



Analysis of Data Governance Implications on Big Data

Lomso Trom^(✉) and Johannes Cronje

Cape Peninsula University of Technology, Cape Town, South Africa
johannes.cronje@gmail.com

Abstract. This paper explores the governance dynamics of big data, the possible governance challenges and gaps that come with the adoption of big data in organizations. On its arrival, big data technology brought along a massive hype and disruption in the world of business. The unique ability to process large quantities of information, analyze it and make well informed data driven decisions is the main reason for such a disruption. Globally businesses, governments organizations have come to realize the gains of big data and they are currently developing plans to implement and expand big data technologies. Even though these organizations are busy exploring different plans to leverage big data, little or no attention is paid to big data governance and thus lack of big data governance remains the ghost that continues to haunt companies. This paper focuses on the arising governance dynamics on big data. Apart from the issue of increasing data scope, big data also complicates the governance processes, policies, people roles and these necessitates the need for new approach, methods, controls and policies that are best suited to govern big data.

Keywords: Big data · Data governance · Big data governance

1 Introduction

Data is growing at a very rapid rate, this is due to larger, medium and small companies generating, using, and storing large amounts of data. Currently there is more data types and sources than ever before with this variety volume are expanding daily [32].

Big data definitions have suddenly evolved and this has brought confusion to some entities. There is an ongoing revolution, this is the generation and analysis of massive high dimensional data quantities and this revolution continuous to shape our lives. Organizations are engaged in research, to explore how big data can be used to solve long standing business challenges.

Organizations are now owning data with enormous volume, veracity and rapid velocity which then beyond the capacity traditional data processing systems. According Wu [49] the Big Data era requires organizations to search for ground breaking technologies to process the high velocity information to produce competitive intelligence.

Modern organizations are not only interested to know what happened and why it happened, but they also eager to know what is happening at this moment and what is

going to occur in the future. Fernandez [13] claims that the common understanding of big data as just ‘data’ has led to experts questioning how is used and ethical concerns. Fernandez [13] further suggest that even though big data carries several benefits for different companies, it is not immune to criticism and must be thoroughly for ethical implications.

According to Ghosh [17] insufficient knowledge about the big data subject, big data privacy issue, inability to handle large amounts of data are part of the most common challenges that are faced by big data implementations. Fatt and Ramadas [12] concur with the latter, and further emphasize that issues such as privacy, security, standards and governance have to be addressed in big data.

Big data necessitates a lot of precautions, controls and governance measures to protect the basic rights of the people and the data asset. These controls include a major shift in regulating big data from collection to the end of its lifecycle [3]. Despite information such as health that may be vulnerable from hackers or theft, the data in large organizations such as telecoms and financial institution is at risk and the clients are unaware [12].

Tesfaye [45] indicates that the specific properties of big data (volume, variety, velocity, veracity, and value) introduce new types of risks and these risks are due to lack of governance, management, architecture, usage, quality, security, and privacy.

This paper focuses on the governance implications surrounding big data. Apart from the issue of increasing data scope, big data also complicates the governance processes, policies, role and challenges prompting for innovative governing methods.

2 Problem Statement

Data remains a vital asset for majority of organization globally, but recently its prominence has increased dramatically. Organizations in various sectors of business have been exploiting data to gain competitive advantage while having good data governance in place to mitigate the risks associated with the use of this data.

Big Data has brought about disruption and this business entities have the opportunity to gain more insight of their industry developments, customers, products and performance. Big data does not only present opportunities but it also present challenges [31].

Suitable solutions must be applied to overcome these challenges and gain valuable insights from big data. Companies therefore need to find ways to cope with massive amounts of unstructured data that demands a shift from old governance processes, policies and methods to govern big data. Even though data is considered an asset, if the environment ill governed, the quality is poor, and inconsistent then that spells a huge loss in value for the company.

Traditional data governance is tailored for structuralized data and that makes it impossible to cater for unstructured data. New technologies have created the demand for a shift from old ways.

3 Literature Analysis

3.1 Traditional Data as an Asset

According to [10] in this era of information, data is recognized as an indispensable enterprise asset and therefore data is the lifeblood of the modern economy. Data has developed become a fundamental part of contemporary businesses and it has reshaped the businesses processes. Stander [44] states that the rapidly changing nature of business encourages organizations to fully exploit all of the assets at their disposal to gain competitive edge over the competitors. One of the major resources for improving productivity, developing new technology and improving processes is data and information.

