ARTICLE IN PRESS

JBR-09221; No of Pages 8

Journal of Business Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Business Research

Factors influencing big data decision-making quality☆

Marijn Janssen *, Haiko van der Voort, Agung Wahyudi

Delft University of Technology, Faculty of Technology, Policy and Management, Jaffalaan 5, 2628 BX Delft, The Netherlands

ARTICLE INFO

Available online xxxx

Keywords:
Big data
Big data analytics
Big data chain
E-government
Governance
Decision-making
Decision-making quality

ABSTRACT

Organizations are looking for ways to harness the power of big data (BD) to improve their decision making. Despite its significance the effects of BD on decision-making quality has been given scant attention in the literature. In this paper factors influencing decision-making based on BD are identified using a case study. BD is collected from different sources that have various data qualities and are processed by various organizational entities resulting in the creation of a big data chain. The veracity (manipulation, noise), variety (heterogeneity of data) and velocity (constantly changing data sources) amplified by the size of big data calls for relational and contractual governance mechanisms to ensure BD quality and being able to contextualize data. The case study reveals that taking advantage of big data is an evolutionary process in which the gradually understanding of the potential of big data and the routinization of processes plays a crucial role.

1. Introduction

Big data (BD) has rapidly moved to being a mainstream activity of organizations. Tapping into large-scale, fast-moving, complex streams of datasets has the potential to fundamentally transform the way organizations make their decisions. Big data refers to datasets that are both big and high in variety and velocity, which makes them difficult to handle using traditional tools and techniques (Elgendy & Elragal, 2014). The ability to take advantage of all available information has become a critical ability for organizational success (Olszak, 2016). The creation of value from data requires combining large datasets originating from different and heterogeneous data sources (Janssen, Estevez, & Janowski, 2014). Big data is closely related to Big Data Analytics (BDA) which are needed to create value of the data (Elgendy & Elragal, 2014; Holsapple, Lee-Post, & Pakath, 2014).

Data is often generated by other organizations, by users on social media or provided by devices of the Internet of Things (IoT). In practice there is often a whole chain of activities in which various actors plays a role (Anderson, 2015). The variety of data sources, the need to combine various sources, and the use of BDA often requires the collaboration between organizations and departments to create a flow of activities. Organizational silos are locking the use of big data for decision-making (Economist Intelligence Unit, 2012). Data collection, processing and use is not done within a single department or organization. Instead data is collected by many parties and organizations might use collaborations and partnerships for acquiring the resources and capabilities for analyzing BD. All these actors and steps increate the difficulty of using BD for decision-making.

The value of BD often originates from the ability to take better decisions (Economist Intelligence Unit, 2012). The quality is not solely dependent on the data, but also on the process in which the data is collected and the way data is processed. BD and BDA often requires bringing together multiple actors from different disciplines and diverse practices to examine the underexplored relationships between types of data (Janssen & Kuk, 2016). Each activity may be carried out by different actors with different capabilities and skills.

The involvement of a variety of organizations results in a flow or chain of activities which can be labeled as the '*big data chain*'. A BD chains begins with collecting the data from the sources and ends when data-based decisions are taken. A big data chain consists of subsequent activities that can be distinguished analytically. The term 'chain' refers to the analytical view taken on the collaboration (Stank, Keller, & Daugherty, 2001). In reality, there are many data sources, variations in flows and decisions. In such a chain there are many efforts to increase the quantity and quality of published data over time. These include removing noise, converting selected datasets into machine readable and linked data and adding meta-data (Kitchin, 2014). These activities can affect how BD can be used for decision-making. The chain perspective is hardly taken as an analytical view of looking at BD. A search using the keywords "big data" and "chain" results only in a few hits. Brown, Chui, and Manyika (2011) are the only one who mention the chain in relationship to big data to express the need to cooperate with supply chain partners and the role other parties can play in generating data.

The goal of the research is to identify factors influencing BD decision-making quality. Often it is assumed that BD results in better decisions,

☆ The author is grateful for proofreading the paper by Karin Boersma and the comments by anonymous reviewers on prior drafts of this article.

* Corresponding author.
E-mail address: m.f.w.h.a.janssen@tudelft.nl (M. Janssen).

http://dx.doi.org/10.1016/j.jbusres.2016.08.007
0148-2963/© 2016 Elsevier Inc. All rights reserved.

but it is unclear which factors influence the decision-making quality and how decision-making quality can be improved by organizations. As BD and BDA become more common, understanding the big data chain and factors influencing decision-making quality chain becomes of paramount importance for organizations.

This structure of this paper is as follows. First, factors influencing BD a decision-making quality as found in the literature. Thereafter the research method in which a BD chain is analyzed in a large, administrative organization is presented. Based on the case study factors influencing decision-making quality are derived. Finally, business implications are discussed and conclusions are drawn.

2. Big data for decision-making

Several steps for the BD process starting with data capturing and resulting in decision-making can be found in the literature. For example, Bizer, Boncz, Brodie, and Erling (2012) identify six steps; data capturing, data storage, data searching, data sharing, data analysis, and data visualization. In contrast Chen and Liu (2014) only use three steps; data handling, data processing, and data moving. Marx (2013) proposes five steps; problem definition, data searching, data transformation, data entity resolution, answer the query/solve the problem. Whereas others used other names for denoting similar steps. For example Zhou, Chawla, Jin, and Williams (2014) use data collection, data storage, data management, data manipulation, data cleansing, and data transformation. Although steps are identified little attention is paid to who executes these steps and the effects of one step on the other steps.

