Future government data strategies

Data-driven enterprise or data steward?: Exploring definitions and challenges for the government as data enterprise

Van Donge, W.; Bharosa, N.; Janssen, M. F.W.H.A.

DOI
10.1145/3396956.3396975

Publication date
2020

Document Version
Final published version

Published in
Proceedings of the 21st Annual International Conference on Digital Government Research

Future government data strategies: data-driven enterprise or data steward?

Exploring definitions and challenges for the government as data enterprise

W. van Donge
Delft University of Technology, Delft, The Netherlands
wendy@digicampus.tech

N. Bharosa
Delft University of Technology, Delft, The Netherlands
n.bharosa@tudelft.nl

M. F. W. H. A. Janssen
Delft University of Technology, Delft, The Netherlands
m.f.w.h.a.janssen@tudelft.nl

ABSTRACT

Comparable to the concept of a data(-driven) enterprise, the concept of a 'government as data (-driven) enterprise' is gaining popularity as a data strategy. However, what it implies is unclear. The objective of this paper is to clarify the concept of the government as data (-driven) enterprise, and identify the challenges and drivers that shape future data strategies. Drawing on literature review and expert interviews, this paper provides a rich understanding of the challenges for developing sound future government data strategies. Our analysis shows that two contrary data strategies dominate the debate. On the one hand is the data-driven enterprise strategy that focusses on collecting and using data to improve or enrich government processes and services (internal orientation). On the other hand, respondents point to the urgent need for governments to take on data stewardship, so other parties can use data to develop value for society (external orientation). Since these data strategies are not mutually exclusive, some government agencies will attempt to combine them, which is very difficult to pull off. Nonetheless, both strategies demand a more data minded culture. Moreover, the successful implementation of either strategy requires mature data governance – something most organisations still need to master. This research contributes by providing more depth to these strategies. The main challenge for policy makers is to decide on which strategy best fits their agency's roles and responsibilities and develop a shared roadmap with the external actors while at the same time mature on data governance.

CCS CONCEPTS

• **General and reference**; • **Document types**; • **General conference proceedings**;

KEYWORDS

Data-driven government, data stewardship, data enterprise, data governance, e-government

dg.o '20, June 15–19, 2020, Seoul, Republic of Korea

ACM ISBN 978-1-4503-8791-0/20/06…$15.00
https://doi.org/10.1145/3396956.3396975

ACM Reference Format:
W. van Donge, N. Bharosa, and M. F. W. H. A. Janssen. 2020. Future government data strategies: data-driven enterprise or data steward?: Exploring definitions and challenges for the government as data enterprise. In *The 21st Annual International Conference on Digital Government Research (dg.o '20), June 15–19, 2020, Seoul, Republic of Korea.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3396956.3396975

1 INTRODUCTION

In order to fulfil their tasks in societies, government agencies need to execute a strategy for collecting, storing, using, and sharing data for public services. A strategy can be defined as: "*a unified, comprehensive, and integrated plan, designed to ensure that the basic objectives of the enterprise are achieved*" [1]. There are at least five trends recognised in the literature that require public agencies to take a closer look at their roles and responsibilities, and redefine their data strategy. First is the rising level of expectations regarding public service delivery. Inspired by the high level of personalization and ease of use achieved by the famous Big Tech companies (i.e. Google, Amazon, Facebook, Baidu, Tencent), tech savvy citizens, politicians and policy makers expect better public services [2]. Second is the huge amount of data available to government agencies [3]. Over the past decades in which paper forms and processes have been transformed to digital databases and automated workflows, government agencies have become big data organisations [4, 5]. Third is the emergence of stricter laws and regulations for data collection, processing and sharing. For instance, the General Data Protection Act (GDPR) demands that data processing should only use as much data as is required to successfully accomplish a given task (data minimization principle) [6, 7]. Additionally, data collected for one purpose cannot be repurposed without explicit consent (purpose limitation). Fourth is that government agencies are expected to become more efficient and spend less taxpayer money [8]. As public service delivery is a significant part of a government's budget, it becomes increasingly important to find ways to provide public services at lower cost. Finally, the rise of GovTech is going to have a major impact on public services [9]. GovTech generally refers to the landscape of private-sector start-ups and small and medium enterprises (SMEs) that deliver technological products and services that can be used in public services [9]. Examples include services such as providing high level of assurance eIDs for authentication and electronic document signing, apps that help citizens to apply for public housing and apps that help people with debts with financial planning . For the accurate and convenient use of such apps, private tech providers (often start-ups) need access to

government services (e.g. citizen number authentication) and data sources (e.g. citizen registry, income registry).

Against this backdrop, public agencies must rethink their data strategies. One of the key concepts that surfaces in discussions and policy agenda's is the concept of data-driven government. Similar to data enterprises, becoming more data-driven seems to be a leading data strategy for government agencies. Since literature on this concept is scarce [10]–[12], we lack a more in depth understanding of the characteristics of a government as data enterprise. The objective of this research paper is to clarify what it means to be data(-driven) government and identify which challenges are on the horizon. Therefore, this paper seeks answer to the following question: what are the characteristics of a data(-driven) government and what are the challenges associated with it?