GlueReply [16] confers that Data is a very valuable commodity and a large number of organizations recognize that their data is an asset, even though less is done to capitalize on the data. The idea of data as an asset is also echoed [34] who states that Data is now a form of capital which is on the same level as financial capital in terms of generating new digital products and services. This data development poses an effect to every company's business strategy, as well as for the supporting computing architecture.

Big data provides an opportunity for businesses to excel when managed effectively by enabling the organization to make sense of this newly produced knowledge and thus extract value from big data. The organizations ability to leverage big data and make key data driven decision will become the defining factor to gain competitive edge [19] This current trend of data-driven decision making has grown to be a captivating area of studies and analysis for researchers [51].

Some of the business benefits of big data as mentioned by [46] is the reduction of doubts by performing comparisons and analyzing many possible scenarios. Tykheev [46] believes that through pattern recognition and addition of fresh ideas can also be generated.

Satish and Yusof [40] revealed big data has been used to provide better customer insight loyalty and behavioral analysis to track data history such as the shopping behavior so that custom offers can be made. Brynjolfsson et al. [2] claims that organizations that have succeeded in their efforts to incorporate data into their decision making processes have managed to significantly improve their overall productivity, achieved higher profitability, improved financial performance and witnessed rise in market share.

3.2 Big Data Defined

Daily, humongous bytes of data are generated from digital media such as videos, pictures, social media posts, website posts, intelligent sensors, sales transaction records, smart phone GPS signals and this is defined as big data [50]. Laudon and Laudon [26] asserts that big data can also be described as data sets that are too large and complex for a regular DBMS to capture, retrieve and analyze.

According to [25] big data is an information technology term which refers to high volume, velocity, variety. That implies big data is a huge dataset with fast produced data and different types of formats. Giant corporations such IBM, Google and Facebook have already realized the huge potential value that is presented by big data. They all profit by utilizing advanced analyzing tools or inventing new business model to extract value from their big data.

ACCA and IMA [1] defines big data as a process of collecting, storing and analyzing large amounts of data. Big data is fast data that is collected, transformed and processed immediately. A common factor in these definitions is the frequent use of words that describe something of a greater scale such as “large”, “humongous”, “massive” and “complex”. The Big Data era requires organizations to search for ground breaking technologies to process the high velocity information to produce competitive intelligence [6].

Currently, big data has several definitions since the phenomenon is both of technical and sociological nature. Most commonly, people make use of the term to describe the piles of data they have been exposed to, consistently generated by our digital lives. It often involves the collection and analysis of data of both large and low value information [24].

3.3 Big Data Revolution

According to [4] large organizations have been ripping the fruits of big data for such a long time and the smaller and medium sized companies are following suites due to cost reduction and provided ease to manage data. Big data is the next generation of data warehousing and business analytics and is poised to deliver top line revenues cost efficiently for enterprises [33].

Schmarzo [41] claims that the big data pushes more of business transformation than Information technology transformation and further asserts that by effectively using big data, company are able to make timely and informed decision than in the past. Apart from the extensive hype that has been spawned by the topic of big data, the primary objective is something that organizations have been aspiring to for such a long time and that is to improve decision making [38].

Big data is constantly generated by everything around us continually. Therefore, it makes the possibility to develop data driven businesses that gather, store and analyze data for improving performance and profitability as well as to solve business challenges and through innovation [24]. Ricci and Kalyva [39] assert that more organizations are following the trend in pursuit of the big data prospects, such as more informed decision making, improved customer insight and value improvement to gain competitive advantage.

LaValle et al. [27] maintains that modern organizations are not only keen to know what happened and why it happened, but they also eager to know what is happening right now and what is most likely to happen in the future. According [48] one of the most probable prospect of big data for businesses and their customers is the opportunity to better understand customers, allowing for more customized offers resulting to increase in sales and customer loyalty.

Big Data is generally defined as factor that can make a huge impact on the organization's operating model at an early stage or late in their big data journey [20]. For companies to effectively leverage big data, the starting point before implementation should be identifying the key business challenges, understand their business model, identify their data sources and learn the data structure [5].

Kaisler et al. [22] suggests that the probable value gained from the exploitation of big data is not just being able to process large volumes of data efficiently, but it is in the ability to analyze the different types of data and new types of data (structured as well as unstructured) in such a way as to develop meaningful insight.

3.4 Data Governance

Data governance is about enabling and encouraging good behavior regarding data, and limiting behaviors that create risks. Data governance is key to understanding the meaning, usage, and quality of big data. Without it, you'll struggle to create the insights and information needed to run a data-driven organization [7].