BD is commonly characterized with three or more Vs: Volume, Velocity, Variety (McAfee & Brynjolfsson, 2012). Gandomi and Haider (2015) add three other Vs to this list; Value, Variability and Veracity. These Vs makes the datasets difficult to handle in traditional ways (Elgendy & Elragal, 2014). BD often originates from many sources which are often beyond the control of a single actor like social media and devices. Veracity refers to that data may be incomplete, out-of-date and contains noise (Gandomi & Haider, 2015). BD sources have a variety of data quality. Data quality is a multidimensional concept describing properties of the information such as accuracy, timeliness, completeness, consistency, relevance and fitness for use (Miller, 1996; Strong, Lee, & Wang, 1997). Data quality can be viewed as set of dimensions describing the quality of the information produced by the information system (DeLone & McLean, 1992). Past research about the use of data shows that data quality influences decision-making quality (Keller & Staelin, 1987; O'Reilly, 1982). As such, BD quality might influence decision-making quality.

An effective BD chain requires to build capabilities and capacity for data management and BDA (Chen & Hsieh, 2014). BDA capabilities include descriptive, exploratory, inferential, predictive, causal and mechanistic techniques (Schutt & O'Neil, 2013). For that reason, various methods are employed such as natural language processing, text mining, linguistic computation, machine learning, search and sort algorithms, syntax and lexical analysis, and so on. BD is often related to predictive analytics which comprises a variety of techniques that predict future outcomes to uncover patterns and find relationships in data (Gandomi & Haider, 2015). Past research in data processing shows that the organizational capability to process information impacts its performance (Galbraith, 1973; Premkumar, Ramamurthy, & Saunders, 2005). Activities for processing BD and BDA capabilities likely influence decision-making quality.

The capabilities of each organizational entity involved in the BD chain influence the outcomes. Capabilities include skills and processes that transform inputs into outputs of greater value (Wade & Hulland, 2004). The ability of organizational entities and firms to collect, prepare and analyze BD might be different. Firms may possess a capability that is idiosyncratic to the firm or difficult to imitate due to path dependencies (Dierickx & Cool, 1989). For example, they might have experts with specialized knowledge in BDA that are rare on the market (Davenport & Patil, 2012). Firms might also have developed a customized infrastructure enabling them to take advantage of BD.

Raghunathan (1999) defines decision-making quality as the accuracy and correctness of decisions. Decision quality may improve or degrade when information quality and processing improves (Raghunathan, 1999). As data becomes larger, more complex, and more inexplicable, the limited mental capacities of humans pose difficulties in deciphering and interpreting an unknown environment (Sammut & Sartawi, 2012). In BD there might be little understanding of what the data actually means and in which context data is collected. Lack of knowledge about BD sources influences decision-making quality.

Decision-makers should be able to interpret the outcomes of BDA and should not be manipulated by fancy graphics (Huff, 1993). Raghunathan (1999) found that decision quality improves if the decision-maker has knowledge about the relationships among problem variables. In contrast, the decision quality of a decision-maker may degrade if the decision-maker does not understand the relationships. Interactions with those who collected and processed data produce better decisions than those without (Burleson, Levine, & Samter, 1984). This may equally apply for BD, suggesting that interactions with others persons involved in the BD chain involved result in higher decision-making quality.

The previous overview shows that the quality of the decision depends on the quality of the inputs and on the quality of the process that transforms the inputs into outputs. Factors affecting the decision-making quality of BD include the characteristics and quality of the BD sources, the quality of the BDA process, the BDA capacity and capabilities of persons involved in collecting and processing BD, and the availability of an BD infrastructure. In addition, research in data processing shows that the ability of decision-makers to understand the data and collaborating with others in the BD chain results in better decision quality.

3. Research approach

Literature shows that there is a wide range of factors influencing the BD decision-making. Particular in situations in which several actors are involved and it is hard to oversee all steps of the BD chain, the quality of decisions might be compromised. A deep understanding of the context is necessary to understand factors affecting it (Davenport, 2010; Goodhue, Wybo, & Kirsch, 1992). In a similar vein, a deep understanding of the BD chain is required to understand the factors influencing decision-making quality. Therefore an indepth case study within a large information-processing organization was conducted.

A qualitative approach based on a case study research (Yin, 1989) was adopted to gain a deep understanding of the factors influencing decision-making quality. The case study research methodology is particularly well-suited for investigating organizational issues (Benbasat, Goldstein, & Mead, 1987). A single case study can contribute to scientific development through a deep understanding of the context and by capturing experiences (Flyvbjerg & Budzier, 2011). Deep understanding is necessary to identify a broad range of factors influencing decision-making quality, whereas understanding experiences results in the identification of mechanisms for improving the decision-making quality. As decision-making quality is dependent on the decision-maker, the data collecting and processing, all these aspects were taken into account when analyzing the case study.

The number of cases that could reveal factors influencing BD for decision-making quality and using BDA for decision-making were found to be limited. This was further complicated as some of the cases considered did not want to disclose their practices. The Dutch Tax organization was selected as this organization was willingly to share their practices and much information was available publicly. Furthermore, this large information processing organization is considered as a frontrunner in the use of BD and BDA within the Dutch government.

Please cite this article as: Janssen, M., et al., Factors influencing big data decision-making quality, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.08.007

The Tax organization manages a complex BD chain and already used BDA in their decision-making processes, which has resulted in new insights and large cost-savings. These examples have been reported in the news. The case study was investigated using interviews and investigating websites, documents, reports and media. Both high-identify factors influencing decision-making.

4. Case background

Gamage (2016) provides an overview of governments from all over the world adopting big data strategies. Public sector organizations routinely collect large volumes of data in domains like health, business activities, crime, safety, security, weather, pollution, traffic, tax and income. The Tax organization processes immense amounts of data that are related to the millions of persons and companies who pay taxes. The BD chain consists of many tasks performed by various parties that need to be managed. BD and BDA are used by the Dutch Tax organizations to reduce costs and improve compliancy. The Dutch Tax organization has a legal basis to use data to improve the assessment of tax filing and collection of taxes. They have initiated various BD-driven programs to realize this. Their ambition is to hire 1500 new staff members who are able to make better assessment of tax filing of persons and organizations which should result in an additional income between 750 million and 2 billion Euro. At the same time 5000 staff members should become obsolete. The whole focus of the employees shifts towards BDA to detect tax evasion and fraud.

4.1. Big data chain

The Dutch Tax organization has 92 links with other organizations to retrieve data about 10 million citizens and 1,2 million organizations paying taxes. Relevant data sources of other public organizations include base registries like the citizens, business and vehicle registry, but also crime information or surveillance information. In additional they want to employ data published by citizens on social media like Facebook. Many public organizations collect structured and unstructured data for their own purposes which can be used to enrich the internal data of the Tax organization. The Tax organization manages the BD chain by collecting data from other public and private organizations and combining the data with their internal systems.

The use of BD and BDA was initially ad-hoc and dynamic. The data was available, but how it could be used was not clear. Often four departments were involved which respectively collected, prepared, analyzed and made decisions (see Fig. 1). The BD collection was often an ad-hoc process based on personal relationship, which incidentally required the making of agreements with other organizations. Sometimes already cleansed data was provided, whereas other time raw data was provided that needed to be prepared before BDA could be used. The unstructured and ad-hoc processes for collecting data pose high requirements on the data quality and the data processing capabilities.

Next the data was prepared by a separate department who was in charge of the data quality and performed activities related to improving the data quality, data enrichment and creating a dataset that is suitable for using analytics. These activities often encountered challenges with the quality of the data and extensive pre-processing was necessary to find and remove mistakes and to create an understanding of the data.

Once the data is ready, the BDA department started analyzing the data to identified patterns and so on. This department closely collaborated with the data quality department, as often datasets needed to be further enriched or questions were raised which required additional data. "*There are so many information sources and each of them seemed to be different, often providing us a headache when trying to combine them*" was mentioned by an interviewee. Data is used for detecting generic patterns about populations, but also for finding patterns of individuals and organizations. Often generic patterns are mirrored on individuals to find anomalies. A challenge is that all these sources provide BD in different ways and manners and might contain mistakes and noise.

Finally, the results are prepared for use by *decision-makers*. The head of the BDA department played a key role as he often presented the results to the board. At later stages the big data chains will be integrated in the operational processes. The decision-makers in these processes will be the inspectors who used the BD and BDA results for making inspection decisions. This requires clear guidelines as their decision-making authority is limited. A decision-maker complained about the complexity "*there is always a 'but' when the provide the data. Why can't they just provide a single recommendation*" followed by an appreciation of the insight gained to improve tax filling and collection and the cost savings accomplished. Also it was commented that "*decision-making became more difficult as combination with other BD might show a different pattern*".

BD sources originate from different places within and outside the Tax organization. Data capturing methods are often based on collecting as much as data as possible because value of the data is mostly prior unknown. Collecting and getting access to high quality data was found to be a major issue. A key aspect found was the transferring of knowledge about the data among the organizational entities involved in the BD chain. A *transfer point* is the handing over of data from one organizational entity to another organizational entity. Transfer points also exist within the Tax organization, as there are many siloed departments owning different types of data. These transfers may be more complicated than perceived initially, as not only the actual data need to be transferred but also the knowledge about the data. Fig. 1 shows the transfer points between four main activities. The arrows stress the transfer points and the need for coordination between these activities. The collection and transfer of data is guided by relational or transactional governance mechanism.

Transferring entails more than just data transfers. For example the cleansing of data might have removed distortion in early phase, which might reduce insight. Distortions might be early warning indicators for the need to investigate in more detail. The sharing of knowledge between the organizational entities and the making of clear agreements about how the data is transferred and how the data is already is processed was found to be essential for being able to use BD in the subsequent steps.