This paper proceeds as follows. Section two presents the research approach which builds on literature review and semi-structured expert interviews. Section three presents the findings of the literature review on data-driven government and related concepts, allowing for the development of a semi-structured interview protocol. Section four reveals the main findings from the interviews. This paper concludes with a discussion on the main insights and avenues for further research.

2 RESEARCH APPROACH

To unravel the concept of data (-driven) government and to identify the challenges, we employ two research instruments: (1) systematic literature review and (2) expert interviews. Both are discussed next.

2.1 Systematic literature review

A systematic literature review was performed on data(-driven) government and enterprise. The goal is to define the concept of a data(-driven) government more in depth by compiling a list of the defining characteristics. In general, literature reviews yield an overview, synthesis, and a critical assessment of previous research, to identify knowledge gaps or construct novel research problems and directions [13]. Systematic literature review is different from a traditional general review as it follows a more replicable, scientific, and transparent process [14]. We performed a systematic literature review by searching on keywords using Scopus. This database is now widely used and includes most major peer-reviewed articles from all the academic fields [15]. This search was performed mid November 2019. Only academic journals and conferences were included in the search, we did not review non-academic reports.

A set of closely related keywords were used as search terms. All keywords were searched using quotation marks to retrieve precise results during the search. The search terms were: "data-driven government", "data-driven organisation", "data-driven enterprise", "data oriented government", "data minded government", "government as a data enterprise" and "government data strategies". The search results are shown in Table 1. The second column presents the total number of hits, the third column presents the hits after screening titles, keywords, and abstracts. Results show that there is more attention for the data-driven organization than for the data-driven government.

Table 1: – Search results using Scopus

Keyword	Total hits	After screening
Data-driven government	5	5
Data-driven organization	55	31
Data-driven enterprise	24	13
Data enterprise	77	29
Data oriented government	0	0
Data minded government	0	0
Government as a data enterprise	0	0
Government data strategies	0	0

During our search, the search term "data-driven government" gave only five hits on Scopus. This shows a current lack of research particularly focusing on data-driven government, although there exists research on concepts that have similarities with government data (e.g. open data and big data). Therefore, research with other search terms were found to be needed and the research was broadened to data-driven organization and enterprise. Each article was first screened by reading its title, abstract and keywords. The main inclusion criteria are: (1) focus on data collection, analysis, exchange, or processing within or between organizations and (2) related to a public-service or -task. After following the screening process, we finally included 78 papers to review. While reviewing the papers, we identified some recurring characteristics. These are presented in section three.

2.2 Expert interviews

2.2.1 Method. Given the scarcity of research on data-driven government (see table 1), it was necessary to work inductively and collect insights from practice. Therefore, we wanted to interview experts in the field of data-driven government. Since data-driven government is a new concept, we expected that experts will have different ideas of what it means to be data-driven. Moreover, they might have more accurate concepts/terms and examples that could capture the essence of a government data strategy in more precisely. In order to collect these concepts, the interviews needed to be as open as possible [16]–[18]. Therefore, we chose to perform one-on-one semi-structured interviews. This method provides a loose and flexible interview which allowed to have a dialogue [19] with the respondents during the interviews. This approach enables respondents to reflect on their own experiences allowing new concepts to emerge [16, 20].

2.2.2 Sampling. The following criteria guided the respondent selection process:

- the respondent should be actively involved in analysing or designing government data strategies/architectures/policies.
- the sample should cover various layers of government: central and local.
- the sample should include multiple disciplines: public administration, policy development, law, ethics, information technologies.

- the respondents should be from both the public as the private sector, but focused on the public sector.
- should contain a variety of demographics in terms of gender.

Based on the criteria listed above twelve respondents were invited for an interview. From these twelve, eleven responded that they were open to an interview. Only eight interviews were completed when submitting this draft of this paper. An overview of the respondents and the interview protocol is available om request.

2.2.3 Interview questions. Each interview questions started with the definition of a 'data enterprise'. Next we discussed the relation between government and data. Here we explored in what way the definition of a data enterprise is applicable to the government. Next we explored the challenges and barriers, and concluded with a vision towards the future. The questions of the interviews were open ended since the aim of the interviews was to get spontaneous and in-depth responses [16, 17, 20]. The interview protocol was tested with two people and had proven to give meaningful data on the research questions. It also provided the flexibility to really explore the subject of government and data. During these tests we did find that the concept of 'data-driven enterprise' was too narrow, and it was better to use the concept of 'data enterprise'.

We asked the experts to define the concept 'data enterprise', talk about their experiences with data and the government, explain their views of the government as data enterprise, share their visions on the future of data and the government and elaborate on the challenges governments face. We asked them to relate to their own experiences. The interviews were audiotaped, transcribed and send to the respondent to be validated. Two respondents have elaborated and given more details as a response to the validation email. The interview questions can be requested at the authors.