On the other hand, [43] defines data governance as the effective management of people, processes, and technology that enables the organization to leverage data as a valued enterprise asset. Tait [47] states the Processes for managing data should be well defined and effectiveness, as well as compliance to these processes needs to be monitored constantly. The common denominator to the above definitions is their emphasis on planning, managing people, processes and leading a data driven organizational culture.

Establishing strong governance has been considered as one of the major challenges in big data value creation [30]. Ndamase [36] claims that an effective data governance program supports the development of formal data standards and policies and it also ensures oversight over data so that decision makers are able receive quality, consistent and timely data to respond or act to the challenges and opportunities of the organization.

Ernst & Young [11] confers that good governance plays a critical role in determining the success of big data initiatives in an organization. Ernst & Young [11] further states that good governance involves formal procedures, consistent guidance and sound decision making by the management. There is a lot of policy considerations that need to made by organizations that aspire to capitalize on big data in both either public or private sector [46].

A major challenge that organizations face when trying to govern big data, is that big data is relatively beyond the organizational lines and mainly external. Thus governing becomes a thorny task due to complications such as unclear ownership and custody of data.

There are many ways in which an organization can be governed. Examples of governance mechanisms include strategies, goals, policies, plans and standards. Different governance mechanisms are used to deliver value and minimize risks and governance has an impact on achieving the strategic goals of an organization and therefore differs from the managerial function, which aims to achieve operational goals [36].

3.5 The Governance Risks Associated with Big Data

Whenever a new technology trend like big data is introduced in an organization, new risks accompanying the new trend are bound to appear and they need to be quickly addressed [15]. Smeda [42] concurs that big data is still a developing technology that carries along its own set of distinct risks. The governance aspect big data remains a huge and there is not enough guidance for new organizations.

Privacy

Big data is relatively used to make various decisions and predictions about a lot of entities including people. Granted the sensitive nature of different big data sources, privacy of data and security presents a huge risk, even though some of the data does not consist of personal information [14].

Kaisler et al. [22] claims that at the moment there is an existing gap due lack of regulations or policies that monitor the collection of unstructured data by companies, precisely social media data. When collected, the data might not seem to be personal, but there is a high possibility of the data being traced back to the relevant person and this might be violation of privacy [52].

The use of personal data without individual consent may mean that security is compromised, privacy is invaded and data is misused, it could have detrimental effects on the business such as damage of reputation or legal consequences [23]. Data can be so valuable for the business, but data ownership and privacy issues have to be put into consideration before big data results are used. Data types such as medical data are very complicate and ownership cannot be easily determined. Big data may present intellectual property concerns such as access privileges or copyright, sharing excessive information externally may have severe consequences [11].

According to Khan et al. [21] privacy is a greatest concern in big data, it has become common for some security agencies to use the data about people for their personal gain without being granted permission. Khan et al. [21] suggest that perpetrators should be identified and formal policies that cater for all user privacy challenges be developed to ensure that private data is not be misused or leaked.

Quality

There are three goals of data quality governance. The starting point of data quality improvement is gaining insight on the value of data within the organization, followed by introduction of right levels of control, then comes detecting and prioritizing data issues to resolve [29].

According to [28] a major data quality challenge in big data is the various data types such as (unstructured, semi-structured and structured) drawn from a variety of data sources. Rajpurohit [37] argues a single data quality metric cannot be relevant to all the types, thus unique metrics have to be defined for all types of data type.

Accurate and high quality data is crucial to the success of the business and guarantees that data-driven decisions are mainly based on facts. Furthermore, organizations must also make a decision of where this data fits into the overview organizational value chain in which they are operating to completely uncover and understand the value stored in their data [14].

Security

The rapid increase in volume and the growing demand of data has produced more complications and these cannot be managed by traditional data governance. The introduction of big data was tipped to solve volume related problems. However, further problems manifested and these are security, quality and privacy [35].

Big data requires security solutions with the ability to monitor all access and critical data. This include being able to detect any suspicious patterns or activities when matched against the frequent and normal activities. By nature, Big data consists of numerous instances, layers, and technologies each serving various purposes in sets of big data. Thus, providing maximum security for big data is regarded as a very complex task than when compared to the traditional data security solutions.

Access control is a technique that has been used for a long time in the data space, to ensure the security of a system. In the big data space, the access control issue comes from the fact that only basic measures of access control are used [35].

When working with big data one is required to be proficient in specific skills and abilities. Statistical knowledge is certainly one of these most vital skills, even though the ability to use new processing methods other than traditional statistical techniques is encouraged [8].

The indicated that a lack or shortage in the availability of big data expertise is a major challenge. ACCA and IMA [1] revealed that organizations are yet to upgrade their competencies, and therefore the staff may lack the understanding of handling massive data sets and understand the structure as well as business processes.