Several times interviewees mentioned that organizational entities were hesitant to share their data and knowledge with others. In particular for situation in which the use of the data was not completely clear in advance and the BDA was searching for possible patterns. Social

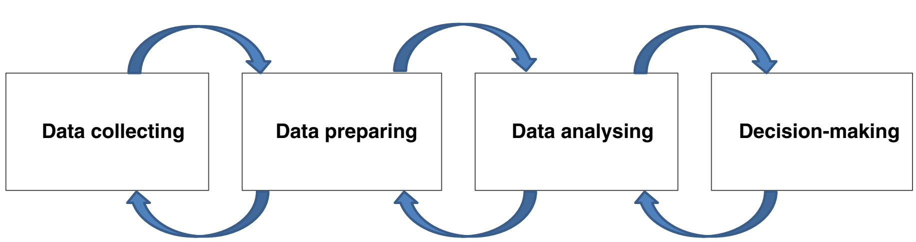

Fig. 1. Steps and transfer points in the big data chain.

Please cite this article as: Janssen, M., et al., Factors influencing big data decision-making quality, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.08.007

processes for trust building among parties were necessary. Sometimes parties refused access to or providing data. Finding out what the meaning of the BD was and the implications of the analysis proved to be challenging. Deep understanding of the context of the tax applications was found to be necessary. A data analyst mentioned "*How do we understand the big data? We had to team-up with various domains experts and only during the discussion we discovered the meaning of the pattern we found.*". Collaboration and knowledge sharing among departments and domain experts were found crucial for using BD. In particular the working together in teams consisting of fiscal and tax domain experts on the one hand and BDA experts on the other and proved to be important.

4.2. Big data for decision-making

BDA are used to detect patterns and for finding factors that contribute to incorrect or fraudulent tax filing and for identifying persons who did not pay their taxes. Often it was a cumbersome searching process before the examples listed hereafter were found. A variety of data sources were used and the BD veracity and validity often was an issue. For example data from persons, who did not pay or were unable to pay their taxes, was combined with data from the social security registers and the vehicle administration. Combining this data resulted in the detection of persons who claimed that they were unable to pay their taxes, but owned an expensive car. Also persons were identified who received social benefits, but were not able to pay their taxes. Also tax payers data was combined with unstructured data on social media like Facebook, Twitter and Marktplaats.nl. This data was used in various ways. For example, to check whether somebody who claimed to have no or limited income employed other activities to earn money. Also to analyze whether persons employed business activities and did not pay taxes about these business activities. The outcomes were used for guiding inspection decisions and have resulted in higher tax income.

BDA patterns show generic relationships which might not always hold for individual cases. For example, somebody trading goods could also mean that somebody was just selling their own belongings to able to buy new good. Another example is that names were mixed up. Making wrong judgement has already resulted in complaints and one case has hit the news. The Tax organization should be reserved in using the BD for certain purposes, as this can have a undesired societal or political impact.

Profiles are created of persons and organizations to detect if there are any risks concerning the providing of the information and the collection of taxes. These profiles can be used for improving tax filing and preventing mistakes. For example, BDA results showed that people in a divorce provided often information that was not correct and incomplete. In a similar vein BDA results showed that self-employed persons who have become ill or unemployed often made mistakes in their tax filing. This has resulted in remedial activities including creating awareness of common mistakes and the simplification of forms. Less mistakes resulted in the use of less resources by the Tax organizations to identify and correct mistakes.

BDA applications are often dependent on data from other public agencies to know their social security and marital status. Sometimes large and varied datasets are necessary, but for this example a simple notification of a status change is sufficient. The receiving of such an alert requires the making of clear agreements between parties about how the data should be delivered. Sometimes BD was transferred without having in-depth knowledge about the dataset structures, the context the data was collected and its data quality. This poses a challenge for the BDA department, as more information needed to be collected to understand the data. Often interactions with the data providers was necessary.

For persons and organizations who did not pay or were not able to pay their taxes *dynamic monitoring* is introduced. Persons and organizations who did not pay their taxes are monitored for conducting economic activities. For example, BDA discovered a person having social security posting all kinds of advertisements on an e-market. These BDA applications require the reliance on information sources that are external to the government and are volatile. Also persons having a high risk profile can be monitored in this way.

4.3. Organization

BD and BDA were new and innovative activities for the Tax organization. Although initially the BD and BDA activities were initiated within the operational department, but after some time a new organizational entity was found. This new entity was operated separately from the operational departments and could work in an agile manner, because this organization was not bound by institutionalized patterns, procedures and principles. Being a separate entity helped to attract a high skilled workforce and to search for new ways of using BDA and to pilot new practices. The new organization entity was directly controlled by the board of directors to enable quick decision-making about the directions to be taken. Apart from the agility this had also the advantage that this allowed to make a conscious trade-off between the government and individual needs. BD changes the balance of power between the government and individuals. Government should avoid exercising power and the use of BD and BDA should be careful considered. For this the board also consulted other stakeholder groups to ensure that the various interests were taken into account when deciding to use BD. In particular if the use of BD outweighs the privacy of the individual was found to be a frequently occurring trade-off.

The outcomes of the BD and BDA resulted in the need to standardize and routinize these activities and to integrate them in the operational administrative processes to enable real-time use of BDA. At the same time there was a need to educate, change and to empower staff. The organization has initiated this process, but the fully automation of the use of BD in real-time for the operational processes was not implemented, as considerable technical challenges were encountered. In particular the integration of the BDA in business process management systems supporting the administrative work was found to be challenging. Another issue was the knowledge transfer of how the data is collected and manipulated. The velocity of the data and the need to understand the contextual meaning makes it often difficult to routinize the work and embed the use of BD and BDA in the operational processes.

5. Factors influencing decision-making quality

The case study provides a rich insight into the many factors influencing decision-making quality. Although no longitudinal study was conducted, the documents and interviews showed that the influence of these factors changed over time. The use of BD has become more mature and institutionalized with the organization. The adoption of BD and BDA was an evolving process which started with ad-hoc activities and in which organizational changes were necessary to take advantage of the possibilities. Over time agreements were made with other parties to acquire the data needed. New staff were hired and a new department was created. A separate organizational department that operates independently of the operational department was founded to take advantage of BD. This resulted in the ability to create value using BDA within a relatively short time frame.

By taking a chain perspective a large number of interrelated factors that influence the decision-making quality were identified. Factors influencing the decision-making quality are listed and described in Table 1. The data quality of the many sources of data influences decision-making. Subsequently this information needs to be processed. The quality of the systems, the integration of processes for handling the data and contractual and relational governance for ensuring data quality and knowledge transfer were found to be key factors. The better the systems are integrated and are suitable for handling BD the easier it becomes to process BD. The capabilities of the workforce were found to be important, as staff should have the rights skills and competences.

Table 1
Factors influencing the decision-making quality in the case study.

Factors	Description
Contractual governance	The making of agreements and contracts with BD providers is used to increase the data quality. Agreements among organizations are used to ensure mutual understanding of BD, to create clear responsibilities and procedures, and to improve communication.
Relational governance	Relational governance is necessary for building trust among organizational entities and for ensuring the sharing of relevant knowledge that is necessary to interpret BD. Good relational governance includes communication and knowledge exchange which is necessary to understand and process data.
BDA capabilities	Analyzing BD can contain dozens of variables and parameters. It was difficult to find the right tools for analyzing. Which techniques can be possibly used and how BD can be visualized is a challenge. This was often a long search process in which knowledge of BD, BDA and the domain was necessary.
Knowledge exchange	Both data and knowledge about the data needs to be transferred. Knowledge about how the data is collected and processed is necessary for being able to interpret the data and to understand how it can be used. Once BDA analyst have more knowledge about the context the use of BDA and the finding of patterns and relationships becomes easier.
Collaboration	The ability to collaborate among BD providers, BDA analysts and decision-makers is a key condition to overcome fragmentation and create a BD chain. Furthermore, the inability to collaborate with data providers and to acquire the data can block the creation of valuable applications.
Process integration and standardization	The ability to integrate processes and to standardize tasks and data results in enhancing the BD chain. This results in lower efforts and cost to use BD and BDA. This is important condition for standardizing and routinizing the use of BD.
Routinizing and standardization	By routinizing BD chain the BD velocity is improved. This helped inspectors to make decisions in real-time.
Flexible infrastructure	Having a flexible infrastructure determines the ability and the amount of effort necessary to handle and process the data. Systems integration improves the handling of BD. Initially much manual work was necessary which resulted in long lead times for arriving at results.
Staff	Finding specialists who can deal with BD, have knowledge of BDA, and are able to communicate with business persons to interpret the results are a key conditions. These people are scarce. Partnership with companies enabled to use people from outside the Tax authority.
Data quality of the BD sources	BD provides little value if it is not accurate and people are not able to interpret the decisions. Wrong decisions can even be more costly. Wrong decision had even a societal impact, as this resulted in questions asked by politicians.
Decision-maker quality	Decision-makers should be able to interpret the outcomes of the analytics and understand the implications. In the case it was found that the more experienced decision-makers were, the better and the faster decisions could be made.

Education was needed and sometimes external staff was hired based on partnership with BD companies. Governance is a complex factor as it is affected by factors like communication, trust, decision-making responsibilities and procedures, which can be views as antecedents of governance. Both relational and contractual governance mechanisms were found. The first was dominating at the early stages, whereas the latter becomes important when BD chain becomes institutionalized.

The description in the table shows that many factors are interrelated and can reinforce each other. A chain dependency is that BD source quality influences the need for data processing activities. Sometimes several processing steps were necessary before the data could be used. The combination of data source quality and the data processing ability influences the decision-making quality. Furthermore, governance mechanisms can be used to get a better grip on and even improve BD quality. Making clear agreements about the quality of BD can improve decision-making quality in the end. Governance is aimed at posing the right conditions for data processing and ensuring that the right data is collected at the right quality.

Many aspects of BD influence the decision-making quality. For example BD volatility can result in incorrect outcomes. The collection of vehicle data by speed trap cameras for the purpose of speed violations can be used to identify if persons who are leasing a car for their business also use this car for their personal use. This requires a different quality of data as the place and time when the data is collected are the crucial variable, instead of the speed of the car and the speed limit which are necessary for providing fines. Although this information is collected, the place often is not stored in the data, as measurements are often fixed and datasets are handled individually. When this information is aggregated the place is not included and needs to be added afterwards which requires the asking for additional information. This process was found to be prone to error. The location of the speed trap camera might changes over time. The way data is collected and data sources are subject to change and further complicates the processing of data. Their main issues found in the case study with BD were related to the Velocity, Validity and Veracity and have to do with the following issues.

1. *Processing and manipulation*: The velocity of the data can mean that only one part of the data is provided which might give a different picture than when the whole dataset can be viewed. For example data about some persons are not communicated, although this data might be available. Some groups might be more fraudulent, but this is kept hidden to avoid stigmatization. This type of data might not be communicated, but this influences the outcomes of BDA and ultimately the decision-making.
2. *Noise*: Data is incorrectly connected to each other. Identities of persons are mixed up, an wrong place is mentioned, or some data from different periods are connected to each other. For data analytics revealing patterns this might not be a problem, but for inferring about individual cases it is.
3. *Error*: The context in which the data is collected is often not known and only the source has this information. If there are any changes in the way data is collected and this is not communicated this results in erroneous results. Not enough understanding of the context and the changes in the context is available. For example, this can result in connecting data collected in different years with each other.

These challenges are not easily to deal with due to the volume of the data. One of the interviewees mentioned that '*the amount of data influences the possible level of control*'. Meaning that the more data is used the less attention can be paid to ensuring the correctness of the data. Part of the causes is that BD sources are used that are often collected for other purposes. What is measured is used, instead of what kind of data needs to be collected to answer the problem.

The processing of data contains a number of steps ranging from data cleansing, combining, aggregating to analyzing. One interviewee mentioned "*The combining of data is probably the most complex step but also the most neglected step*". The combinations might result in various limitations that need to be taken into account. The naming conventions, data definitions, place and time of measurement, and granularity of data all play a role when combining data. For instance, although the term 'net income' is standardized over the past decades from over 50 definitions to a handful, there are still several different definitions used in different contexts, like taxes and pensions. This makes it complex to aggregate the data. Sometimes assumptions are made or data is left out based on the judgement of the person in charge of this. In addition, not having the same unique identifiers among data sources

Please cite this article as: Janssen, M., et al., Factors influencing big data decision-making quality, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.08.007

makes it sometimes even impossible to combine the data. Apart from technical issues, also privacy legislation can prevent data sharing. This is confirmed in the literature as understanding the privacy of these datasets is key to their broad use (de Montjoye, Radaelli, Singh, & Pentland, 2015).

Apart from the skills of the decision-makers, the experience of decision-makers proved to be important. This confirms the findings of Raghunathan (1999) on decision-making that simultaneous improvement in data quality and decision-maker quality results in higher decision quality. The first time decisions were made, there was unbridled enthusiasm over the results, but also much uncertainty over their possible use and the interpretation of the analysis. Specifically, there were worries about violating legislation, like privacy, and whether the decision would prove to be correct. After a while the decision-makers better understood the limitations and the potential for their decision-making.

Often manual work was done to get the data out of the systems. In the long term the ambition is to automate the handling of BD to integrate them in the operational processes. The creation of such a BD pipeline requires the integration of the disparate BD chain processes and the standardization of the data. Such a data pipeline is necessary for integrating BDA in the administrative processes. However, the current situation is far from this ideal, as many manual steps need to be taken and the existing business process management systems need to be updated. Process integration is expected to take several years.

The case study confirmed the findings of Gamage (2016) finding the right people having the rights skills was viewed as a major challenge. Much efforts were invested in attracting the right people. Attracting and employing the right persons became possible through the creation of a separate organizational entity. "*There is a huge gap between the capabilities of the people needed and the capabilities our staff has*" was a statement that was expressed by several persons. The situation is even worsened as each department had different types of persons who sometimes had a hard time to understand each other. Nevertheless collaboration among departments was found to be necessary for using BD and BDA for improving decision-making quality. Whereas the information quality department was traditionally focused on maintaining the quality of their internal data they were confronted with external data sources having unknown quality that were out of their direct control. They were not used to handle this.

In the BD chain many actors had different capabilities. The actors involved are different players which are often independent, and process and manipulate the data before transferring. They might have different strategies and be unaware of the possible use of BD by others. The quality department might remove all exceptions and anomalies from the dataset in order to improve the quality, whereas the data scientist might want to look for these kinds of patterns. A lack of understanding about how data is collected, what processes are being implemented, and what the data might be used for might disrupt organizational processes.

Apart from having the knowledge to handle data also relational capabilities were found to be important. The persons needed to collaborate to gain a deeper understanding of what can be done with the data. Also communication of the findings, including the validity and the implications, is a crucial aspect. Finally, understanding the context in which the Tax organization operates in, including relevant legislation is important, as its sets the boundaries to what is socially and legally allowed.

Initially there were no formal agreements about the use and sharing of data. At the beginning a one-time access to data sources was necessary as situations were only analyzed once. In these cases gaining access to the data was often dependent on personal relationships. Sometimes jurists needed to be involved to ensure that data sharing would not violate any legislation or policy. Data was extracted from the operational system once and thereafter used. This approach affected the timeliness of the BD and resulted in the need to improve the BD velocity.

For a continuous collecting of data a well-functioning BD chain requires stability of transactions. For this relationship governance was introduced in the case study. Relationship governance describes any activity that enhances mutual relationships in order to facilitate transactions and build trust (Bevir, 2012). Using data from not-trusted sources having unknown quality is risky and this should be avoided for a public organization that should represent the societal interest. The need for access to some information sources, like the base registries, and the continuous inclusion of data from external sources in the operational processes poses the need to have formal agreements between the data suppliers and Tax organizations. These formal agreements should warrant the continuous access to the data, information quality but also that formats or other aspects are not changed. Relational contracts consisted of contracts and Service Level Agreements (SLAs). Relational contracts dictate the specification of the data, its qualities and its form. As such, the relational contracts regulate the transfers between these activities if they are carried out by different actors. Also within the own organizations agreements were developed as they were not used to sharing BD.

6. Business research implications

The factors identified in this research can be used to improve the BD decision-making quality by organizations. Decision-making based on BD requires the organization of the activities to acquire and use BDA to analyze data. Organizations have to develop sophisticated processes to understand the context to have the precise meaning to make BD usable for decision-making. Big data can offer some unusual inflection points for new insights and understandings and has the potential to improve decision-making. This requires a good overview of possible data sources and how they are combined to understand the context. Creating such a BD chain requires that BD experts from external parties collaborate closely with domain experts to understand the analysis and implications of BD. The collaboration aspects were found to be a crucial aspect for employing BD and BDA need further research attention. Also creating an overview of BD sources and the evaluation of the quality of the resources is an area which needs more research.

Embedding and institutionalizing BD and BDA as a part of the operational organizations was found to be difficult. Instead a new department was introduced. Having a new department enabled the Tax organization to work in an agile manner, as how data could be used, which data was necessary, which types of BDA could be used and the possible purposes were not clear. This also enabled to change quickly and adapt new practices. The existing organization was found to be bound by institutionalized patterns, procedures and principles which would make it a cumbersome process to take advantage of BD and BDA within a short time frame. Furthermore, by establishing a new organization the management was able to attract new, highly skilled workforce which had the right capabilities instead of having to rely on the capabilities of the existing workforce. A next step will be the integrating of BD and BDA into existing processes which requires that existing staff are able to deal with the collecting and processing of data. The expectation is that this requires considerable effort and reorganization. This practice is domain dependent and more research into the integration of BD and BDA into organizations is needed.

BD is noisy, messy, constantly-changing and comes in different formats. The veracity of BD makes the creation of value from BD challenging. There are datasets from many systems, some incomplete, with noise and errors, collected at different point of times. These have to be copied, cleansed, translated and unified which might affect the potential for decision-making. For this reason the BD chain needs to be managed. However, several actors are outside the hierarchical scope and there was limited control on the activities they perform. Also relying entirely on contractual governance was no option, because formal agreements explicating their responsibilities are not suitable due to the velocity of BD. Relationship governance, based on trust and informal rules, was found to be of vital importance to facilitate the transfers in the BD chain. Decision-making quality depends heavily on the capability, trust and willingness of actors to play their role in the chain well. Increased collaboration among BD chain participants leads to lower total cost and enhanced decision-making quality. Building trust involves personal qualities, but also skills and knowledge. Staff should have the capability and willingness to safeguard for instance accuracy, timeliness, completeness, consistency, and relevance of BD which is no sinecure. Especially delivery of data in real time requires constant attention.

The case study clearly shows the evolutionary nature of adopting BD and BDA approaches, which is surrounded by uncertainty and complexity. This nature becomes paramount when looking at the governance. Initially the activities were primarily based on informal relationships. The quality and the data analytics departments contacted each other informally and there was much dependence on informal communication. There was no governance of transfer points and no control of BD chain. As the number of data sources increased and the awareness of the need for information quality improved, SLAs were made, and contractual governance became more important for assuring a continuous stream of up-to-date data. This evolvement posed also new requirements on the skills of persons. Whereas at the start number crunching was important, after some time the communication skills gained importance. Kitchin (2014) provides a large variety of skills required from BD professional. Persons having all these qualities are scarce or may not exist at all. This requires organizing staff in such a way that appropriate skills are available.

7. Conclusions

Value from BD and BDA is generated by improving decision-making quality. Despite its importance, there is limited research on the use of BD for decision-making to date. Often it is assumed that BD and BDA results in better decisions, but this might be a too simplistic assumption. There are many interrelated factors influencing decision-making quality. In addition, conceptualizing BD and BDA as a single process operated by a single data scientist is too simple as showed in the case study. Taking a chain perspective allows to analyze both the activities that are carried out in a BD chain and the organizations that carry them out, and to understand the interdependencies between those activities. This results in a more deeper understanding and more diverse set of factors influencing decision-making quality.

BD sources vary and have different characteristics which influences how BD needs to be processed and how BDA can be used. There are many factors influencing decision-making quality which need to be addressed simultaneously to improve decision-making quality. Process transformation and integration, development of skills, retaining experience and human resources, ensuring data quality, flexible systems, collaboration, knowledge exchange, decision-maker quality, building trust and managing relationships are among the key factors found. The main challenge found was not dealing with the volume, but the ability to understand the BD and use BDA to create value by dealing with the variety, velocity, veracity and validity of data. For these processes need to be in place to deal with these characteristics of BD. Decision-making quality is not just dependent on BD and BDA, but also about the ability to manage the BD chain.

The findings show that the quality of the source data, the processing of the data and how the transfer of the data is handled influences the quality of decision-making. These findings highlight the need to develop appropriate and effective contractual and relational governance mechanisms for managing the BD chain. It also shows the need to devise governance mechanisms for managing the collection and processing of BD. Governance should ensure access to the BD sources, creating insight into the quality of BD, and understanding of the meaning and limitations of BD. Although the variety of BD and the use of BDA enhance the ability to detect fraud and can be used to prevent the making of mistakes, it can also exacerbate discrimination. In government the balance of power between government and the individual needs to be careful balanced. In the case study the board was involved in this decision-making. In addition, also other parties were consulted when necessary.

The main limitation of this study relates to the use of a single case study within one domain based on interviews and documents. Finding a suitable case study was found to be challenging, however, the case study revealed experiences that are not found in the literature yet. Further research can focus on investigating cases from different domains and on generalizing the factors influencing decision-making.

References

Anderson, C. (2015). *Creating a data-driven organization.* O'Reilly Media, Inc.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*, *11*(3), 369–386.

Bevir, M. (2012). *Governance; A very short introduction.* Oxford: Oxford University Press.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: Four perspectives — Four challenges. *SIGMOD Record*, *40*(4), 56–60. http://dx.doi.org/10.1145/2094114.2094129.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. *McKinsey Quarterly*, *4*(2011), 24–35.

Burleson, B. R., Levine, B. J., & Samter, W. (1984). Decision-making procedure and decision quality. *Human Communication Research*, *10*(4), 557–574. http://dx.doi.org/10.1111/j.1468-2958.1984.tb00032.x.

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.

Chen, M., & Liu, S. M. Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209.

Davenport, T. H. (2010). Business intelligence and organizational decisions. *International Journal of Business Intelligence Research*, *1*(1), 1–12.

Davenport, T. H., & Patil, D. (2012). Data scientist. *Harvard Business Review*, *90*, 70–76.

de Montjoye, Y. -A., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, *347*(6221), 536–539. http://dx.doi.org/10.1126/science.1256297.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95. http://dx.doi.org/10.1287/isre.3.1.60.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science*, *35*, 1504–1511.

Economist Intelligence Unit. (2012). The deciding factor: Big data & decision making. In C.G. reports (Ed.), (pp. 24): Economist Intelligence Unit.

Elgendy, N., & Elragal, A. (2014). Big data analytics: A literature review paper. In P. (Ed.), *Advances in data mining. applications and theoretical aspects: 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, July 16–20, 2014. Proceedings* (pp. 214–227). Cham: Springer International Publishing.

Flyvbjerg, B., & Budzier, A. (2011). Why your IT project may be riskier than you think. *Harvard Business Review*, *89*(9), 601–603.

Galbraith, J. R. (1973). *Designing complex organizations.* Reading, MA: Addison-Wesley.

Gamage, P. (2016). New development: Leveraging 'big data' analytics in the public sector. *Public Money & Management*, *36*(5), 385–390. http://dx.doi.org/10.1080/09540962.2016.1194087.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007.

Goodhue, D. L., Wybo, M. D., & Kirsch, L. J. (1992). The impact of data integration on the costs and benefits of information systems. *MIS Quarterly*, *16*(3), 293–311.

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. Decision Support Systems, 64, 130–141. http://dx.doi.org/10.1016/j.dss.2014.05.013.

Huff, D. (1993). *How to lie with statistics.* New York: W W Norton & Company.

Janssen, M., & Kuk, G. (2016). Big and Open Linked Data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 3–13. http://dx.doi.org/10.1080/10919392.2015.1124005.

Janssen, M., Estevez, E., & Janowski, T. (2014). *Interoperability in big, open, and linked data—Organizational maturity, capabilities, and data portfolios.* IEEE Computer Society.

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research*, 200–213.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences.* Sage.

Marx, V. (2013). Biology: The big challenges of big data. *Nature*, *498*(7453), 255–260.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*(90), 60–66 (68, 128).

Miller, H. (1996). The multiple dimensions of information quality. *Information Systems Management*, *13*(2), 79–82.

Olszak, C. M. (2016). Toward better understanding and use of business intelligence in organizations. *Information Systems Management*, *33*(2), 105–123. http://dx.doi.org/10.1080/10580530.2016.1155946.

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

Premkumar, G., Ramamurthy, K., & Saunders, C. S. (2005). Information processing view of organizations: An exploratory examination of fit in the context of interorganizational relationships. *Journal of Management Information Systems*, *22*(1), 257–294.

Raghunathan, S. (1999). Impact of information quality and decision-maker quality on decision quality: A theoretical model and simulation analysis. Decision Support Systems, 26(4), 275–286. http://dx.doi.org/10.1016/S0167-9236(99)00060-3.

Sammut, G., & Sartawi, M. (2012). Perspective-taking and the attribution of ignorance. *Journal for the Theory of Social Behaviour*, *42*(2), 181–200.

Schutt, R., & O'Neil, C. (2013). *Doing data science: Straight talk from the frontline.* O'Reilly Media, Inc.

Stank, T. P., Keller, S. B., & Daugherty, P. J. (2001). Supply chain collaboration and logistical service performance. *Journal of Business Logistics*, *22*(1), 29–48.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103–110.

Wade, M., & Hulland, J. (2004). Review: The resource-based view and information systems research: Review extensions and suggestions for future research. *MIS Quarterly*, *28*(1), 107–142.

Yin, R. K. (1989). *Case study research: Design and methods.* Newbury Park, CA: Sage Publications.

Zhou, Z. H., Chawla, N. V., Jin, Y., & Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives. *IEEE Computational Intelligence Magazine*, *9*(4), 62–74.

Please cite this article as: Janssen, M., et al., Factors influencing big data decision-making quality, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.08.007