2.2.4 Analysis. The interview transcripts are compiled using Microsoft Word and are available on request. Based on the three components of the interview protocol – definition, drivers and challenges – we created an overview of the answers given by respondents. The answers were compared across respondents and multiple groups were developed. Section four presents the resulting groups of answers. This is the first step of the data analysis process, further research will focus on using quantitative coding techniques using Atlas.TI.

3 LITERATURE REVIEW

The goal of the literature review is to define the concept of a data-driven government more in depth, i.e. list the defining characteristics. While the concept of a government as data enterprise, or data-driven government, is increasingly being used in practice, our literature review found that we still lack sound definitions in literature. As depicted in table 1, there is a small body of work on data-driven organisations. For instance, Patil states [21] : "*A data-driven organization acquires, processes, and leverages data in a timely fashion to create efficiencies* [21]*, iterate on and develop new products, and navigate the competitive landscape.*" We found similar definitions on data-driven organisations in other papers that underline the focus on competitive advantage [22]. Yet, the notion of competitive advantage is not very useful for government agencies that – by law – have a monopoly over specific public services. When we dig deeper, on variables in data-driven organisations, there is an emphasis on data processes, organisation, decision making, and culture. Table 2 outlines the characteristics found in literature.

Table 2: – Characteristics

	Characteristics of data-driven organisations	Source
1	Have a managed and aligned architecture of processes, services, tools and roles that govern the organisation.	[23]
2	Have a strong data culture (also known as a data minded culture).	[24]
3	Use data as the basis for operational (task specific) decision-making, as well as tactical strategic decision making.	[10]
4	Have a mature data governance process.	[24]
5	Continuously search for data quality improvements.	[25]
6	Depend on both internal and external data sources (open data).	[26]
7	Use a well-defined set of data-metrics to monitor overall organisational performance.	[27]
8	Emphasize value creation based on data for all actors: management, employees, clients and vendors.	[27]
9	Have tooling and expertise for business intelligence and data analytics (qualitative and quantitative).	[27]
10	Have an explicit organisation structure with chief data executives (e.g. chief information or data officer), data stewards and data scientist.	[27]

The characteristics listed in Table 2 are useful for defining a data-driven organisation (or data-driven enterprise, we use these terms interchangeably). Essential characteristics are having a mature data process, adding value to data, a data minded view to the data process, which is all about using it to its full potential, and finally using the outcome for decision-making. The question is to what extend they also characterise a data(-driven) government. The following sections elaborate on this question.

4 INTERVIEW FINDINGS

In this section we explore the extent to which a government can be described as a data enterprise. We see similarities and differences with the concept of data enterprise as found in the literature and the outcomes of the interviews. It is difficult to define a data enterprise, and it is even more difficult to define a government as data enterprise because of the many variables that come into play. In

the next paragraph we elaborate on the characteristics of a data enterprise as stated by respondents, and look into what extent a typical government agency fits these characteristics. Next we elaborate on the drivers observed by experts for government agencies to transform into a data enterprise. We close with the challenges for this transformation.

4.1 Definitions

When defining the data enterprise, the interviews reveal multiple similarities and differences compared to the literature. The respondents agree that value creation through data and a data minded culture are the defining characteristics of a data enterprise. The main difference is a distinction in the use of the data.

Some respondents stated that a data enterprise revolves around data – data is the core business. It is not only an asset to their business process, it is the main objective of the business to gather, store, exploit or process data. This is the type of data enterprise which performs (parts of) the data process on behalf of others. The data enterprise earns money by gathering, processing, storing, and sharing data. On the other hand, some respondents mentioned a data enterprise is an organisation which exploits data to improve their own business, where data is just a means to an end. For example a car company which uses data to improve its product, sales, or production line. In this dichotomy the difference can be seen with the characteristics from the literature, where the focus is on mature data processes and data exploitation within an organisation. Whereas interviewees, noticed a possible separation in the parts of the data process and the final use. Based on the interviews and literature we cannot distil one definition of a data-enterprise. We can however identify four characteristics of a government agency as data enterprise. The first characteristic is to have a data process in order to exploit data for decision making. One of the respondents stated the difference between the 'Google kind of companies' and the government, is that data for a government is not at the centre of their business model. However other respondents pointed out that being data-driven is very important within the government, and maybe the government is one of the most data-driven organisations, next to banks, insurance and pension funds. As one interviewee stated: "the primary task of the government is to deliver public value, which are mostly services. These services are established because of data. When you submit for a driver's licence, there are many registries where your data is subtracted, and this results in a personal product.".

Since there are over 400 governmental organisations in the Netherlands, there are differences between them. There are organizations which use data to optimize the internal processes and services. Also, there are organizations which use the data to create new services and processes, and try to derive these out of the data. And finally, there are organisations which have parts of the data process as the core of their business.