4 Conclusion

The rapid growth in big data implementation and initiatives across different sectors of business provides a hint of what the future may look like. To that end, the various risks and challenges that exist or having the potential to arise, must still be addressed. As more organizations adopt big data and encounter these challenges, the above mentioned governing issues such as privacy, security and quality will top the list of priorities.

At the moment big data is amongst popular concepts globally. It has a huge potential and there are various ways to apply and govern it. Even though some businesses and government entities adopting this amazing technology, a large number of entities are sleeping on piles of valuable data with no clue or realization of its value [46]. This supplements the need to study and explore big data governance to cover existing gaps.

In the age of big data, the traditional data governance approach faces a lot of difficulties and it no longer applies. Enaxisconsulting [9] asserts that for an organization be effective in the current big data era, data governance must be entrenched from the principle that all data is ready to be governed, and data that has the greatest potential to provide value to the organization should be governed.

The process of governing big data must not become static or unwillingly performed occasionally. It must become continuous, dynamic initiatives that are able act on new

data collections and the changing data regulations. According to Soares [43] “Big data governance is part of a broader information governance program that formulates policy relating to the optimization, privacy, and monetization of big data by aligning the objectives of multiple functions.”

The governing unit gains new lessons as the business ability to use data matures. The ultimate goal is not to put constraints, but to ensure that the data collected by the organization is relevant, makes sense to the decision makers and that it adds value [18]. Big data has the sufficient potential to help grow businesses, create an advance customer and supplier understanding and optimize the business operations. But for organizations to gain its utmost value and minimize risks, a data governance strategy has to be put in place.

To mitigate and management the risks that big data poses, a big data governance framework with involves the broader organization role players is needed. The role players may be producers or consumers of big data, the governing unit must ensure that a positive atmosphere and culture around governance is initiated.

This paper has outlined existing big data governance dynamics, challenges, opportunities and risks that are may arise from the adoption of big data. The paper seeks expand the limited understanding of the governance dynamics presented by big data.

A comprehensive analysis of these dynamics will result to effective governance of big data with more awareness of the dynamics. The outcomes of this paper could be used to intensify governance of big data in various industries including the public sector.

References

1. ACCA and IMA: DigitalDarwinism: Thriving in the Face of Technology Change (2013). <http://www.accaglobal.co.uk/content/dam/acca/global/PDF-technical/other-PDFs/Five-mins-on-Digital-Darwinism.pdf>
2. Brynjolfsson, E., Hitt, L.M., Kim, H.H.: Strength in numbers: how does data-driven decisionmaking affect firm performance? (2011)
3. Broeders, D., Schrijvers, E., van der Sloot, B., van Brakel, R., de Hoog, J., Hirsch Ballin, E.: Big Data and security policies: towards a framework for regulating the phases of analytics and use of Big Data. *Comput. Law Secur. Rev.* **33**(3), 309–323 (2017)
4. Buttler, P.: 10 Challenges to big data Security and Privacy - Dataconomy (2017). <http://dataconomy.com/2017/07/10-challenges-big-data-security-privacy/>. Accessed 5 June 2018
5. CGMA: From insight to impact. Unlocking the opportunities in big data (2013)
6. Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: from big data to big impact. *MIS Q.* **36**(4), 1163–1170 (2012)
7. Colibri and MapR: The Need for Big Data Governance—MapR (2016). <https://mapr.com/resources/need-big-data-governance/>. Accessed 15 June 2018
8. De Mauro, A., Greco, M., Grimaldi, M.: What is big data? A consensual definition and a review of key research topics. In: AIP Conference Proceedings, vol. 1644, pp. 91–102 (2015)

