

Benefits, business considerations and risks of big data

by
Jorina Smeda

*Assignment presented in partial fulfilment of the requirements for the degree
MCom (Computer Auditing) at Stellenbosch University*



Supervisor: Mr LP Steenkamp
Faculty of Economic and Management Sciences
School of Accountancy

Ma!&@2015

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the authorship owner thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: Maart 2015

Abstract

Big data is an emerging technology and its use holds great potential and benefits for organisations. The governance of this technology is something that is still a big concern and an aspect for which guidance to organisations wanting to use this technology is still lacking.

In this study an extensive literature review was conducted to identify and define the business imperatives distinctive of an organisation that will benefit from the use of big data. The business imperatives were identified and defined based on the characteristics and benefits of big data. If the characteristics and benefits are clear, the relevant technology will be better understood. Furthermore, the business imperatives provide business managers with guidance to whether their organisation will benefit from the use of this technology or not.

The strategic and operational risks related to the use of big data were also identified and they are discussed in this assignment, based on a literature review. The risks specific to big data are highlighted and guidance is given to business managers as to which risks should be addressed when using big data. The risks are then mapped against COBIT 5 (Control Objectives for Information and Related Technology) to highlight the processes most affected when implementing and using big data, providing business managers with guidance when governing this technology.

Opsomming

'Big data' is 'n ontwikkelende tegnologie en die gebruik daarvan hou baie groot potensiaal en voordele vir besighede in. Die bestuur van hierdie tegnologie is egter 'n groot bron van kommer en leiding aan besighede wat hierdie tegnologie wil gebruik ontbreek steeds.

Deur middel van 'n uitgebreide literatuuroorsig is die besigheidsimperatiewe kenmerkend van 'n besigheid wat voordeel sal trek uit die gebruik van 'big data' geïdentifiseer. Die besigheidsimperatiewe is geïdentifiseer en gedefinieer gebaseer op die eienskappe en voordele van 'big data'. Indien die eienskappe en voordele behoorlik verstaan word, is 'n beter begrip van die tegnologie moontlik. Daarbenewens bied die besigheidsimperatiewe leiding aan bestuur sodat hulle in staat kan wees om te beoordeel of hulle besigheid voordeel sal trek uit die gebruik van hierdie tegnologie of nie.

Die strategiese en operasionele risiko's wat verband hou met die gebruik van 'big data' is ook geïdentifiseer en bespreek, gebaseer op 'n literatuuroorsig. Dit beklemtoon die risiko's verbonde aan 'big data' en daardeur word leiding verskaf aan besigheidsbestuurders ten opsigte van watter risiko's aangespreek moet word wanneer 'big data' gebruik word. Die risiko's is vervolgens gekarteer teen COBIT 5 ('Control Objectives for Information and Related Technology') om die prosesse wat die meeste geraak word deur die gebruik van 'big data' te beklemtoon, ten einde leiding te gee aan besigheidsbestuurders vir die beheer en kontrole van hierdie tegnologie.

Contents

1	Introduction	1
1.1	Background to the study	1
1.2	Problem statement.....	3
1.3	Aims and objectives	3
1.4	Motivation.....	4
1.5	Methodology.....	5
2	Big data definition, characteristics and benefits	6
2.1	Big data defined	6
2.2	Characteristics of big data	8
2.2.1	Data volume	8
2.2.2	Data velocity	9
2.2.3	Data variety	10
2.2.4	Data value	11
2.2.5	Data complexity.....	11
2.3	Benefits of big data	12
2.3.1	Increased value delivery	13
2.3.2	Better predictions	13
2.3.3	Improved ability to adapt to change	14
2.3.4	Benefits of scale	14
3	Business imperatives	16
3.1	Introduction	16
3.2	Pro-active management.....	17
3.3	Innovation.....	18
3.4	Agility.....	19
3.5	Scalability	19

3.6	Impact of big data characteristics on business imperatives	20
3.7	Conclusion	21
4	Strategic and operational risks related to big data	22
4.1	Introduction	22
4.2	Strategic risks.....	23
4.2.1	Integration risk.....	23
4.2.2	Interoperability risk	25
4.2.3	Security risk.....	27
4.2.4	Scalability risk.....	28
4.2.5	Retrofit risk	29
4.2.6	Impact of strategic risks on business imperatives	29
4.3	Operational risks	30
4.3.1	Planning risks	30
4.3.2	Design risks	32
4.3.3	Build/setup/configure risk	32
4.3.4	Operating risks	33
4.3.5	Maintenance risks	34
4.4	Conclusion	34
5	Risks mapped to governance framework	36
5.1	Introduction	36
5.2	COBIT 5	37
5.3	Risks mapped against COBIT 5	39
5.3.1	Processes most affected by implementing big data	46
6	Conclusion	48
7	References.....	51

List of figures

Figure 1.1 Investments or planned investments in big data	2
Figure 5.1 Illustration of the impact of strategic and operational risks on COBIT 5 processes	46

List of Tables

Table 3.1 Impact of big data characteristics on business imperatives	20
Table 4.1 Impact of strategic risks on business imperatives	30
Table 4.2 Planning risks	31
Table 5.1 Strategic risks mapped against COBIT 5.....	40
Table 5.2 Operational risks mapped against COBIT 5	43

1 Introduction

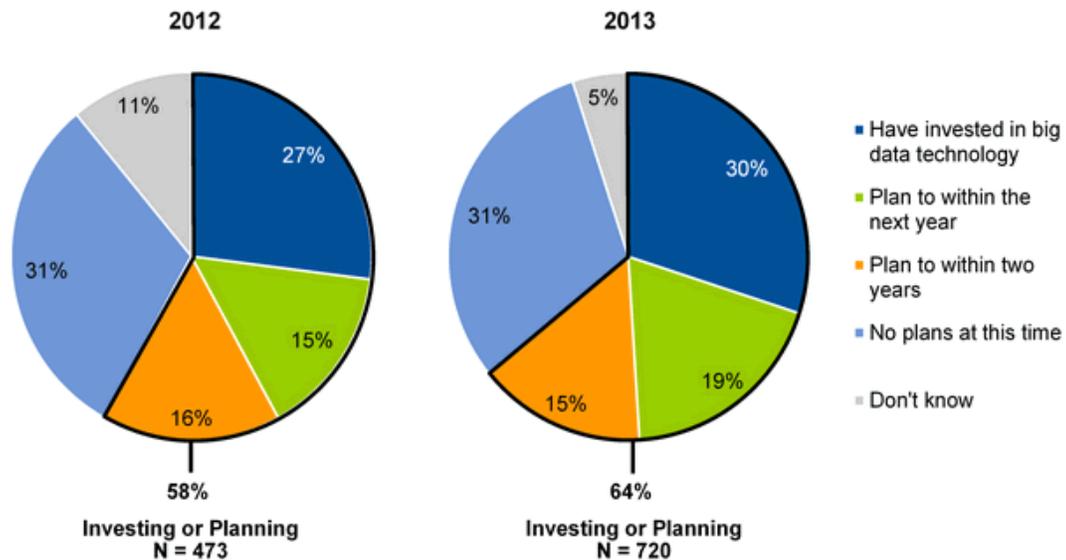
1.1 Background to the study

Businesses that are aware of where they are in terms of the adoption of analytics have a competitive advantage because they are better prepared to turn challenges into opportunities (Steve, Eric, Rebecca, Michael & Nina, 2011). In a world where the amount of data to which companies have access increases significantly every year; this becomes an important driver for business success. The competitive advantage that a business can gain from the use of analysing large amounts of data is based on the fact that it is able to provide better value to its customers by making better decisions based on the information obtained through analysing the data. According to Way and Yuan (2012) this results in higher return on equity for these organisations and over time so-called “big data” has the potential to become a new type of corporate asset.

Big data is defined by the Tech American Foundation as “large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information” (TechAmerica Foundation, 2012).

In recent years an increasing number of companies have invested in big data to improve their value delivery. The illustration in Figure 1.1 indicates how companies’ investment in big data has grown in 2013 compared to 2012, based on a survey done by Gartner in 2013 (Kart, Heudecker & Buytendijk, 2013).

Figure 1.1 Investments or planned investments in big data



Kart, Heudecker & Buytendijk, 2013

Value delivery is one of the governance objectives of the Control Objectives for Information and Related Technology (COBIT 5). According to COBIT 5, enterprises exist to create value to their stakeholders and therefore value delivery is one of the governance objectives of every business (ISACA, 2013). The fact that the use of big data results in better value delivery gives rise to a situation where the risks related to the use of this technology will have to be governed in order to ensure that this governance objective is achieved.

According to a Gartner circle study the number one concern and challenge with big data is the governance thereof (Buytendijk & Heiser, 2013). Big data is a developing technology that has its own set of unique risks. Businesses have access to vast quantities of data, which, when analysed effectively, can lead to a lasting competitive advantage (Johnson, Neff & Stuart, 2012). As technology evolves, incremental risks to the business arise which may have a negative impact on the business.

In order to govern the implementation and use of a new and evolving technology such as big data, the strategic and operational risks related to big data will have to be identified. By mapping these risks against a control framework management will

be provided with an indication of which processes will require specific consideration during the implementation of big data to ensure that this process is governed appropriately.

1.2 Problem statement

Using technologies such as big data results in strategic and operational risks for the business that should be governed to mitigate these risks. Much has been written on the general risks of big data, but not much on strategic and operational risks of this technology to provide guidance to business managers for the governance thereof.

In order to identify the strategic and operational risks of big data, consideration should first be given to the business imperatives distinctive of an organisation that will benefit from the use of this technology. This will provide business managers with guidance on whether this technology will be a suitable option for their organisation or not.