The second characteristic is having data as a business unit, being part of the data process. A respondent noticed The Central Bureau of Statistics (CBS), a public organisation which collects and publishes statistics on the Dutch society. Their main product is data. However there are also organisations which comply to both the first and second characteristics. A respondents defined Rijkswaterstaat, a public organisations responsible for management and development of main roads, waterways, and water-systems, as a data enterprise. According to the respondent the only way Rijkswaterstaat can do their task correctly is by having a digital copy of reality and from this direct the policy and determine decisions. But beside using data for their own processes and decisions, it also shares this data with other (public) organisations for them to use. Making it both input for their own process, as unit for others process.

The third characteristic is the creation of value. An example of adding value is the partnership with ABN Amro, a Dutch Bank, and the University of Amsterdam, on the detection of human trafficking. Some indicators were determined, and with these seven effective cases of human trafficking were detected. This is a collaboration between the private sector, a government agency and a university which creates public value based on data. Here all parties contributed with data, in order to provide a public service.

The fourth characteristics of data enterprises is being data minded. One of the respondents noticed it is hip and happening to look at data-minded companies like Google and Uber. Since the government has a lot of data too, you might wonder whether a government should behave like a big data enterprise. However there is a big difference between the vision, strategy and their target. The objective for a company is to exists and earn money. For a government the objectives are centred around regulating society and the open and fair provision of social services. Even though many respondents noticed a government should not become exactly Google or Uber, a government can become more data minded. To become data minded one has to "look in a positive way to the opportunities of data and be able to translate this to the organisation. This asks for leadership, steering and vision." Government agencies are gaining interest in data and digitalisation, even though this interest is still low.

4.2 Drivers for formulating new data strategies

When talking about the current situation of the government and data, all respondents highlight the potential of data, and how the government is not using it optimal at this point. When analysing the interviews, there are several drivers of change which ask for a redefinition of the data strategy of the government.

The first driver observed is that the society which becomes more self-serving, either by the forces of the private market, or by citizen collaboration. A respondent found that: "There are many initiatives resulting from collaborations in society to take over roles of the government. Where people themselves or companies provide some services which the government otherwise would have done. The market therefore asks for data from the government to provide in opportunities they see in society."

Besides adding value to society, government data also has value for the private sector. One respondent stated that open data is crucial for the private sector. "Government data has a huge financial value and as the Dutch government we should strive the be as open as possible". Another respondent noticed that companies rather have raw (unprocessed) data from the government, instead of analysed reports. An appeal is made to the government for becoming a data steward, referring to the overall responsibility for the data lifecycle in public service delivery.

Not only the market and society collaborate to create value. The government is more and more involved in public-private initiatives. As stated by one of the respondents: "When looking at how the government can get smarter by using data, you see a movement in the last decennia from more reactive and hierarchical steering, to the opposite, referring to an organisation which is more proactive and steers via partnerships." Many respondents stated that the government realizes it cannot do it by itself, and more collaboration is sought out.

The third driver is that people would like to gain more insight or control over their personal data. This request can first be traced to the fact that information is registered at many places. A respondent noticed it is not clear to many where all the personal data is stored. A second reason might be the fact that citizens require data from the government, like extracts from the Chamber of Commerce . A respondent argued that the hard line between public and private is vanishing. If a citizen needs data or a service it does not matter to him what the source of that data or service is. There is a risk; "Giving citizens the perception they have control over their data can be misleading". In cases where the government has a monopoly on using force, legal enforcement, tax collection, data is necessary in order to perform their tasks. In the end, people cannot control which data is known to the government. The respondent therefore stated that it is not about ownership of the data, but responsibility. The one stewarding the data has to comply with certain laws, to make sure the object of the data is not harmed.

These first three drivers ask the government to take a more active role in becoming a data steward. However, the fourth driver has a focus on the government exploiting data itself. With the expanding amount of data, companies like Google, Facebook and Uber, can provide personalized services. Since it is possible in the private sector, the expectations of citizens rise to also get personalized services provided by the public sector. A respondent stated: "we have much higher expectations. Everyone has a smartphone, with which you can do everything, have all digital services. These expectations transfer to the government and its service provision. Why should I go to a desk to identify myself? In fact, why do I have to have a physical driver's license, when I can have it with me on my phone?."

This has a strong connection to the issue of personal data control discussed previously. There is a tension between both wanting to control your 'own' data and expecting services of the government. One respondent however stated: "as a government you want to make use of technology. However, your position is different. As a tax office, you must perform certain tasks. We cannot always think of what the client wants. People must provide data, so the inspector can collect taxes. What we can do, is use this data to its extent, so these clients are as much helped, for example in filling out their tax forms. (..) We can design the process, so it fits as much as possible with the person's needs. However, the focus still stays at performing the legal tasks".

The government itself also uses data to both internally optimize their processes and to provide several services to society. As respondents stated, we all agreed to a social contract with the government. Where the government has the right, and the task, to use data in order to perform their public tasks. The government has a lot of data, and since technology is developed to a certain degree, we can

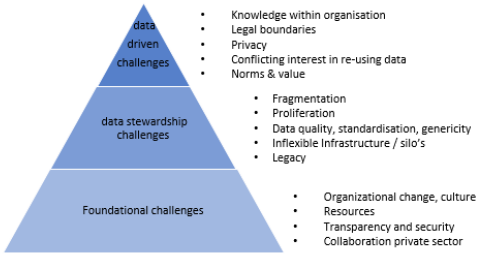

Figure 1: – Challenges for rolling out government data strategies

do a lot with data. The government therefore has a lot of opportunities, and we see a trend in experimenting in the use of this data. A respondent summarizes this as: "There is a lot of experimentation within the government. Everyone has a data lab, sometimes in collaborating parties. There is experimented, however the level of experimentation is not very high, not of the artificial intelligence application where people often talk about."

However not everyone is convinced it is the task of the government to experiment with data. Respondents stated that even though the government has a lot of data, and the possibilities are unlimited, the question remains whether it is the task of the government to process this data. In their opinion the government should rethink what the responsibilities and tasks of the government are, and what is necessary to carry out these tasks and responsibilities. When the market is able to provide services to 95% of the population, the government's task is to take care of the remaining 5% that is not or cannot be catered by the private sector, create a safety net for the cases where it goes wrong, initiate initiatives which the market does not start, set the standards, and regulate the licenses to produce such services.

4.3 Challenges

The drivers shape a reality in which governments must rethink their data strategy and redefine the roles they take within the data process. First, we define the challenges which are specific for the data stewardship strategy. Then we elaborate on specific problems for data exploitation. Finally there are some overarching challenges which the respondents identified. The next sections discuss the challenges more in depth.

4.3.1 Foundational challenges. Common problems within governmental organisations are realizing change, resource allocation (both financial and capacity of employees), transparency on data usage, and organising collaboration with the private sector.

The first challenge is to get the organisation to change its culture and processes. As respondents stated, if an organisation wants to implement a new strategy, the way of working must change. In order to have a data minded culture, people have to look at data differently to grasp its full potential. Also a respondent stated: "When looking at change-theory, when dealing with change, there must be a burning platform. Within the government this is less present. In my opinion a critical mass must be reached to push

change through.". When one wants to change the government into a data-minded organisation, a large part of the organisation has to be behind this change. It can be a challenge for the data minded people calling for change, to create this burning platform. According to a respondent it is all about mandate, the urge to score points, and the willingness to succeed. This is a cultural aspect which will be hard to change.

Besides the creation of a critical mass, organisations must also facilitate the opportunity for people to change the way they work. Here, the challenge the government face is resources. Respondent: "You have to provide data, the right quality of data, standardized data, people with knowledge who know what they are talking about and have people to perform these tasks." Additionally, a respondent stated that at one point in time there will be the moment you will implement the tools and technologies currently available. In order to do so one must map the current state and sort this system out, before a change can be implemented. This is a lot of work, and often comes with high costs. And even though it might be an improvement in technological development, the added value might not be as high for the organization. The costs always precede the benefits, making it an investment without exactly knowing the outcome. Besides, the benefits of having an organized data management system in place and exploiting data to the fullest, does not always lie with the governmental parties involved in these changes, rendering any data strategy into a hard sell.

A respondent stated that the transparency and control of data – which is fundamental for government agencies - can decrease when letting the market take over public services. An example of this statement is the Ockto-app – a service that enables you to log-in with your government ID to onto the MyGoV portal called "MijnOverheid". The app allows you to give data obtainable via the "MijnOverheid" portal to third parties, for example a bank or insurance company. The question is to what extent the user of the app really sees what happens in the app, and for which purposes. The app would say you have given your consent to the app to collect and share data, but as multiple respondents noted you sometimes hit the consent button without reading the terms and conditions.

Even though respondents noticed there is more collaboration between government agencies and the private sector, one respondent stated: "It is very difficult to work with the government if you do not have a large network of existing contacts." To compete in large tenders is difficult for small start-ups. The overhead is large and the chance of winning from a large company is small. However as other respondents noticed the government can learn a lot from start-ups and organisations which are successful in the field of data. Therefore it will be a challenge to involve the private sector through classical tendering schemes.

4.3.2 Additional challenges of data stewardship. The main characteristics of data stewardship are to assign responsibility over the governed data, collect and document meta-data (definitions, business rules, etc), and manage data in the best interest of all, in order to improve the quality of data. However, there are several problems the government faces with following the data stewardship strategy.

As one respondent stated data is scattered all across the government. Whereas the national registers (e.g. with citizen data, company data, income data, building data and so on) were meant to be singular data sources, many organisations still download a duplicate of the entire data set every day in order to work with the data. These data systems are not yet equipped to work efficiently. As respondents noticed we are nowhere near using a system of singular data registry, that facilitates data reuse. As another respondent stated, there are a lot of registries, but to what extend are they necessary? The government must look into whether this data is necessary to be registered at the government, and think about how to reuse data, also from other sources.