9. Enaxisconsulting: Governing and managing data in the big data era – Enaxis Consulting (2017). <http://www.enaxisconsulting.com/governing-and-managing-data-in-the-big-data-era-blog/>. Accessed 6 June 2018
10. Enofe, M.: Data Management in an Operational Context: A study at Volvo Group Trucks Operations. Lund University School of Economics and Management, Sweden (2017)
11. Ernst & Young: Big data Changing the way businesses compete and operate Insights on governance, risk and compliance [ebook], pp. 11–14 (2014). http://www.ey.com/.../changing_the_way_businesses_operate/.../EY-Insights-on-GRC-Big-d. Accessed 11 June 2018
12. Fatt, Q., Ramadas, A.: The usefulness and challenges of big data in healthcare. Healthcare-communications. imedpub.com (2018). <http://healthcare-communications.imedpub.com/the-usefulness-and-challenges-of-big-data-in-healthcare.pdf>. Accessed 25 Oct 2018
13. Fernandez, A.: Shaping the future use of big data: towards an ethical use of big data technologies in online marketing. Masters. University of Twente (2017)
14. Franks, B.: Characterisation of a dual detector probe for medical physics applications. Masters thesis. University of Wollongong (2012)
15. Gerber, P.: Addressing the incremental risks associated with social media by using the COBIT 5 control framework. Masters thesis. University of Stellenbosch (2015)
16. GlueReply: The Valuation of Data as an Asset: a Consumption-Based Approach. Reply, 12 June 2015 (2014). www.reply.com/Documents/13903_img_The-valuation-of-data-as-an-asset.pdf. Accessed 05 June 2018
17. Ghosh, P.: Data Management Trends in 2018 - DATAVERSITY. DATAVERSITY (2018). <http://www.dataversity.net/data-management-trends-2018/>. Accessed 25 Oct 2018
18. Hannula, J.: Big data management: the challenges of governance—SmartShift. Smartshift-tech (2015). <http://smartshifttech.com/big-data-management-challenges-governance>. Accessed 15 June 2018
19. Henke, N., et al.: The age of analytics: competing in a data-driven world. McKinsey Global Institute (2016)
20. Hurwitz, J., Nugent A., Halper, F. Kaufman, M.: Big Data for Dummies. Wiley, Hoboken (2013). <http://www.wiley.com/>
21. Khan, N., et al.: Big data: survey, technologies, opportunities, and challenges. Sci. World J. (2014)
22. Kaisler, S., Armour, F., Espinosa, J., Money, W.: Big data: issues and challenges moving forward. In: 46th Hawaii International Conference on System Sciences (2013)
23. Kshetri, N.: Big data's impact on privacy, security and consumer welfare. Telecommun. Policy **38**(11), 1134–1145 (2014)
24. Kuurila, J.: The role of big data in Finnish companies and the implications of big data on management accounting. Master thesis, University of Jyväskylä (2016)
25. Laney, D.: 3-D data management: controlling data volume, velocity and variety. META Group Res. Note **6**, 1 (2001)
26. Laudon, K.C., Laudon, J.P.: Management Information Systems, Managing the digital firm, 13th edn. Prentice Hall, Upper Saddle River, pp. 36–37, 498–499, 254–269 (2014)
27. LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big data, analytics and the path from insights to value. MIT Sloan Manag. Rev. **52**, 20–32 (2011)
28. Liell-Cock, S., Graham, J., Hill, P.: IT Governance aligned to KING III (2009)
29. Loshin, D.: Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. Morgan Kaufmann, Burlington (2013)
30. Marchibroda, J.: A Policy Forum on the Use of big data in Health Care (2013)

31. Mathur, N., Purohit, R.: Issues and challenges in convergence of big data, cloud and data science. *Int. J. Comput. Appl.* **160**(9), 7–12 (2017). <https://pdfs.semanticscholar.org/43f1/c1fddaf8acf6e6aaf1ec5fbce7236ac228e9.pdf>. Accessed 24 Oct 2018
32. Mathes, C.A.: Big data has unique needs for information governance and data 73 quality. *J. Manag. Sci. Bus. Intell.* **1**(1), 12–20 (2016)
33. Minelli, M., Chambers, M., Dhiraj, A.: Business Analytics. *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, pp. 99–125 (2013)
34. MIT Technology Review: The Rise of Data Capital (2015)
35. Moreno, J., Serrano, M., Fernández-Medina, E.: Main issues in big data security. *Future Internet* **8**(3), 44 (2016)
36. Ndamase, Z.: The impact of data governance on corporate performance: the case of a Petroleum Company. University of Cape Town (2018)
37. Rajpurohit, A.: Must-Know: what are common data quality issues for Big Data and how to handle them? Kdnuggets.com (2017). <https://www.kdnuggets.com/2017/05/must-know-common-data-quality-issues-big-data.html>. Accessed 14 Aug 2018
38. Regalado, A.: Data and decision making. *MIT Technol. Rev.* **117**(2), 61–63 (2014)
39. Ricci, A., Kalyva, F.: Unlocking the full potential of Big Data: a change management approach. Master's Programme in Management. Lund University (2015)
40. Satish, L., Yusof, N.: A review: big data analytics for enhanced customer experiences with crowd sourcing. *Procedia Comput. Sci.* **116**, 274–283 (2017)
41. Schmarzo, B.: *Big Data: Understanding How Data Powers Big Business*. Wiley, