018). *Country Report United Kingdom 2018*. Retrieved from https://ec.europa.eu/info/sites/info/files/2018-european-semester-country-report-united-kingdom-en.pdf

Fichman, R. G., & Kemerer, C. F. (1997). The assimilation of software process innovations: An organizational learning perspective. *Management science*, *43*(10), 1345-1363.

Frambach, R. T., & Schillewaert, N. (2002). Organizational innovation adoption: A multi-level framework of determinants and opportunities for future research. *Journal of Business Research, 55*(2), 163-176.

Gerring, J. (2006). *Case Study Research: Principles and Practices*. Cambridge: Cambridge University

Giest, S. (2017). Big data for policymaking: fad or fasttrack?. *Policy Sciences*, *50*(3), 367-382.

Government Digital Service. (2018). *Guidance: GOV.UK Verify*. Retrieved May 3, 2018 from https://www.gov.uk/government/publications/introducing-govuk-verify/introducing-govuk-verify#govuk-verify-overview

Government Economic Service. (2013). *GES Professional Standards*. Retrieved from https://files.civilservicejobs.service.gov.uk/admin/fairs/apptrack/download.cgi?SID=b3duZXI9NTA3MDAwMCZvd25lcnR5cGU9ZmFpciZkb2NfdHlwZT12YWMmZG9jX2lkPTQxOTkyNSZ2ZXJpZnk9NjE3N2NjYmY0MzUwMGEzZjVmMDAxZTM5NzJkYjZlZTI=

Government Knowledge and Information Management. (2009). *GKIM Professional Skills Framework*. Retrieved from

http://www.nationalarchives.gov.uk/documents/information-management/gkim-skills-framework.pdf

Government Office for Science. (2016). *The Government Science and Engineering (GSE) Profession Strategy: Our Blueprint for the future of the GSE profession.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/557334/gs-16-17-government-science-engineering-gse-strategy.pdf

Government Operational Research Service. (2014). *Operational Research Competencies*. Retrieved from http://www.operational-research.gov.uk/recruitment/competencies/

Government Science and Engineering. (2013). *Science and Engineering Professional Framework*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/301275/science-engineering-skills-framework.pdf

Government Social Research Professions. (n.d.). *About us: The Government Social Research profession supports the development, implementation, review and evaluation of government policy*. Retrieved from https://www.gov.uk/government/organisations/civil-service-government-social-research-profession/about

Government Statistical Service. (n.d.). *The Government Statistician Group (GSG) Competency Framework*. Retrieved from https://gss.civilservice.gov.uk/wp-content/uploads/2016/03/GSG-Competency-Framework_March2016-1.pdf

Grimsley, M., & Meehan, A. (2007). e-Government information systems: Evaluation-led design for public value and client trust. *European Journal of Information Systems*, *16*(2), 134-148.

Hameed, A., Counsell, S., & Swift, S. (2012). A meta-analysis of relationships between organizational characteristics and IT innovation adoption in organizations. *Information & Management, 49*(5), 218-232.

Harvey, C. (2018). How to Select a Big Data Application. *Datamation*. Retrieved May 7, 2018, from https://www.datamation.com/big-data/how-to-select-a-big-data-application.html

Head, B. W. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration*, *67*(1), 1-11.

HM Revenue & Customs. (2014). *Annual Report and Accounts 2013-14*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/330670/HMRC-annual-report-2013-14.pdf

HM Revenue & Customs. (2015, July 17). *HMRC ICT: 1 January 2015 to 31 March 2015*. Retrieved April 25, 2018, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/443187/HMRC_exceptions_to_the_Government_spending_moratorium_1_Jan_2015_-_31_Mar_2015_-_ICT.csv/preview

HM Revenue & Customs. (2017). *Annual Report and Accounts 2016-17*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/635587/HMRC_Annual_Report_and_Accounts_2016-17_web_.pdf

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 147-169.

Hood, C. (1991). A public management for all seasons?. *Public administration*, *69*(1), 3-19.

Howlett, M. (2009). Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian public administration*, *52*(2), 153-175.

Howlett, M. (2015). Policy analytical capacity: The supply and demand for policy analysis in government. *Policy and Society*, *34*(3-4), 173-182.

Howlett, M., & Newman, J. (2010). Policy analysis and policy work in federal systems: Policy advice and its contribution to evidence-based policy-making in multi-level governance systems. *Policy and Society*, *29*(2), 123-136.

Hsu, A. (2015). Measuring policy analytical capacity for the environment: A case for engaging new actors. *Policy and Society, 34*(3-4), 197-208.