1.3 Aims and objectives

The purpose of this study was to identify the strategic and operational risks related to the use of big data for decision making purposes. A set of business imperatives that are key when a business considers the use of this technology was also identified to help business managers evaluate whether this technology is appropriate for use in their business.

This study identified a number of strategic and operational risks when implementing and using big data. These risks were mapped to COBIT 5 to identify the processes most affected by implementing the big data, giving business managers an indication of which process should be governed carefully to ensure that the gap between business and IT is appropriately bridged.

The purpose of this study was firstly to identify the business imperatives characteristic of an organisation that will benefit from the use of big data, to give business managers guidance as to whether their business will benefit from the use of this technology or not. Secondly, the strategic and operational risks related to big data needed to be identified and mapped against the COBIT 5 framework. By mapping the risks to the COBIT 5 framework business managers could be given an indication of which processes will be affected most when implementing big data, thus providing guidance on which processes will have to be governed more carefully when implementing big data.

This study focussed on the use of big data for making business decisions in order to obtain a competitive advantage. The technical details on procedures and software used to perform data analysis were not included. By identifying the business imperatives this study will assist business managers in assessing whether big data will be an option for their specific business and the strategic and operational risks and the mapping thereof will provide them with guidance on how to govern this technology.

Only strategic and operational risks related specifically to big data were included in the study and the list of risks provided can by no means be considered to be a complete list of risks relating to big data.

1.4 Motivation

According to Buytendijk and Heiser (2013) big data is an emerging discipline that is still lacking best practices. There is thus a need for guidance on how to govern the implementation of this technology. Before implementing new technology an organisation must evaluate whether the technology is suitable and beneficial to the organisation.

The business imperatives identified in this study will provide guidance to business managers in their decision on whether or not big data is a suitable technology for

their organisation. The list of strategic and operational risks highlights the risks of implementing this technology and will assist businesses to bridge the gap between business and IT by improving the alignment between the particular businesses and their IT goals.

By mapping these risks to the COBIT 5 framework the processes most affected when implementing big data are highlighted, thus providing guidance to business managers on how to govern the use of this technology effectively.

1.5 Methodology

A literature review (Chapter 2) was used in defining big data and in guiding the discussion on the characteristics of the technology. A list of benefits when using this technology within a business was also compiled.

Also based on a literature study, the characteristics identified and presented in Chapter 2 were then used to compile a list of business imperatives distinctive of a business that will benefit from the use this technology in Chapter 3. This list of business imperatives and a further literature study were used to identify the strategic and operational risks specific to big data as presented in Chapter 4.

The risks referred to above were mapped against the COBIT 5 control framework to identify the COBIT 5 processes most affected by implementing big data to provide management with guidance when governing this technology. This mapping against the COBIT 5 framework is provided in Chapter 5.

2 Big data definition, characteristics and benefits

2.1 Big data defined

By increasing the value delivery of a business big data is becoming very appealing to business managers. Over the past decade the volume of data collected and stored by organisations has multiplied, reflecting the principle of Moore's law (Anderson & Roberts, 2012). According to Moore's law the storage capacity of storage mediums and the processing power available double approximately every two years (Moore's Law and Intel Innovation, 2013). Many business managers do not understand the term 'big data' correctly and incorrectly assume that if they have a large database it means they are using big data (Longbottom, 2012). One of the reasons for this error is that the term big data is very widely and loosely defined. In order to understand exactly what big data is, it has to be defined properly.

According to Kaisler, Armour, Espinosa, and Money (2013), big data was originally defined as a volume of data that cannot be processed efficiently by means of traditional databases and tools. However, this definition focussed on structured data alone. As storage mediums developed over the years the volume of data as referred to in the definition was able to increase making it possible to process bigger and bigger volumes of data.

With the development of new technology the data stored by organisations moved from being mainly structured data to now also including unstructured data (Anderson & Roberts, 2012). Structured data refers to data that has a pre-defined structure and that can easily be saved in a relational data base. Unstructured data, on the other hand, refers to data without a pre-defined structure that cannot be easily stored in a relational database, for example emails, videos and images (From Insight to Impact, 2013). When defining big data, both these sources of data, structured as well as unstructured, must be taken into account. Structured data, according to Data Science Central, is data that is stored in a database where each field in the database has a name and the relationship between the fields is defined (Walker, 2012). Data

Science Central defines unstructured data as data not stored in a relational database, for example data produced by The Internet of Things, blogs, videos, social media and emails, to name a few (Walker, 2012). These more complex forms of unstructured data are clearly included in the big data definition as described by MIS Quarterly. They describe big data as complex data sets that are so large that they require specific and more advanced forms of data storage, management and analysis (Chen, Chaing & Storey, 2012).

Information Week defines the technology of big data as building new analytic applications based on new types of data, in order to better serve one's customers and derive a better competitive advantage (Bertolucci, 2013). The value delivery to be gained from the use of this technology is not just in being able to process large volumes of data efficiently, but in being able to analyse the different types of data and new types of data (structured as well as unstructured) in such a way as to develop actionable information (Kaisler *et al.*, 2013).

The International Data Corporation (IDC) defines big data technologies more formally in a way that highlights the different characteristics of big data: "Big data technologies describe a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis" (Gantz & Reinsel, 2011).

From these definitions it is evident that using big data adds value to the business by increasing the usefulness of the data in decision making and giving the business a competitive advantage. According to COBIT 5, organisations exist to create value and will have value creation as a core governance objective (ISACA, 2013). Using big data as a driver to create value will result in a need for the governance of the technology.

2.2 Characteristics of big data

From the IDC definition of big data it is evident that big data is characterised as a wide variety of data in large volumes that is captured at a high velocity (Gantz & Reinsel, 2011). Kaisler *et al.* (2013) add two more components to these, namely the value of the data and the complexity of the data.

The characteristics of big data can be summarised as follows, based on the abovementioned:

- Data volume
- Data velocity
- Data variety
- Data value
- Data complexity

In order for business managers to understand the benefits of big data and to determine whether their organisation will benefit from big data these characteristics must first be defined. Through defining the characteristics of big data, a list of benefits of big data can be compiled that will indicate to business managers if their organisation will benefit from this technology. The characteristics of big data are defined and discussed in detail in section 2.2.1 to 2.2.5.

2.2.1 Data volume

According to the Harvard Business Review, 2.5 exabytes of data are created daily (McAfee & Brynjolfsson, 2012). That is about half the amount of data created from the beginning of civilisation until 2003 (Schmidt, 2010). One of the main reasons why so much data is created nowadays is that much of it is unstructured data. Where previously most of the data an organisation had access to was structured data captured by the organisation, whereas nowadays organisations also have access to, *inter alia*, real-time data from activity on their websites (Russom, 2011), as well as unstructured data available through social media. These other sources of

data are created much faster than structured data increasing the volume of data created daily.

Since the volume of data available to organisations is substantial there is an opportunity to use and analyse this data in a way to derive useful information from it. Big data will enable organisations to do this, because it will allow them to derive useful information from much bigger volumes of data than would be the case without this technology.

Big data is also distinguished from normal data by the fact that organisations using it do not have to own all the data; they should only be able to access it (Kaisler *et al.*, 2013). The volume of data available to the organisation through big data is not limited to the amount of data stored in the relational database of the specific organisation, but rather to the amount of data to which the organisation can get access. Consequently a vast number of data sources become available to the organisation. When such data is combined with the organisation's own data the organisation will therefore have access to more and relevant information they would not have had access to if they were not using big data.

Big data will give business managers access to more useful information which will enable the organisation to gain a competitive advantage in the marketplace. Being able to access and use such a big volume of data and successfully use it to gain a competitive advantage is something that is unique to an organisation that makes use of big data.

2.2.2 Data velocity

For data to be created in the volumes as mentioned in section 2.2.1, it must happen at quite a substantial speed. This leads to the second characteristic of big data: the velocity of the data. Data velocity refers to the speed or frequency at which data is generated or delivered (Kaisler *et al.*, 2013; Russom, 2011).

Many organisations are already collecting data in real time from their websites (Russom, 2011), in other words at the velocity at which it is being created. The challenge these organisations face is being able to analyse this data in real time (Russom, 2011). By making use of big data techniques the organisation will be able to not only collect data at the velocity at which it is created, but also to analyse the data in real time, producing useful information that management can use for decision making, all in real time.

Access to this real-time information streams will result in the organisation being much more agile than their competitors (McAfee & Brynjolfsson, 2012). Another advantage of having access to real-time information is the competitive edge the organisation will have in terms of being an innovative organisation. Both agility and innovation are discussed in more detail in Chapter 3.

2.2.3 Data variety

From the definition of big data it is evident that data classified as big data comes from a vast number of sources. According to The Data Warehouse Institute (TDWI) this is one of the aspects that is making big data so topical (Russom, 2011). The main reason for this is the fact that big data is optimised for unstructured data (Johnson, Neff & Stuart, 2012). As mentioned in the previous section unstructured data arises from a variety of different sources.

However, data variety does not only refer to the number of sources, but also to the richness of the data representation (Kaisler *et al.*, 2013). Richness refers to the type of data, for example text, audio and images (Kaisler *et al.*, 2013). In terms of richness, big data also has a much bigger data variety, because it can incorporate unstructured data shared by users via the internet and social media, which includes a variety of images and audio.

Being able to incorporate richer data into their current system by using big data also increases the complexity of the data businesses are able to use in their decision

making processes. The complexity of big data is one of its other characteristics differentiating it from other data. This aspect is further discussed in section 2.2.5.

By making use of powerful analysis of the large variety of data available to organisations through the use of big data will enable the organisation to gain access to automated discovery techniques (Buytendijk & Heiser, 2013). Consequently the organisation will be able to predict the needs and wants of their customers to a certain extent, resulting in a competitive edge in the market.