The forms of data proliferation can be seen in the silo's which have grown within the government. "I see silos everywhere, when talking about data sharing. There is no common infrastructure. Within the government there is a system of registration, which on its own is a nice system. Only, how do you connect with other systems? Then again you end up with silos". One respondent also explained how everything now is arranged for the tasks and responsibilities of organisations, instead of looking at a person with a need. Here the respondent sees an opportunity. The citizen does not see a difference between public and private, when they want to arrange their lives. As a government we can reduce this difference even more for the citizen. In doing so, we do not only have to think on a governmental architecture, but an architecture for the entire data system, where data is shared. In all these cases there is a technical challenge, which respondents reported. The data which is used needs to be of a high quality. The government should provide standardisation and have general purpose data usage in mind. Not only within the government, but also towards the outside world. For example, according to one respondent most government data is shared in PDF. However, this is difficult standard when you want to combine and analyse data, making is more difficult for others to use.

The fourth challenge for data stewardship is legacy. One respondent noticed that for one of the information chains the architecture was developed in 2004/2005, making use of what was possible then. It was sufficient for then, but even though the technologies have developed rapidly since then, we are still dealing with an outdated architecture. To deal with this one could say, let us start over. However, as a respondent noticed: "Sometimes you would like to start all over, just like Estonia did. If you could do it all over, it becomes less difficult. However, in the end Estonia again created legacy within their new systems". Hence, whatever path the government decides on, it will always face the limits of previous choices. It will be a challenge to overcome the past legacy and deal with the future legacy.

4.3.3 Additional challenges of being data-driven. Besides arranging data stewardship, the government also faces several problems with the exploitation of data. Respondents stated that the government has difficulties with implementing technology. "*In practice, it is difficult to invest in AI, machine learning and data science. I notice that organisations are quickly inclined into using hardware and large scaled infrastructure, or on a team which after several years still delivers no value.*"

According to a respondent the amount of experiments exploiting data within the government is growing. However, the respondent also stated that there is a lot of window dressing. Many innovative strategies do not make it beyond experiments and proof of concepts.

"The true challenge is incorporating the innovation with data into the organisation." The government needs to learn, not only from experimenting by itself, but also to share the knowledge with and learning from other (governmental) organisations. " Everyone is struggling, and there is no answer ready yet to how the government can be smart with data. Its starts with sharing knowledge and a vision on what data can mean for the government".

Most respondents noticed the challenge of having knowledge within organisations. In the beginning of the computer at work, when you wanted to analyse you had to decide of the analyse you wanted to do with it, and sent it to a analytics centre with a large computer to wait a day and get the result of the analyse. This meant you worked a lot with the data and understood the data. You had the time to keep a logbook and keep the data quality in order. Now the people using the data are not the ones who have worked with it for years. To create information and draw conclusions prior knowledge is often needed.

Not only should the government deal with implementing innovation, sometimes it turns out some innovation is not possible due to legal boundaries. According to one of the respondents government often find out afterwards what is possible with data, and what is not. Therefore they cannot always use the outcome of data. A another respondent pointed out that a government has stricter legal boundaries to which it must comply, when using data, due to the rule of law and principle of legality. As one of the respondent stated: "It is a matter of power which the government has, which you want to keep in check. Consequently, data usage must be kept within legal boundaries." The government is bounded in the use of data and can only use it to perform predefined tasks.

The next challenge of being data-driven is that the optimization of the (re)use of data can lead to conflicting values. As a respondents stated; "On the one hand you want to protect civilians, on the other hand you want to collect as much data to analyse your services, detect fraud, optimize policy etc." Not only do you want a lot of data when being data-driven, this data also has to be of high quality and truthful, in order to rely on the outcome. This can create a challenge of re-using data. A respondent gave the example of data collection done by the government on the state of the health- and safety department (ARBO) within a company. When companies finds out data might be reused for other data analysis purposes, like penalizing when the ARBO is not arranged properly, companies may have a tendency to portray themselves more positively, affecting the quality of the data.

On the other hand respondents noticed the possibilities of data are not always known in advance and the ethics and privacy of data, is not a reason to kill all well-meant initiatives, which is the easiest thing to do. Respondents stated, that working with data, people should develop a feeling about when an application is truly radical for society. A moment of reflection should be incorporated in the data exploitation process to reflect on the meaning of the use of the data source for that particular purpose, and think of the possible side effects of using the data, in order to gain grip on the situation.

The last challenge with being a data-driven government is the matter of using AI. With AI you try to make models as fit as possible, however this delivers some tension with the norms and values of the government. As one respondent stated, the government cannot make mistakes. When a company can make correct decisions for 99% of the cases this would be a great result. However, if the government would make mistakes 1% of the time, it still is about people is unacceptable. On the other hand, a respondent noticed: "We live in a constitutional state. When we have AI based applications like SyRI, and we do not agree anymore, or we feel discriminated, you go to the judge which checks whether AI made the the correct decisions or not. We have that covered."

5 DISCUSSION

This paper examines the definition of a data enterprise and to what extent the government can be seen as a data enterprise. Here we found a dichotomy in the concept. The first is a data-driven organisation, where data is used within the organisation to optimize processes and services or create new services. This is the data driven government strategy. Based on the literature review and interviews, we present the following tentative definition for a data driven government: *a (semi-)public agency that uses external and internal data for process optimisation and public service delivery.* We pose that a data-driven government agency must satisfy the following characteristics:

- Uses data as the basis for operational (task specific) decision-making, as well as tactical strategic decision making.
- Uses a well-defined set of data-metrics to monitor overall organisational performance.
- Indefinitely puts value creation for all actors based on data on its policy agenda.

The second strategy is data stewardship, where an organisation acknowledges that it only plays one part in the entire data ecosystem. The role of data steward is to make sure the data is set to certain standards, is of high quality, is up to date, can be accessed by those with the rights to access, and shared when needed, in order to add value to the data chain and giver others the possibility to exploit data to its fullest potential. Based on the literature review and interviews, we present the following tentative definition for government as a data steward: a (semi-) public agency that has an explicit data responsibility and a continuous focus on data quality improvement, that allows external (private) parties to access (personal) data based *on predefined conditions.* We pose that a data-driven government agency must satisfy the following characteristics:

- Formalizes responsibility over data.
- Formulates explicit data sharing policies.
- Uses specifications for external data access (for instance via REST APIs or dataset download buttons).
- Continuously strives for data quality improvements.
- Stimulates the use of data, within legal boundaries.

To implement both strategies successfully, governments must have a mature data governance and need to become data minded. The following characteristics apply both for data driven governments and governments as data stewards:

- Has an explicit data governance structure with chief data executives (e.g. chief information or data officer), data stewards and data scientist.
- Has an explicit architecture of processes, services, tools and roles that govern the organisation.

- Cultivates a strong data culture (also known as a data minded culture).

In both strategies we see a shift from a process-oriented mind-set to a society-oriented mind-set. Not only is data used to optimize (internal) processes and systems, the focus also comes to lie on value delivery to society. The monopoly of the government as the sole public service provider is diminishing and the private sector is entering the public services market.

6 LIMITATIONS AND FURTHER RESEARCH

6.1 Limitations

This research has three main limitations. First is the geographic limitation: the experts consulted in the interviews are from the Netherlands and provide insights from the Dutch context. Second is the small number of interviews (total of eight). Even though we found some level of saturation (i.e. requiring observations, ideas and examples), we feel like there are more ideas and examples of data driven governments out there. Third is the explorative nature of this paper in general. While we conclude this paper with tentative definitions for data driven government and the government as data enterprise, we have not tested or validated these definitions in the field. All three limitations provide directions for future research.

6.2 Further research

Putting aside the directions for future research based on the limitations of this paper, we want to highlight three new avenues for future research. First is the development of a conceptual model for both data strategies: data driven government and the government as data enterprise. A conceptual model should help to identify and substantiate the various variables and relationships for each strategy, allowing for more focused theory developments. Second is an empirical comparison of the implementation of both data strategies. As revealed in the interviews, several public agencies are already working on implementing these strategies, which opens the path for in depth and comparative case studies. It can be very interesting for policy makers if such a study identifies the practical ramifications of the different strategies. Third, and perhaps most complex from a research perspective, is an international comparison of the substantiation and execution of two data strategies. The complexity here lies in the different cultural and institutional conditions across various countries. The configuration of institutions will determine the options for strategy execution. In order to compare across countries, researchers need to first analyse and map the cultural and institutional differences for the respective countries. An option is to use the institutional analysis framework provided by Williamson [28]. Researchers could for instance focus on an international comparison between European countries, especially since more and more data exchange legalisation is being harmonized across European countries (e.g. the general data protection regulation).

ACKNOWLEDGMENTS

This research is supported by Digicampus – a quadruple helix innovation hub in the Netherlands that focusses on developing future public services. The authors acknowledge Digicampus for funding this research, as well as providing access to experts for the interviews.

REFERENCES

[1] H. Mintzberg, "The Strategy Concept I: Five Ps For Strategy." California Management Review. Volume: 30 issue: 1, page(s): 11-24
[2] "UN E-Government Survey 2018." [Online]. Available: https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2018. [Accessed: 30-Dec-2019].
[3] G. H. Kim, S. Trimi, and J. H. Chung, "Big-data applications in the government sector," Communications of the ACM, vol. 57, no. 3, pp. 78–85, 2014, doi: 10.1145/2500873.
[4] B. Klievink, B. J. Romijn, S. Cunningham, and H. de Bruijn, "Big data in the public sector: Uncertainties and readiness," Information Systems Frontiers, vol. 19, no. 2, pp. 267–283, Apr. 2017, doi: 10.1007/s10796-016-9686-2.
[5] M. Janssen and G. Kuk, "Big and Open Linked Data (BOLD) in research, policy, and practice," Journal of Organizational Computing and Electronic Commerce, vol. 26, no. 1–2, pp. 3–13, Apr. 2016, doi: 10.1080/10919392.2015.1124005.
[6] B. van Loenen, S. Kulk, and H. Ploeger, "Data protection legislation: A very hungry caterpillar. The case of mapping data in the European Union," Government Information Quarterly, vol. 33, no. 2, pp. 338–345, Apr. 2016, doi: 10.1016/j.giq.2016.04.002.
[7] European Union, "Regulation 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data," Official Journal of the European Union, vol. L 8/1, no. November 2000. p. 22, 2001.
[8] M. Janssen and E. Estevez, "Lean government and platform-based governance—{Doing} more with less," Government Information Quarterly, vol. 30, pp. S1–S8, 2013.
[9] T. Filer, "Thinking about GovTech A Brief Guide for Policymakers."
[10] A. Luthfi and M. Janssen, "Open data for evidence-based decision-making: Data-driven government resulting in uncertainty and polarization," International Journal on Advanced Science, Engineering and Information Technology, vol. 9, no. 3, pp. 1071–1078, 2019, doi: 10.18517/ijaseit.9.3.8846.
[11] S. Mouzakitis *et al.*, "Challenges and opportunities in renovating public sector information by enabling linked data and analytics," Information Systems Frontiers, vol. 19, no. 2, pp. 321–336, Apr. 2017, doi: 10.1007/s10796-016-9687-1.
[12] "Driving public sector innovation using big and open linked data (BOLD) | SpringerLink." [Online]. Available: https://link.springer.com/article/10.1007/s10796-017-9746-2. [Accessed: 09-Jan-2020].
[13] D. Finfgeld-Connett and E. D. Johnson, "Literature search strategies for conducting knowledge-building and theory-generating qualitative systematic reviews," Journal of Advanced Nursing, vol. 69, no. 1, pp. 194–204, Jan. 2013, doi: 10.1111/j.1365-2648.2012.06037.x.
[14] S. K. Boell and D. Cecez-Kecmanovic, "On being 'systematic' in literature reviews in IS," Journal of Information Technology, vol. 30, no. 2. Palgrave Macmillan Ltd., pp. 161–173, 28-Jun-2015, doi: 10.1057/jit.2014.26.
[15] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, and E. Delgado López-Cózar, "Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories," Journal of Informetrics, vol. 12, no. 4, pp. 1160–1177, Nov. 2018, doi: 10.1016/j.joi.2018.09.002.
[16] C. Dearnley, "A reflection on the use of semi-structured interviews," Nurse Researcher, vol. 13, no. 1, pp. 19–28, 2005, doi: 10.7748/nr2005.07.13.1.19.c5997.
[17] D. W. Turner, "The Qualitative Report Qualitative Interview Design: A Practical Guide for Novice Investigators."
[18] P. Åstedt-Kurki and R.-L. Heikkinen, "Two approaches to the study of experiences of health and old age: the thematic interview and the narrative method," Journal of Advanced Nursing, vol. 20, pp. 418–421, 1994, doi: 10.1111/j.1365-2648.1994.tb02375.x.
[19] L. S. Whiting, "Semi-structured interviews: guidance for novice researchers.," Nursing Standard, vol. 22, no. 23, pp. 35–40, 2008.
[20] S. Krauss, A. Hamzah, Z. Omar, T. Suandi, I. A. Ismail, and M. Z. Zahari, "Preliminary Investigation and Interview Guide Development for Studying how Malaysian Farmers Form their Mental Models of Farming," 2009.
[21] M. Berndtsson, D. Forsberg, D. Stein, and T. Svahn, "Becoming a data-driven organisation," in 26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, ECIS 2018, 2018, doi: 10.1007/978-3-662-60304-8.
[22] M. Berndtsson, D. Forsberg, D. Stein, and T. Svahn, "Becoming a data-driven organisation," in 26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, ECIS 2018, 2018, doi: 10.1007/978-3-662-60304-8.
[23] I. Alhassan, D. Sammon, and M. Daly, "Data governance activities: an analysis of the literature," Journal of Decision Systems, vol. 25, no. sup1, pp. 64–75, Jun. 2016, doi: 10.1080/12460125.2016.1187397.

[24] B. Heinrich, D. Hristova, M. Klier, A. Schiller, and M. Szubartowicz, "Requirements for Data Quality Metrics," Journal of Data and Information Quality, vol. 9, no. 2, pp. 1–32, Jan. 2018, doi: 10.1145/3148238.
[25] R. Vidgen, S. Shaw, and D. B. Grant, "Management challenges in creating value from business analytics," European Journal of Operational Research, vol. 261, no. 2, pp. 626–639, Sep. 2017, doi: 10.1016/j.ejor.2017.02.023.
[26] E. Masa, P. Busch, G. Guzman, and L. Sanzogni, "Qualitative analysis of big data analytics in the oil & gas industry," in Proceedings of the 31st International Business Information Management Association Conference, IBIMA 2018: Innovation Management and Education Excellence through Vision 2020, 2018, pp. 287–306.
[27] A. T. Chatfield, V. N. Shlemoon, W. Redublado, and F. Rahman, "Data scientists as game changers in big data environments," in Proceedings of the 25th Australasian Conference on Information Systems, ACIS 2014, 2014.
[28] O. E. Williamson, "Transaction cost economics: How it works; where it is headed," Economist, vol. 146, no. 1, pp. 23–58, 1998, doi: 10.1023/A:1003263908567.