Institute for Government. (2017,August 08). *Specialisms in the civil service*. Retrieved April 27, 2018, from https://www.instituteforgovernment.org.uk/publication/whitehall-monitor/whitehall-explained/specialisms-civil-service

Institute for Government. (2018). *Accountability in modern government: what are the issues?* A discussion paper. Retrieved from https://www.instituteforgovernment.org.uk/sites/default/files/publications/IfG%20accountability%20discussion%20paper%20april%202018.pdf

Intelligence Analysis. (n.d.). *About us*. Retrieved from https://www.gov.uk/government/organisations/civil-service-intelligence-analysis-profession/about

Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish basic data program. *Government Information Quarterly*, *33*(1), 89-104.

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of theoretical and applied electronic commerce research*, *9*(2), 100-120.

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78-85.

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.

Lavertu, S. (2016). We all need help: "Big data" and the mismeasure of public administration. *Public administration review*, *76*(6), 864-872.

Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. *Government information quarterly*, *18*(2), 122-136.

Lee, & Xia. (2006). Organizational size and IT innovation adoption: A meta-analysis. *Information & Management, 43*(8), 975-985.

Lewis-Beck, M. S., Bryman, A., & Futing Liao, T. (2004). *The SAGE encyclopedia of social science research methods*. Thousand Oaks, CA: SAGE Publications Ltd

Manzoni, J. (2017). *Big data in government: The challenges and opportunities*. Speech. Retrieved May 3, 2018, from https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities

Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on the Web. *Philosophical Transactions of the Royal Society A,* 371, 1-17.

Margetts, H., & Sutcliffe, D. (2013). Addressing the policy challenges and opportunities of "Big data". *Policy & Internet*, *5*(2), 139-146.

Marks, G., Hooghe, L., & Blank, K. (1996). European integration from the 1980s: State-centric v. multi-level governance. *JCMS: Journal of Common Market Studies*, 34(3), 341-378.

Martin, A. (2006). A European framework for digital literacy. *Nordic Journal of Digital Literacy*, *1*(02), 151-161.

Mathieson, S. A. (2016, August 01). Hadoop starts to trumpet way through UK public sector. *Computer Weekly*. Retrieved April 25, 2018, from https://www.computerweekly.com/feature/Hadoop-starts-to-trumpet-way-through-UK-public-sector

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, *76*(6), 928-937.

Meyer, A. D., & Goes, J. B. (1988). Organizational assimilation of innovations: A multilevel contextual analysis. *Academy of management journal*, *31*(4), 897-923.

Moon, M. J. (2002). The evolution of e-government among municipalities: rhetoric or reality?. *Public administration review*, *62*(4), 424-433.

Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard university press.

Morabito, V. (2015). *Big data and analytics: Strategic and organizational impacts*. New York, NY: Springer.

National Audit Office. (2017). *Report: Capability in the civil service*. HC 919. Retrieved April 27, 2018, from https://www.nao.org.uk/wp-content/uploads/2017/03/Capability-in-the-civil-service.pdf

Neuman, W. (2013). *Social Research Methods: Qualitative and Quantitative Approaches*. Harlow: Pearson Education Limited.

NHS England. (2017). *Annual Report and Accounts 2016/17.* Retrieved from https://www.england.nhs.uk/wp-content/uploads/2017/07/Annual-Report-Full_201617.pdf

O'Malley, M. (2014). Doing what works: Governing in the age of big data. *Public administration review*, *74*(5), 555-556.

OECD, Eurostat & EU. (1997). *Proposed Guidelines for Collecting and Interpreting Technological Innovation: Data Oslo Manual*. (The Measurement of Scientific and Technological Activities). Paris: OECD Publishing.

OECD. (2017). *Government at a Glance 2017* (Country Factsheet: United Kingdom). Retrieved May 3, 2018 from: http://www.oecd.org/gov/gov-at-a-glance-2017-uk.pdf

Office for National Statistics. (2016). *User requested data: Civil service professions by government department 2016*. Retrieved April 27, 2018 from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/publicsectorpersonnel/adhocs/006316civilserviceprofessionsbygovernmentdepartment2016

O'Flynn, J. (2007). From New Public Management to Public Value: Paradigmatic Change and Managerial Implications. *Australian Journal of Public Administration*, 66(3), 353-366

O'Leary, Z. (2004). *The essential guide to doing research*. London: SAGE.

Open source data analysis prototype. (n.d.). Retrieved from http://www.mastodonc.com/casestudies/defra/

Policy Profession. (2018). *Policy Profession Standards: a framework for professional development*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/692329/PolicyProfessionStandardsJan18.pdf

Post, L. C. (2018, February 07). OracleVoice: Big Data Helps UK National Health Service Lower Costs, Improve Treatments. *Forbes*. Retrieved April 24, 2018, from https://www.forbes.com/sites/oracle/2018/02/07/big-data-helps-uk-national-health-service-lower-costs-improve-treatments/#1f554a5d58e3

Pudjianto, B., & Hangjung, Z. (2009). Factors Affecting E-Government Assimilation in Developing Countries. *SSRN Electronic Journal*, 1-14.

Punton, M., & Welle, K. (2015). *Applying Process Tracing in Five Steps*. Number 10 Annex. Center for Development Impact. Brighton: IDS

Purvis, R. L., Sambamurthy, V., & Zmud, R. W. (2001). The assimilation of knowledge platforms in organizations: An empirical investigation. *Organization science*, *12*(2), 117-135.

Ranganathan, C., Dhaliwal, J. S., & Teo, T. S. (2004). Assimilation and diffusion of web technologies in supply-chain management: an examination of key drivers and performance impacts. *International Journal of Electronic Commerce*, *9*(1), 127-161.

Reffat, R. (2003). Developing a successful e-government. In *Symposium on E-government: Opportunities and Challenge* (pp. 1-13). Muscat.

Richards, N. M., & King, J. H. (2014). Big Data Ethics. *Wake Forest Law Review*, *49*, 393.

Rogers, E. (1995). *Diffusion of innovations* (4th ed.). New York: The Free Press.

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

Saran, C. (2014, April 4). HMRC uses Hadoop to tackle corporate tax avoidance. *Computer Weekly*. Retrieved April 25, 2018, from https://www.computerweekly.com/news/2240217592/HMRC-uses-Hadoop-to-tackle-corporate-tax-avoidance

SAS. (2012). *Harnessing big data to mend our Broken Economy*. White Paper. *Retrieved from http://www.sourcingfocus.com/uploaded/documents/PubSector_HPA_BIGDATA_Final.pdf*

SAS. (2014). *SAS provides the DWP with powerful predictive insights in highly complex policy areas*. Retrieved April 25, 2018, from https://www.sas.com/content/dam/SAS/en_gb/doc/CustomerStories/dwp.pdf

Savoldelli, A., Codagnone, C., & Misuraca, G. (2014). Understanding the e-government paradox: Learning from literature and practice on barriers to adoption. *Government Information Quarterly*, *31*, S63-S71.

Schedler, K., Guenduez, A., & Frischknecht, R. (2017). How smart can government be? - Discussing the barriers to smart government adoption. In *IPMN Conference* (pp. 1-17). Shanghai Jiaotong University, China.

Sopra Steria. (2017). *Government Digital Trends Survey 2015-201*. Retrieved April 30, 2018, from https://www.soprasteria.co.uk/docs/librariesprovider41/White-Papers/sopra-steria-government-digital-trends-2015-2017.pdf?sfvrsn=2

Sumner, S. (2017, June 15). Mayank Prakash CIO interview - IT priorities at DWP. *Computing*. Retrieved April 25, 2018, from https://www.computing.co.uk/ctg/news/3011998/mayank-prakash-cio-interview-it-priorities-at-dwp

Tassabehji, R., Hackney, R., & Popovič, A. (2016). Emergent digital era governance: Enacting the role of the 'institutional entrepreneur' in transformational change. *Government Information Quarterly*, *33*(2), 223-236.

TechUK. (2017). *Smarter Services: delivering the next wave of digital transformation in the public sector*. Retrieved April 30, 2018 from https://www.techuk.org/images/smarter_services_DIGITAL_FINAL.pdf

Toshkov, D. (2016). *Political Analysis: Research Design in Political Science*. Basingstoke: Palgrave Macmillan.

United Nations and American Society for Public Administration (2001), *Benchmarking E-Government; A Global Perspective*, retrieved from: https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/English.pdf.

Van Veenstra, A. F., Klievink, B., & Janssen, M. (2011). Barriers and impediments to transformational government: insights from literature and practice. *Electronic Government, An International Journal*, *8*(2-3), 226-241.

Walker, R. (2008). An Empirical Evaluation of Innovation Types and Organizational and Environmental Characteristics: Towards a Configuration Framework. *Journal of Public Administration Research and Theory*, 18(4), 591-615.

Windrum, P., & Koch, P. M. (2008). *Innovation in public sector services: entrepreneurship, creativity and management*. Northampton: Edward Elgar Publishing.

Wu, X., Ramesh, M., & Howlett, M. (2015). Policy capacity: A conceptual framework for understanding policy competences and capabilities. *Policy and Society*, *34*(3-4), 165-171.

Yin, R. K. (2012). *Applications of Case Study Research* (Vol. 34). SAGE Publications.