2.2.4 Data value

The value of big data can be measured by how useful the data is (Kaisler *et al.*, 2013) or how useful the information is that an organisation can extract from that data (Digging Up Value with Big-Data Mining, 2014). The fact that a variety of data sources and also sources that are not necessarily owned by the organisation can be used as part of big data is something that increases the data value.

The usefulness of the data is measured by how predictive the data is and how useful the information is that the organisation can derive from this data. By having access to more data and also more relevant sources of data and by combining these sources with the data sources the organisation has available the usefulness of the information derived from these sources will be improved.

For example: if a courier company is able to incorporate live traffic data when planning delivery schedules and routes, it can improve its fuel efficiency and customer satisfaction by being able to avoid traffic incidences causing delays in delivery time or resulting in longer driving times and the use of more fuel.

2.2.5 Data complexity

The complexity of data is measured by the interconnectedness and interdependence of the data structure (Kaisler *et al.*, 2013). This data can be from different sources,

as long as there some link between the data (Beyer, Lapkin, Gall, Feinberg & Sribar, 2011). If a variety of data is interconnected and/or interdependent, the complexity of the structure is measured as the impact a small change will have on information derived from the data (Kaisler *et al.*, 2013).

This interconnectedness of the data is unique to big data and provides business managers with valuable information for decision making purposes they would not have had access to if the organisation did not make use of big data. It also enables them to be able to adapt easily if changes arise because they can evaluate the effect of the change.

The characteristics discussed in section 2.2 are an elaboration on the definition of big data. These characteristics will give an organisation an indication of what big data entails. To enable an organisation to start evaluating whether it will be able to benefit from the use of this technology the benefits of big data must be identified. In section 2.3 the benefits derived from the abovementioned definition and characteristics are identified and discussed.

2.3 Benefits of big data

From the characteristics of big data discussed in section 2.2 one can compile a list of benefits of big data. This list will enable business managers to start to identify whether big data is a technology from which their organisation can benefit.

The key benefits of big data based on the characteristics thereof are the following:

- Increased value delivery
- Better predictions
- Improved ability to adapt to change
- Benefits of scale

Each of these benefits is discussed briefly in the sections below:

2.3.1 Increased value delivery

Value delivery, according to King III, is optimising expenditure and proving the value of IT in an organisation (King Report on Governance for South Africa 2009, 2013). By making use of big data, organisations will be able to gain intelligence from data and translate this intelligence into business value (McAfee & Brynjolfsson, 2012). The business value of using big data lies in the fact that very large amounts of data from different sources can be analysed effectively to provide actionable information.

Big data techniques are also specifically optimised for effectively transforming unstructured data into useful information (Johnson, Neff & Stuart, 2012), improving the effectiveness of data and information management even further. Being able to transform large quantities of complex data from different sources effectively will result in reduced expenditure for the organisation because less time is spent on analysing the data (Johnson, Neff & Stuart, 2012). By reducing expenditure big data will increase the value of IT for the organisation.

Value delivery is one of the COBIT 5 governance objectives (ISACA, 2013). Big data will enable the organisation to be able to meet this objective more efficiently, as seen above. The influence of big data on governance and the risks associated with this technology are discussed in detail in Chapters 4 and 5.

2.3.2 Better predictions

Big data fundamentally changes the way decisions are made within an organisation (Brown, Chui & Manyika, 2011). By making use of controlled experiments on the available data, the organisation can better predict the behaviour of their customers and use this knowledge to their advantage (Brown, Chui & Manyika, 2011). This results in the organisation adding more value to their customers because they can better predict what their customers want (Kaisler *et al.*, 2013) enabling them to make better business decisions (Johnson, Neff & Stuart, 2012) that will lead to a competitive advantage in the marketplace. By being able to better anticipate the

future behaviour of their customers the organisation will be able to be more proactive in their business approach.

The variety of data sources available to organisations and the appropriate analysis of this data can be a powerful tool for business managers in analysing customer behaviour and trends to make more accurate decisions about innovations in their products and services to give them an edge above their rivals.

2.3.3 Improved ability to adapt to change

The interconnectedness and interdependence of the structures within big data will enable business managers to determine the effect of possible changes more easily (Kaisler *et al.*, 2013). If an organisation is able to determine the effect of possible changes, it will be more ready to adapt or react to these changes if they do occur.

2.3.4 Benefits of scale

Organisations that are able to capture large volumes of valuable data and use the data effectively at scale will have a competitive advantage above other organisations (Brown, Chui & Manyika, 2011). By implementing big data, organisations are able to capture data from different sources within the business and to acquire data from outside sources (partners and customers), giving them the ability to integrate the data and derive useful information from it by using analytics (Brown, Chui & Manyika, 2011). A further benefit of scale of big data, according to Johnson, Neff and Stuart (2012), is the fact that other departments can benefit from the big data system used by one department by gaining access to the data stored or information already generated by that department.

By having open access to big data and to the information and reports already generated from this data the organisation will experience cost and time savings, because there is no duplication of work in various departments. If there is access to all the data from both internal and external sources, more useful information can be

derived from the data. Consequently, decision making will be improved and the organisation will be able to achieve their strategic objectives more easily by gaining a competitive advantage over their rivals through improving their customer satisfaction by creating a better customer experience.

Considering the abovementioned benefits on their own might not be sufficient for an organisation to identify accurately whether they would be able to benefit from the use of big data. These benefits can, however, be used to identify and define the business imperatives distinctive to an organisation that will benefit from the use of big data. By comparing these business imperatives to their own set of business imperatives, organisations will be able to get a more accurate indication whether their organisation will benefit from the use of this technology or not. In Chapter 3 the business imperatives of an organisation that will benefit from the use of big data are identified and discussed in detail.

3 Business imperatives

3.1 Introduction

Every organisation has a set of business imperatives specific to the type of organisation and the industry in which it operates (Boshoff, 2012). An organisation that will benefit from the use of big data will have certain business imperatives that will be more easily and effectively achieved by using big data than it would be without it. These business imperatives related to big data can be identified based on the characteristics and benefits of big data as identified and discussed in Chapter 2.

In order to identify the business imperatives of an organisation, the term 'business imperative' must first be defined. Terblanche (2011) defines business imperatives as the non-negotiable principles set by management that have to be achieved in order for the business to reach its strategic business objectives. Goosen (2012) defines business imperatives as strategic level objectives that are seen as the critical and fundamental business drivers that will enable a company to achieve its objectives and that give the company a competitive advantage.

Based on these definitions business imperatives can be defined as the critical and fundamental business drivers that are imperative in the achievement of the strategic objectives of the company resulting in a competitive advantage in the environment within which the organisation operates. By using these drivers to achieve the strategic objectives of the organisation, the gap between business and IT in the organisation is bridged, resulting in better alignment of the IT and business goals of an organisation.

In this study four main business imperatives, characteristic of organisations that will benefit from the use of big data, were identified by means of a literature review. Business and IT alignment is one of the principles of the King III report on IT governance (Badenhorst, 2009). These big data specific business imperatives will result in better business and IT alignment within an organisation that implements this technology.

The business imperatives distinctive of an organisation that will benefit from the use of big data are the following:

- Pro-active management
- Innovation
- Agility
- Scalability

In sections 3.2 to 3.5 below, these business imperatives are defined and the effect of big data on each of them is discussed in detail.

3.2 Pro-active management

An organisation that has pro-active management as one of its business imperatives is an organisation where management require information in real time in order to react and adjust their decisions according to what is happening at that moment (Boshoff, 2012). According to Géczy (2014), one of the primary goals of many organisations is to have access to actionable information.

As discussed in the section on the benefits of big data, big data provides organisations access to actionable information derived from large volumes of very rich and complex data. Another benefit of using big data is that the effect of changes (big or small) can easily be determined due to the interconnectedness of big data (Kaisler *et al.*, 2013). Giving business managers access to real-time actionable information derived from this data enables them to make more accurate and informed decisions (Goosen & Rudman, 2013). By being able to predict the effect of changes more accurately business managers will be able to be pro-active in their management approach by anticipating possible outcomes of changes; consequently they will be able to act faster and more effectively if the changes do occur.

This attribute is unique to an organisation that has access to and uses big data, because big data allows the organisation access to useful information derived from

the large quantities of rich and complex data, which would not have been possible using ordinary relational databases and database tools. One of the reasons why this is possible with big data is that by using big data, organisations gain access to unstructured data sources. Data from these unstructured sources are a good indication of what is happening in the marketplace and of the customers' perception thereof. Therefore the organisation will gain a competitive advantage because it is able to act pro-actively.

3.3 Innovation

Being the most innovative company in the market will give an organisation a substantial competitive advantage. Being innovative can be defined as being an industry leader when it comes to the development of new products and services and satisfying customer needs. In order to achieve this goal, business managers need systems with high functionality that can provide information in real time (Boshoff, 2012).

Big data provides organisations access to useful information, in real time, derived from a big variety of complex data. The interconnectedness of big data with the resulting benefit of being able to easily determine the effect of changes to elements on the outcome or system (Kaisler *et al.*, 2013) provides business managers with the information they need to make innovative decisions faster and more accurately in order to gain the competitive advantage. The availability of such information in real time will increase this advantage even further.

By analysing big data an organisation gains access to the predictive power of the data (Kaisler *et al.*, 2013). This will enable business managers to predict the behaviour of their customers more accurately, giving them a competitive advantage when developing new products and services. Organisations that make use of big data will be able to use the information gathered from big data not only to identify opportunities in the market, but also to predict what it is the customers want,

enabling them specifically to target their customers market their products and services more effectively to a specific target market.

3.4 Agility

Agility is the ability to adapt to unpredictable changes within the business environment easily and effectively (Van Oosterhout, Waarts & Van Hillegersberg, 2006). Most organisations are considered flexible if they are able to adapt to predictable changes (Vokurka & Flidner, 1998). Agile organisations are more than just flexible because they are able to respond effectively to unpredictable changes because they are continually ready for change (Vokurka & Flidner, 1998).

By making use of big data the organisation will have access to real-time data and information (Russom, 2011). Paired with the improved predictability of big data, this enables the organisation to adapt to possible changes easily because it is able to anticipate them. Another benefit of using big data is the improvement in the organisation's ability to adapt to these changes easily because the effect of changes to the data can easily be determined because of the interconnectedness of big data (Kaisler *et al.*, 2013).

By being able to determine the effect of a possible change, business managers will be able to consider the different possible outcomes of a scenario resulting in their being able to react more effectively if the change does occur.

3.5 Scalability

Scalability can be defined as the ability of a system to handle a growing amount of data in an effective and flexible manner or the ability to accommodate such growth by being enlarged (Bondi, 2000; Géczy, 2014). As discussed in Chapter 2, one of the characteristics of big data is the ability to process large volumes of data efficiently. By making use of big data the organisation automatically has the ability to

handle this large volume of data and to handle even bigger volumes of data than it is currently able to do.

Another advantage of scalability, in a big data context, is the fact that it enables departments within an organisation to make use of the same systems and reports (Johnson, Neff & Stuart, 2012). The benefits of scale related to this is considerable, because the one department already has the data or information available that might be useful to another department. By making use of big data and giving other departments access to the data and reports already generated, costs as well as time can be saved, because there is less duplication.

3.6 Impact of big data characteristics on business imperatives

The definition of the big data business imperatives in the sections above indicates that the different characteristics of big data have an impact on each of these business imperatives in a different way. Table 3.1 indicates the level (high, medium or low) in which each of the characteristics influence the business imperatives that have been identified. The business imperatives are listed horizontally and the characteristics vertically.

Table 3.1 Impact of big data characteristics on business imperatives

Imperatives Characteristics	Pro-active Management	Innovation	Agility	Scalability
Data volume	Low	Low	High	High
Data velocity	Med	Low	Med	Low
Data variety	Med	High	High	High
Data value	High	Med	Med	Low
Complexity	High	High	High	Low

3.7 Conclusion

In this chapter the most significant business imperatives related to big data have been listed and discussed. For an organisation to evaluate whether their business will benefit from the use of big data the organisation will have to identify and define business imperatives for their organisation and compare them to this list. If the business imperatives of the organisation match the business imperatives listed here the organisation will be able to benefit from the use of big data.

4 Strategic and operational risks related to big data

4.1 Introduction

Business imperatives are the key drivers that assist an organisation in achieving its strategic goals (Terblanche, 2011; Goosen, 2012; Boshoff, 2012). An organisation can make use of technology, which fits in with the business imperatives of the organisation, to achieve its business objectives more effectively. However, making use of technology, such as big data, to achieve this purpose involves specific risks related to the technology (IT risk).

IT risk can be defined as the risk that the IT system of an organisation does not support the organisation in such a way that it will effectively achieve its business objectives (Worrel & Bush, 2007). According to Boshoff (2013), IT risk can be divided into two categories: strategic risks and operational risks.

Strategic risks are defined by Bromiley, Rau and McShane (2014) as “the risks inherent to the company’s strategic decisions”. The strategic decisions an organisation make are those decisions that will enable them to achieve their strategic goals. According to the definition of business imperatives provided in Chapter 3, these strategic decisions will be influenced by the business imperatives of the organisation, because the business imperatives are the drivers that assist the organisation in achieving its strategic goals (Terblanche, 2011; Goosen, 2012; Boshoff, 2012). Therefore, strategic risks are those risks that will have an impact on the business imperatives of an organisation.

According to Boshoff (2013), strategic risks can be divided into the following five main categories:

- Integration risk
- Interoperability risk
- Security risk
- Scalability risk
- Retrofit risk

Operational risks can be defined as “the risk of losses resulting from inadequate or failed internal processes, people and systems” (Böcker & Klüppelberg, 2008). In an IT context it will be the risks related to the IT system of the organisation. The operational risks related to implementing a new technology such as big data will be those risks that will have an impact on the different stages of the systems life cycle.

The systems life cycle corresponds with the categories of the Life Cycle according to the COBIT 5 Enabler Dimension (ISACA, 2013) as well as with the five categories into which Boshoff divides operational risks. The five categories are as follows (Boshoff, 2013):

- Plan
- Design
- Build/Setup/Configure
- Operate
- Maintain

In this chapter each of these risks is defined and discussed within the context of big data. An extensive literature review was performed to identify strategic and operational risks specifically related to the use of big data. The strategic risks related to big data are discussed in section 4.2 and in section 4.3 the operational risks are discussed. Table 4.1 in section 4.2.6 indicates the business imperatives that are influenced by each of the identified risks.

4.2 Strategic risks

4.2.1 Integration risk

System integration is the process of combining different computer systems or software applications in a physical or functional manner, in order for it to act as a new coordinated system (Hobday, Davies & Prencipe, 2005). Implementing big data and integrating it into the current system of the organisation poses the risk that the

integration between the current system and the new technology may not be seamless.

Big data requires a solid IT infrastructure to perform as intended (Géczy, 2014). If the current IT infrastructure of the organisation is not sufficient to provide the storage capacity and processing power needed for big data, the organisation will face an integration risk when implementing this technology.

If the organisation is able to overcome the IT infrastructure risk another integration risk is attempting to integrate big data into the current system without acquiring the analytical software specifically designed for big data analysis. This poses a risk that the current database software is lacking in-database analytics needed to analyse big data and that the analytic queries will not be performed fast enough (Russom, 2011). The organisation will have to assess and decide whether it will also implement an analytics tool specifically designed use with big data or whether it will just make use of its current analytical tools to do the analysis. The difference between using a big data specific analysis tool and the current analytical tools available to the organisation is in the response time when performing the analysis. Tools specifically designed for use with big data will have a much faster response time than the current in-database analysis tools most organisations use in their current systems. One aspect that should be noted is that one of the main problems with the decision on whether or not to implement these tools is the lack of understanding by management of how big data analytics can be used to improve business (LaValle, Lesser, Shockley, Hopkins & Kruschwitz, 2011). This is a risk that will have to be addressed from the IT side of the business to help educate business managers on the advantages that can be gained from the effective use of big data. The benefits of big data as discussed in Chapter 2 can also give business managers a better idea of the advantages of implementing big data properly.

Integration is a problem that does not only arise between the current system and the newly implemented technology, but also within the data that is defined as big data. In Chapter 2 one of the characteristics of big data is listed as the variety of data that an organisation will have access to when making use of big data. Kaisler *et al.* (2013) indicates that some of the biggest problems arising from this big variety of

data are the incompatibility of the different data formats, data structures that are not aligned and inconsistent data semantics. This risk is even higher when a combination of structured and unstructured data are being used, because unstructured data can occur in many different formats and some of these formats do not necessarily fit into a relational database. Consequently, data streams that are not so well behaved could cause disruptions and mixed-in data that is often unrelated to the primary data of interest (Kaisler *et al.*, 2013). This will influence the analytics performed on the data to derive useful information from big data.

Business managers will have to be aware of the integration risks between the old system and the newly implemented technology, as well as within the different data formats. Business managers will have to make an informed decision on whether to implement big data specific analytical tools. Such a decision must be made based on the business goals that the organisation wants to achieve by implementing this technology as well as the business imperatives of the organisation.

4.2.2 Interoperability risk

Interoperability can be defined as the interaction between multiple agencies (Way & Yuan, 2012). The process in which opportunities for exchange and re-use of information, external or internal, is maximised by managing systems, procedures and the organisational culture (Way & Yuan, 2012). Implementing a technology such as big data involves hard interoperability classified as interoperability which relies on factors such as hardware, storage and communication protocols to exchange information (Way & Yuan, 2012).

The fact that interoperability involves the managing of systems as well as the organisational culture of the organisation poses two risks when using big data: (1) a skills risk and (2) a compatibility risk. Each of these risks is discussed in more detail below.

4.2.2.1 Skills risk

One of the concerns of big data implementation is the availability of the employees with the necessary skills to facilitate such an implementation as well as employees with the necessary skills to effectively use the technology and analyse the data. Research done by Brown, Chui and Manyika (2011) for McKinsey indicated that in the USA the demand for employees with the necessary big data skills was far bigger than the availability of these types of employees.

The lack of skills when big data must be implemented poses a big interoperability risk because without the necessary skills organisations will be unable to maximise the opportunities inherent to the use of big data.

The organisation will have to ensure that it has employees with the necessary knowledge and skills to implement and operate big data or they will have to consider hiring employees with the skills needed, to ensure that this new technology is utilised to its full potential and that it will add value to the organisation.

4.2.2.2 Compatibility risk

One of the characteristics of big data is the variety of the data. There are no universally acceptable standard to capture this variety of data (JASON, 2008). Big data is also not standardised, and since it comes from so many different sources and some of it can be very complex, there is a risk that the same fact may come from different sources; because of the different standards within the data sources it might not be recognised by the system as the same fact (Green & Panzer, 2014). These different data formats will have to be merged into one database, creating the risk that the formats might not be compatible, especially when data from unstructured sources is used.

The complexity of big data because of the interconnectedness and interdependence amongst the data in the data sources (Kaisler *et al.*, 2013) is another compatibility issue since it can lead to a ripple effect within the big data system if changes are

being made. This is however also one of the advantages of big data, as discussed in Chapter 2, but it is something that business managers must be aware of when they implement big data.

The variety and complexity of the data available through big data is clearly an advantage for the organisation, but it can also pose a compatibility risk that the organisation will have to manage.

4.2.3 Security risk

Security risks from a big data perspective are regarded as the privacy issues related to the data sources used (Kaisler *et al.*, 2013). The main reason for these privacy issues is the way in which unstructured data is collected.

Currently there are no real regulations or protocols regarding the accumulation of unstructured data, specifically data collected from social media platforms (Kaisler *et al.*, 2013). Consequently, privacy issues could arise because it might be possible to trace the data back to individuals, even if it does not seem like personal data (Mayer-Schonberger & Cukier, 2013). This can result in personal data of an individual being used without that individual having given permission for the use of such data.

Another privacy issue that can arise relates to the secondary use of data. An individual might have given consent for data to be used for a primary purpose, and the use of the data for this purpose might lead to manipulated information or a report generated from the data. At a later stage this information or report might be used for another (secondary) purpose. However, the individual granted permission for the primary use of the data, but at that time the secondary purpose was not known yet and the individual did not give permission for the use of the data for a secondary purpose (Mayer-Schonberger & Cukier, 2013)

With regard to permission being given by individuals for the use of their data, another concern is that individuals gave permission that their data can be used, believing that they will remain anonymous, but in fact it is still possible to identify them (Buytendijk

& Heiser, 2013). The anonymisation of the data used as part of big data is a concern that poses a serious security risk (Buytendijk & Heiser, 2013). This risk will have to be addressed by management when implementing big data. By anonymising the data the organisation will be able to prevent the abovementioned security risk and make it impossible for data to be traced back to specific individuals, thus reducing the privacy and personal security risks related to big data.

4.2.4 Scalability risk

In the discussion on the benefits of big data in Chapter 2, scalability was listed as one of the benefits of big data. However, a clear distinction must be made between the scalability benefit and the scalability risk of big data. The scalability benefit that arises from the use of big data is related to the fact that the analysis and reports as well as the larger volume and variety of data available through big data are available to all the departments within the organisation. Since the reports and data are readily available this prevents the duplication of data accumulation as well as the duplication of analysis, leading to cost savings for the organisation and greater efficiency of the different departments.

From a risk point of view scalability is the ability of a system, network, or process to handle a growing amount of work in a capable manner or its ability to be enlarged to accommodate that growth (Géczy, 2014). When implementing big data the volume of data that the organisation will work with and will have to be able to handle will increase significantly. The variety and complexity of the data that the system will have to be able to handle will also increase. All of these factors could cause pressure on the system if it does not have the necessary capacity to handle such volumes and complexity.

In a study done by TDWI Research in 2011 it was found that 23% of participants had scalability problems when using big data (Russom, 2011). Most of participants in this survey indicated that such a problem could be solved by making use of the appropriate big data tools. Management will have to ensure that the big data tools

used when analysing the data are the most appropriate ones for their specific organisation.

4.2.5 Retrofit risk

Retrofit risk is the risk that the system can become obsolete with the introduction new technology or features to an older system (Boshof, 2013). When implementing big data into the current system unstructured data will be added to a structured database. Without the unstructured data being standardised, as discussed in section 4.2.2.2, there is a risk that the current database could become unusable, because it is more than the organisation's infrastructure can handle.

According to Gantz and Reinsel (2011), data centre architectures as well as organisational models will have to evolve in order to accommodate big data applications within the organisation's IT infrastructure.

4.2.6 Impact of strategic risks on business imperatives

Table 4.1 indicates which strategic risks have an impact on which business imperatives. The business imperatives are listed horizontally and the risks vertically, an X indicating when a risk has an impact on an imperative.

Table 4.1 Impact of strategic risks on business imperatives

Imperatives Risks	Pro-active Management	Innovation	Agility	Scalability
Integration	X	X	X	
Interoperability	X	X	X	
Security	X	X	X	
Scalability			X	X
Retrofit	X	X	X	X

4.3 Operational risks

In a big data context the operational risks are risks that are for the most part related specifically to the big data tools implemented. This is especially the case for the build/setup/configure, operating and maintenance risks. This study does not deal with the specific big data analysis tools available, for these categories of operational risks general risks related to the building, setup, configuration, operation and maintenance of big data are discussed, but no details on the specific risks are given, because they relate to the analytical tools that fall outside the scope of this study.

4.3.1 Planning risks

Planning risks are the risks related to the planning phase of implementing the new technology. In the planning phase the creation and use of information resources are defined (ISACA, 2013).

When identifying the planning risks for implementing big data the strategic risks identified in section 4.2 were used as a basis. The reason for this approach is that the planning risks arise because of the strategic risks identified and determining

these planning risks will assist business managers in identifying the key areas that will need specific focus when planning the implementation of big data.

In Table 4.2 the planning risks related to each of the strategic risks are listed.

Table 4.2 Planning risks

Strategic risk	Planning risks
Integration risk	The risk that sufficient planning was not done to ensure that big data would be able to integrate with the current system of the organisation
Interoperability risk	<ul style="list-style-type: none"> • The risk that sufficient consideration was not given to whether the organisation has access to the skills necessary for a big data implementation • The risk that sufficient planning was not done to ensure that the system will be able to deal with the complexity of big data • The risk that the standardisation of data from different sources was not planned for
Security risk	The risk that the organisation did not determine how they will handle the privacy of the data used when implementing big data by giving consideration to how the data will be anonymised in such a way that private and personal data of individuals is protected
Scalability risk	The risk that consideration was not given to how the large volume and variety of complex data will affect the current system of the organisation
Retrofit risk	The risk that an evaluation was not done on the current IT infrastructure to ensure that it would be able to handle the implementation of a technology such as big data

Table 4.2 indicates how the planning risks are directly related to the strategic risks as discussed in section 4.2. Management will have to be aware of these risks and

during the planning phase they will have to identify ways to respond to these risks in order to manage them in such a way that the implementation can be successful.

4.3.2 Design risks

When implementing a technology such as big data there are not many general design risks because most of the risks related to the design of the system will already have been addressed during the planning phase. The main design risk is that the design for the systems and components of big data requires an understanding of both the needs of the users and the technology (Kaisler *et al.*, 2013).

The needs of the users as well as the expectations of management must be clearly defined and proper care should be taken to ensure that these needs are in line with what the technology can provide and that these needs are indeed met when the system is designed.

Another design risk is a lack of skills within the organisation to design a system that will be compatible with big data. One of the big risks when implementing a technology such as big data is the lack of employees with the necessary knowledge and skills to design the system that is needed to implement big data (Buytendijk & Heiser, 2013).

The organisation will have to evaluate whether it has current employees with the skills sets necessary to be able to bridge the design risk or whether it should consider contracting in employees with these skills.

4.3.3 Build/setup/configure risk

In a big data context the risks related to build/setup/configure are specifically related to the big data analytic tool the organisations choose to implement. There are,

however, general risks that are universal regardless of which tool the organisation chooses to implement.

As is the case with the design risk, having employees with the necessary skills to implement these tools is of utmost importance. Without employees with the necessary skills available the organisation will either have to employ someone with these skills or they will have to make use of consultants to assist them in the building, setup and configuring of these tools. A third alternative would be to provide current employees with training in order for them to acquire the skills needed.

Management will have to consider which of these options will be most beneficial for the organisation in the long run, based on the objectives of the organisation.

4.3.4 Operating risks

Risk regarding the operation of the big data tools is also to a certain extent related to the specific tool that the organisation chooses to implement.

The skills risk as discussed in section 4.3.3 is also present as an operating risk, because if the organisation does not have employees available with the necessary skills sets to operate the big data tools, this may lead to business risk indicators being missed (Jackson, 2013). Another operating risk related to the lack of skills is that organisations may be capturing and storing data that they cannot use or do not need (Jackson, 2013), resulting in wasted storage space and time.

There are also some big data processing (operating) risks that are universal across all the big data tools available. One of these is the extensive parallel processing power needed to effectively process the exabytes of data available through big data (Kaisler *et al.*, 2013). If the system is not capable of doing this it could cause a mismatch between the speed at which data is transmitted and the speed at which the system is able to store the data (Green & Panzer, 2014). Such a situation will result in a lag in the system that will influence the efficiency of the system.

Another processing risk is the issue of quantity versus quality. A decision will have to be made to determine which data is relevant and which not and management will have to make the decision to determine how much data is enough (Kaisler *et al.*, 2013). Both of these decisions will help to ensure that only reliable and accurate data is used and that the most valuable data is used for decision making.

The abovementioned risks will have to be managed by the organisation irrespective of which big data tools are being implemented.

4.3.5 Maintenance risks

As with the build/setup/configure and the operating risks, maintenance is specific to the big data tool implemented.

The general risk of a lack of skills available is something that will also be a risk to the maintenance of the big data system. To mitigate this risk the organisation can use the options as discussed in sections 4.3.3 and 4.3.4. Another option to bridge the skills risk related to the maintenance of the system is to consider outsourcing the maintenance. This is an option because maintenance is not a process that will happen every day, but it is something that must be done periodically or as a problem arises and it may be cost-effective for the organisation to outsource this function.

4.4 Conclusion

In the first part of this chapter the strategic risks related to big data were discussed and it was shown which business imperatives are influenced by each of these risks. In the second part of the chapter the operational risks were discussed. It is, however, important to understand that the operational risks are of a more technical nature and to a large extent directly linked to the specific big data analysis tool implemented by the organisation. The technical specifications of the big data

analysis tools are not included in this study and only general operational risks were listed in section 4.3.

The operational planning and design risks do, however, have a strong link with the strategic risks and these risks influence each other. Business managers will have to take time to ensure that these risks are addressed appropriately in order to ensure that the business objectives are achieved.

The impact of the risks listed in this chapter on the big data business imperatives as identified in Chapter 3 indicate that these risks will have to be governed in order to ensure that the business objectives are achieved and that the strategic business goals and the IT goals are aligned when implementing big data. In Chapter 5 the risks identified in this chapter are mapped to a governance framework to identify the process of the framework most influenced by implementing big data. In this way guidance will be given to business managers on the required governance when implementing and using big data.

5 Risks mapped to governance framework

5.1 Introduction

In Chapter 4 strategic and operational risks related to the use of big data were identified and defined. In this chapter the risks identified in Chapter 4 are mapped against a governance framework to evaluate the impact of the use of this technology and to identify the processes where the impact of the use of big data is the highest and which must be governed specifically in order to achieve the business objectives.

In the King III report IT governance is defined as the effective and efficient management of IT resources to enable an organisation to achieve its corporate objectives (King Report on Governance for South Africa 2009, 2013). According to Gartner (2009), corporate governance provides the structure for allocating authority and monitoring performance to ensure that business strategy is reached. This function of governance can therefore be applied specifically to IT.

To govern the use of technology effectively the organisation must make use of an appropriate governance framework. There are many such frameworks available, but one must consider which of these will be most appropriate for implementing a specific technology.

For this study the Control Objectives for Information and related Technology (COBIT) was used. The latest version of COBIT, namely COBIT 5, was released during 2012, and this is the framework used in this study. The reason why COBIT was used is that it is a framework that provides better control over IT investments. Implementing new technology such as big data is a considerable IT investment, making this the most relevant framework to use. In section 5.2 more detail is provided on COBIT 5 and in section 5.3 the strategic and operational risks as identified in Chapter 4 are mapped against COBIT 5.

5.2 COBIT 5

The COBIT framework is an internationally accepted best practice framework. Best practices frameworks are used to give organisations the ability to have better control over their IT investments by improving the performance and value of these investments (Hardy, 2006). In COBIT 5 it is emphasised that management of an organisation is responsible for managing IT (De Haes, Van Grembergen & Debreceeny, 2013).

Implementing technology such as big data will result in a substantial IT investment for an organisation. Sufficient governance and management processes must thus be in place in order to ensure that this investment is governed and managed appropriately.

The COBIT 5 framework is divided into a number of processes which are discussed in more detail later in this section. In order for the implementation of big data to be governed and managed properly the organisation must identify the impact of the implementation of big data on the COBIT 5 processes. This can be achieved by mapping the strategic and operational risks related to big data against the COBIT 5 framework to identify the COBIT 5 processes most affected by the implementation of big data. The mapping of the risks against COBIT 5 is done in section 5.3, but first more detail on the COBIT 5 processes are discussed in the rest of this section.

COBIT 5 makes a distinction between governance and management. According to COBIT 5, governance ensures that stakeholder needs are evaluated to determine the objectives that must be achieved and it also provides guidance through monitoring performance to ensure that these needs are met (ISACA, 2013). It describes management as the drivers behind planning, building, running and monitoring activities in such a way that the business strategy is achieved (ISACA, 2013).

The COBIT 5 reference model consists of 37 governance and management processes. These processes are divided into five domains: one for governance and four for management. The COBIT 5 domains are (ISACA, 2013) the following:

- **Governance:**
 - Evaluate, Direct and Monitor (EDM)
- **Management:**
 - Align, Plan and Organise (APO)
 - Build, Acquire and Implement (BAI)
 - Deliver, Service and Support (DSS)
 - Monitor, Evaluate and Assess (MEA)

The five domains are described below:

- **Align, Plan and Organise (APO):** The processes in this domain assist in the identification of IT and it can contribute to the achievement of the goals of the business (De Haes, Van Grembergen & Debreceeny, 2013).
- **Build, Acquire and Implement (BAI):** The processes in this domain identify detail requirements for managing the IT investment projects or program as a whole (De Haes, Van Grembergen & Debreceeny, 2013).
- **Deliver, Service and Support (DSS):** The processes in this domain deal with the delivery of the actual IT services required (De Haes, Van Grembergen & Debreceeny, 2013).
- **Monitor, Evaluate and Assess (MEA):** This domain includes the processes to assess whether the requirements set in all the processes of APO, BAI and DSS domains are complied with (De Haes, Van Grembergen & Debreceeny, 2013).
- **Evaluate, Direct and Monitor (EDM):** This is the only domain that forms part of governance only. The processes in this domain refer to the overall governance of all the processes included in the management domains.

In section 5.3 the processes in each of these domains are indicated in Tables 5.1 and 5.2.

In section 5.3 the strategic and operational risks as identified in Chapter 4 are mapped against COBIT 5 to identify the processes most affected by implementing big data. This is done to identify which COBIT 5 processes must be governed and managed more specifically when implementing big data to reduce IT risk.

5.3 Risks mapped against COBIT 5

The two main types of risks (strategic and operational) are mapped separately against COBIT 5 and the results are combined in a graph in Figure 5.1 in section 5.3.1. In Table 5.1 the strategic risks and in Table 5.2 the operational risks are mapped.

The mapping was done per COBIT process for each risk category as defined in Chapter 4. The processes affected by each risk are indicated with a cross in the column beneath each risk.

This mapping has been done to identify the processes most affected by implementing big data, giving management an indication of which COBIT 5 processes must be managed and governed more closely to ensure IT investment with implementing big data is managed effectively.

Table 5.1 Strategic risks mapped against COBIT 5

COBIT Domain	COBIT process		Integration risk	Interoperability risk	Security risk	Scalability risk	Retrofit risk
Evaluate, Direct and Monitor	EDM01	Ensure Governance Framework Setting and Maintenance					
	EDM02	Ensure Benefits Delivery				X	
	EDM03	Ensure Risk Optimisation	X	X	X	X	X
	EDM04	Ensure Resource Optimisation		X			
	EDM05	Ensure Stakeholder Transparency					
Align, Plan and Organise	APO01	Manage the IT Management Framework					
	APO02	Manage Strategy	X	X	X	X	X
	APO03	Manage Enterprise Architecture	X	X	X	X	X
	APO04	Manage Innovation	X	X			
	APO05	Manage Portfolio					
	APO06	Manage Budget and Costs					
	APO07	Manage Human Resources		X			
	APO08	Manage Relationships					

COBIT Domain	COBIT process		Integration risk	Interoperability risk	Security risk	Scalability risk	Retrofit risk
Align, Plan and Organise	APO09	Manage Service Agreements					
	APO10	Manage Suppliers					
	APO11	Manage Quality					
	APO12	Manage Risk	x	x	x	x	x
	APO13	Manage Security			x		
Build, Acquire and Implement	BAI01	Manage Programmers and Projects					
	BAI02	Manage Requirements Definition					
	BAI03	Manage Solutions Identification and Build					
	BAI04	Manage Availability and Capacity		x		x	
	BAI05	Manage Organisational Change Enablement					
	BAI06	Manage Changes					
	BAI07	Manage Change Acceptance and Transitioning					
	BAI08	Manage Knowledge		x			
	BAI09	Manage Assets					
	BAI10	Manage Service Agreements					

COBIT Domain	COBIT process		Integration risk	Interoperability risk	Security risk	Scalability risk	Retrofit risk
Deliver, Service and Support	DSS01	Manage Operations					
	DSS02	Manage Service Requests and Incidents					
	DSS03	Manage Problems					
	DSS04	Manage Continuity					
	DSS05	Manage Security Services					
	DSS06	Manage Business Process Controls					
Monitor, Evaluate and Assess	MEA01	Monitor, Evaluate and Assess Performance and Conformance	x	x	x	x	x
	MEA02	Monitor, Evaluate and Assess the System of Internal Control					
	MEA03	Monitor, Evaluate and Assess Compliance With External Requirements			x		

Table 5.2 Operational risks mapped against COBIT 5

COBIT Domain	COBIT process		Planning risk	Design risk	Build/Setup/ Configure risk	Operational risk	Maintenance risk
Evaluate, Direct and Monitor	EDM01	Ensure Governance Framework Setting and Maintenance					
	EDM02	Ensure Benefits Delivery					
	EDM03	Ensure Risk Optimisation					
	EDM04	Ensure Resource Optimisation					
	EDM05	Ensure Stakeholder Transparency					
Align, Plan and Organise	APO01	Manage the IT Management Framework					
	APO02	Manage Strategy					
	APO03	Manage Enterprise Architecture					
	APO04	Manage Innovation					
	APO05	Manage Portfolio					
	APO06	Manage Budget and Costs					
	APO07	Manage Human Resources	x				
	APO08	Manage Relationships					

COBIT Domain	COBIT process		Planning risk	Design risk	Build/Setup/ Configure risk	Operational risk	Maintenance risk
Align, Plan and Organise	APO09	Manage Service Agreements					
	APO10	Manage Suppliers					
	APO11	Manage Quality					
	APO12	Manage Risk					
	APO13	Manage Security					
Build, Acquire and Implement	BAI01	Manage Programmers and Projects	x	x			
	BAI02	Manage Requirements Definition	x				
	BAI03	Manage Solutions Identification and Build	x	x			
	BAI04	Manage Availability and Capacity			x	x	x
	BAI05	Manage Organisational Change Enablement					
	BAI06	Manage Changes			x		
	BAI07	Manage Change Acceptance and Transitioning					
	BAI08	Manage Knowledge	x	x	x	x	x
	BAI09	Manage Assets					
	BAI10	Manage Service Agreements					

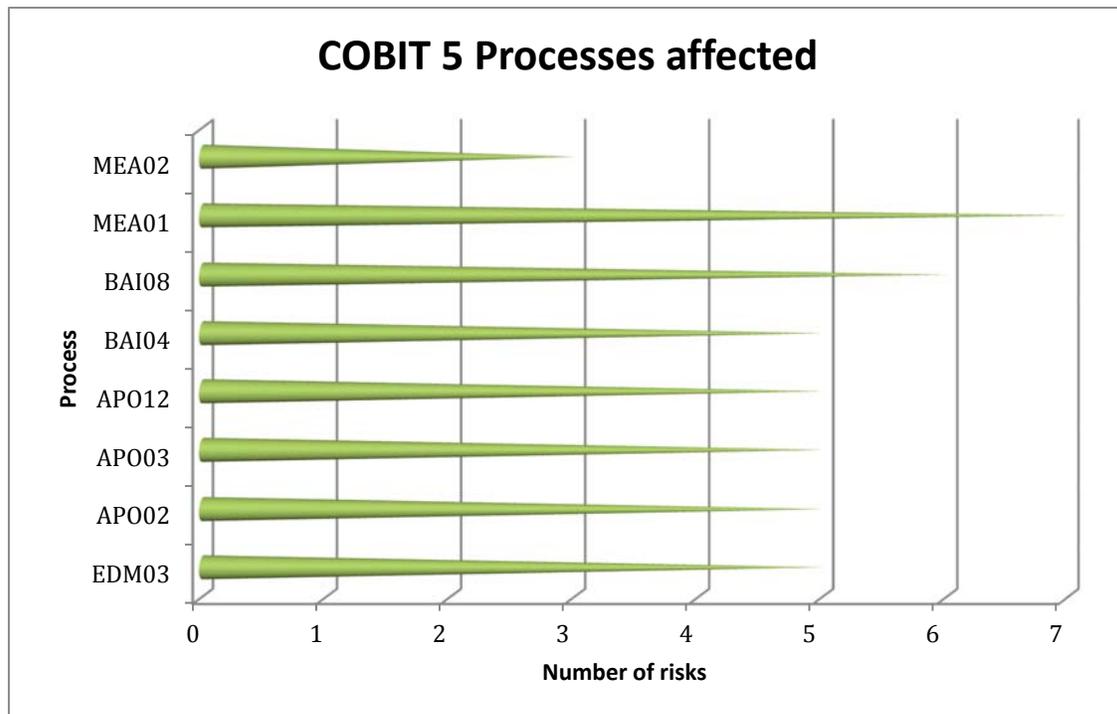
COBIT Domain	COBIT process		Planning risk	Design risk	Build/Setup/ Configure risk	Operational risk	Maintenance risk
Deliver, Service and Support	DSS01	Manage Operations			x	x	
	DSS02	Manage Service Requests and Incidents					
	DSS03	Manage Problems					x
	DSS04	Manage Continuity					x
	DSS05	Manage Security Services					
	DSS06	Manage Business Process Controls					
Monitor, Evaluate and Assess	MEA01	Monitor, Evaluate and Assess Performance and Conformance	x	x			
	MEA02	Monitor, Evaluate and Assess the System of Internal Control	x	x	x		
	MEA03	Monitor, Evaluate and Assess Compliance with External Requirements	x				

5.3.1 Processes most affected by implementing big data

The mapping of the strategic and operational risks related to implementing big data against the COBIT 5 processes, as done in Tables 5.1 and 5.2, indicates that some of the COBIT 5 processes are affected significantly when big data is implemented. The implementing of big data is clearly something that will have an implication for IT governance and management, based on this conclusion.

In order to identify which of the COBIT processes must be governed and managed specifically when implementing big data, the processes most affected by big data must be identified. In Figure 5.1 below the COBIT 5 processes most affected by big data are illustrated graphically, indicating which COBIT 5 processes require extra care when governing and managing the implementation of big data. For the purpose of this study a process qualifies as being significantly affected by the implementation of big data if it is affected by more than two risks.

Figure 5.1 Illustration of the impact of strategic and operational risks on COBIT 5 processes



The mapping in section 5.3 indicates that in implementing big data, business managers must take extra care in governing the following COBIT 5 processes to ensure sufficient management and governance over the IT investment made when implementing big data:

- Ensure Risk Optimisation (EMD003)
- Manage Strategy (AP002)
- Manage Enterprise Architecture (AP003)
- Manage Risk (AP012)
- Manage Availability and Capacity (BIA04)
- Manage Knowledge (BIA08)
- Monitor, Evaluate and Assess Performance and Conformance (MEA01)
- Monitor, Evaluate and Assess the System of Internal Control (MEA02)

Managing these processes efficiently will ensure that management respond to the strategic and operational risk specifically related to big data in an appropriate manner. This will result in better alignment of the business and IT goals of the organisation.

6 Conclusion

Big data is an emerging technology of which the foremost concern and challenge is governance (Buytendijk & Heiser, 2013). In order for technology to be governed effectively the risks of using the technology must be identified and defined.

The purpose of this study was to identify and discuss specific IT risks, strategic and operational to provide guidance to business managers on how to effectively govern these risks related to big data. This need for such a study was identified after undertaking an extensive literature review and concluding that previously mostly general risks have been discussed in literature, but not many discussions on the specific risks related to big data are available.

Identifying and understanding of the risks are however of little value to an organisation without first understanding the term big data and evaluating whether big data will indeed be a suitable solution for their organisation. Therefore this study started with first defining the term big data. Big data was defined as: using high volumes of complex and valuable data created at a high velocity from a variety of sources to add value to the business by increasing the usefulness of the information used in decision making and giving the business a competitive advantage. From this definition it was evident that big data has specific characteristics (volume, velocity, variety, value and complexity). These characteristics result in benefits to the organisation unique to the use of big data. Both the characteristics and the benefits were then discussed in this study. By using this approach it provided a basis to define the business imperatives distinctive of an organisation that will be able to benefit from the use of big data. These business imperatives are the following: proactive management, innovation, agility and scalability. The business imperatives will not only serve as an indication of whether or not an organisation can benefit from the use of big data, but it will also assist the organisation in achieving alignment between its IT and business goals related to big data.

Business imperatives are influenced by IT risks; therefore the business imperatives were used as a basis to identify specific strategic IT risks related to the use of big

data. The strategic risks related to the use of big data were identified using five categories, namely: (1) integration risks, (2) interoperability risks (which were further divided into skills risk and compatibility risk), (3) security risk, (4) scalability risk and (5) retrofitting risk. By identifying and discussing big data specific risks related to each of the strategic risk categories it provides guidance to organisations regarding specific risks that should be appropriately addressed and mitigated to ensure that the strategic goals of the organisation are reached when implementing and using big data. This in return will result in better alignment between the IT goals and the business goals of the organisation.

IT risks however consists of two overall categories of which strategic risks are one and operational risks the other. In the next part of the study the operational risks related to the use of big data were discussed. Operational risks with regard to big data are, however, linked more to the analysis tools used to perform the analysis of the data. This study did not deal with the specific tools available to analyse big data and therefore only general operational risks were discussed. The only category of operational risks that are directly affected by the strategic risks is the planning risks and these risks were discussed based on the categories of strategic risks as were identified earlier in the study. This approach provides business managers with guidance on risks that must be specifically addressed during the planning phase of the implementation of big data to ensure that the strategic goals are aligned with the IT goals of the organisation.

Although the identification and discussion of the strategic and operational risks provides business managers with some degree of guidance with regards to governing the implementation and use of big data it is the effect that these risks have on the processes of the COBIT 5 framework that will provide business managers with a more valuable form of guidance on how to govern the implementation and use of big data. In the last part of the study the strategic and operational risks were mapped against the COBIT 5 framework in order to identify the COBIT 5 processes most affected by big data. From the mapping it was concluded that there are eight processes in the COBIT 5 framework that are most affected by big data (EDM03, APO02, APO03, APO12, BAI04, BAI08, MEA01 and MEA02). By identifying these processes most affected by big data business managers are given guidance on

which processes must be carefully managed when implementing and using big data to ensure better alignment of IT and business goals.

Through this study it became evident that management will have to consider the risks specifically related to big data carefully to ensure that alignment between business and IT goals is achieved. By first defining the business imperatives distinctive of an organisation that will benefit from the use of big data and then using these business imperatives as basis to identify the strategic and operational risks related to big data, this study serves as guidance to organisations when considering the use of big data. It further provides guidance on specifically which COBIT 5 process should receive extra care when implementing and using big data to ensure the alignment of business and IT goals.

7 References

Anderson, R. & Roberts, D. 2012. *Big Data Strategic Risks and Opportunities*. [Online]. Available: http://www.crowehorwath.net/uploadedFiles/Crowe-Horwath-Global/tabbed_content/Big%20Data%20Strategic%20Risks%20and%20Opportunities%20White%20Paper_RISK13905.pdf [2013, May 20]

Badenhorst, M. 2009. *Making sense of IT Governance: The Implications of King III*. [Online]. Available: http://pdf.aminer.org/000/244/975/a_governance_model_for_managing_outsourcing_partnerships_a_view_from.pdf [2014, June 06]

Bertolucci, J. 2013. *Information Week. Big Data: A Practical Definition*. [Online]. Available: <http://www.informationweek.com/big-data/news/big-data-analytics/big-data-a-practical-definition/240160412> [2013, September 04]

Beyer, M.A., Lapkin, A., Gall, N., Feinberg, D. & Sribar, V.T. 2011. *Gartner. 'Big Data' Is Only the Beginning of Extreme Information Management*. [Online]. Available: <http://my.gartner.com/portal/server.pt?open=512&objID=260&mode=2&PageID=3460702&resId=1622715&ref=QuickSearch&stkhw=big+data+risks> [2013, July 08]

Böcker, K. & Klüppelberg, C. 2008. *Multivariate Models for Operational Risk*. [Online]. Available: <http://mediatum.ub.tum.de/doc/1079220/1079220.pdf> [2014, December 04]

Bondi, A.B. 2000. Proceedings of the 2nd international workshop on Software and performance. *Characteristics of scalability and their impact on performance*. [Online]. Available: http://delivery.acm.org/10.1145/360000/350432/p195-bondi.pdf?ip=146.232.43.52&id=350432&acc=PUBLIC&key=646D7B17E601A2A5%2EC011CE1E941E2524%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=541122672&CFTOKEN=36266527&__acm__=1408715730_112bad3f8769e6a9794cd684dd350d7b [2014, August 22]

Boshoff, W. 2012. IT Governance and IT Assurance. Masters in Commerce (Computer Auditing) lecture slides, Stellenbosch University, Stellenbosch.

Boshoff, W. 2013. IT Governance and IT Assurance. Masters in Commerce (Computer Auditing) lecture slides, Stellenbosch University, Stellenbosch.

Bromiley, P., Rau, D. & McShane, M.K. 2014. *Can Strategic Risk Management Contribute to Enterprise Risk Management? A Strategic Management Perspective*. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512477 [2014, December 04]

Brown, B., Chui, M. & Manyika, J. 2011. *McKinsey Quarterly. Are you ready for the era of 'big data'?*. [Online]. Available: http://www.t-systems.com/solutions/download-mckinsey-quarterly-/1148544_1/blobBinary/Study-McKinsey-Big-data.pdf [2012, November 20]

Buytendijk, F. & Heiser, J. 2013. *Gartner. Privacy and Ethical Concerns Can Make Big Data Analytics a Big Risk Too*. [Online]. Available: <http://my.gartner.com/portal/server.pt?open=512&objID=260&mode=2&PageID=3460702&resId=2358315&ref=QuickSearch&stkw=big+data+risks> [2013, July 08]

Chen, H., Chaing, R.H.L. & Storey, V.C. 2012. *MIS Quarterly. Business Intelligence and Analytics: From Big Data to Big Impact*. [Online]. Available: <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=4fd673ad-92ac-4fc8-baf8-263d7ac579d3%40sessionmgr111&vid=2&hid=103> [2014, April 30]

De Haes, S., Van Grembergen, W. & Debreceeny, R.S. 2013. *Journal of Information Systems. COBIT 5 and Enterprise Governance of Information Technology: Building Blocks and Research Opportunities*. [Online]. Available: <http://web.b.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?sid=5f2535d7-4fe6-4b20-980b-9566db543eea%40sessionmgr114&vid=1&hid=103> [2014, September 10]

Digging Up Value with Big-Data Mining. 2014. [Online]. Available: http://www.isc-events.com/bigdata14/isc_blog/items/digging-up-value-with-big-data-mining.html [2014, December 04]

From Insight to Impact. Unlocking opportunities in big data. 2013. CGMA Report.

Gantz, J. & Reinsel, D. 2011. *Extracting Value from Chaos*. [Online]. Available: <http://www.itu.dk/people/rkva/2011-Fall-SMA/readings/ExtractingValuefromChaos.pdf> [2013, November 06]

Gartner. 2009. *IT governance must be driven by corporate governance*. [Online]. Available: http://www.gartner.com/it/content/1229500/1229528/it_governance_must_be_driven_corp_gov.pdf [2013, September 12]

Géczy, P. 2014. *Big data characteristics* [Online]. Available: http://macrotheme.com/yahoo_site_admin/assets/docs/8MR36Pe.97110828.pdf [2014, May 21]

Goosen, R. 2012. *The development of an integrated framework in order to implement information technology governance principles at a strategic and operational level for medium-to-large sized South African business*. Unpublished masters assignment. Stellenbosch: Stellenbosch University.

Goosen, R. & Rudman, R. 2013. *South African Journal of Business Management. The development of an integrated framework in order to address King III's IT governance principles at a strategic level*. [Online]. Available: http://reference.sabinet.co.za/webx/access/electronic_journals/busman/busman_v44_n4_a8.pdf [2014, June 25]

Green, R. & Panzer, M. 2014. *The Interplay of Big Data, WorldCat and Dewey*. [Online]. Available: <https://journals.lib.washington.edu/index.php/acro/article/view/14677/12319> [2014, May 28]

Hardy, G. 2006. Information Systems Controls Journal. *Guidance on Aligning COBIT, ITIL and ISO 17799*. [Online]. Available: <http://m.isaca.org/Journal/Past-Issues/2006/Volume-1/Documents/jpdf0601-Guidance-on-Aligning.pdf> [2014, September 10]

Hobday, M., Davies, A. & Prencipe, A. 2005. Industrial and Corporate Change. *Systems integration: a core capability of the modern corporation*. [Online]. Available: <http://icc.oxfordjournals.org/content/14/6/1109.short> [2014, December 03]

ISACA. 2013. *COBIT 5*. [Online]. Available: <http://www.isaca.org/COBIT/Pages/COBIT-5-Framework-product-page.aspx> [2013, June 13]

Jackson, R.A. 2013. *Big Data**. [Online]. Available: <https://iaonline.theiia.org/big-data> [2013, November 15]

JASON. 2008. *Data Analysis Challenges*. [Online]. Available: <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADB346913> [2014, September 12]

Johnson, J., Neff, T. & Stuart, A. 2012. *Big Data The Risks and Rewards Locked in Vast Oceans of Data*. [Online]. Available: <http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/big-data-oceans.pdf> [2013, May 19]

Kaisler, S., Armour, F., Espinosa, J.A. & Money, W. 2013. *Big Data: Issues and Challenges Moving Forward*. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6479953&tag=1> [2013, October 31]

Kart, L., Heudecker, N. & Buytendijk, F. 2013. *Gartner. Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype*. [Online]. Available: [http://www.gartner.com/document/2589121?ref=QuickSearch&stkw=big%20data%](http://www.gartner.com/document/2589121?ref=QuickSearch&stkw=big%20data%20)

20adoption&refval=145513152&qid=afa0ccca1398979e95a4397278f71660 [2014, October 30]

King Report on Governance for South Africa 2009. 2013. [Online]. Available: <http://african.ipapercms.dk/IOD/KINGIII/kingiiiireport/?Purge=true> [2014, March 07]

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S. & Kruschwitz, N. 2011. *MIT Sloan Management Review. Big Data, Analytics and the Path From Insights to Value*. [Online]. Available: http://www.ibm.com/smarterplanet/global/files/in_idea_smarter_computing_to_big-data-analytics_and_path_from_insights-to-value.pdf [2013, August 21]

Longbottom, C. 2012. *Quorcirca. How to make sense of the big data universe*. [Online]. Available: <http://www.quocirca.com/media/articles/072012/728/CW%20-%20big%20data.pdf> [2014, March 14]

Mayer-Schonberger, V. & Cukier, K. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt Publishing Company.

McAfee, A. & Brynjolfsson, E. 2012. *Harvard Business Review. Big Data: The Management Revolution*. [Online]. Available: <http://automotivedigest.com/wp-content/uploads/2013/01/BigDataR1210Cf2.pdf> [2014, May 23]

Moore's Law and Intel Innovation. 2013. [Online]. Available: <http://www.intel.com/content/www/us/en/history/museum-gordon-moore-law.html?wapkw=moore> [2013, May 20]

Russom, P. 2011. *TDWI Best Practices Report. Big Data Analytics*. [Online]. Available: ftp://129.35.224.12/software/tw/Defining_Big_Data_through_3V_v.pdf [2013, July 05]

Schmidt, E. 2010. Google, Privacy and the New Explosion of Data. Techonomy conference. 4 August. Lake Tahoe.

Steve, L., Eric, L., Rebecca, S., Michael, S.H. & Nina, K. 2011. *MITSloan Management Review. Big Data, Analytics and the Path From Insights to Value* [Online]. Available:
http://www.ibm.com/smarterplanet/global/files/in_idea_smarter_computing_to_big-data-analytics_and_path_from_insights-to-value.pdf [2013, March27].

TechAmerica Foundation. 2012. *Demystifying Big Data A Practical Guide To Transforming The Business of Government*. [Online]. Available:
<http://www.greenplum.com/sites/default/files/techAmerica-BigDataReport-FINAL.pdf> [2013, May20]

Terblanche, J. 2011. *An IT Governance framework for the public sector*. Unpublished masters assignment. Stellenbosch: Stellenbosch University.

Van Oosterhout, M., Waarts, E. & Van Hillegersberg, J. 2006. European Journal of Information Systems. *Change factors requiring agility and implications for IT*. [Online]. Available:
<http://search.proquest.com.ez.sun.ac.za/docview/218787362/fulltextPDF?accountid=14049> [2014, June 25]

Vokurka, R.J. & Fliedner, G. 1998. Industrial Management & Data Systems. *Journey towards agility*. [Online]. Available:
<http://www.emeraldinsight.com/journals.htm?articleid=849907&show=abstract> [2014, July 10]

Walker, M. 2012. *Structured vs. Unstructured Data: The Rise of Data Anarchy*. [Online]. Available: <http://www.datasciencecentral.com/profiles/blogs/structured-vs-unstructured-data-the-rise-of-data-anarchy> [2013, November 06]

Way, S.C. & Yuan, Y. 2012. *Towards a Context-Aware Multi-Party Emergency Coordination System Framework*. [Online]. Available:
<http://www.iscramlive.org/ISCRAM2012/proceedings/296.pdf> [2014, December 01]

Worrel, J.L. & Bush, A.A. 2007. *Perceptions of Information Technology Risk: A Delphi Study*. [Online]. Available: <http://ashley-bush.com/pdf/AMCIS2007WorrellBush.pdf> [2014, December 04]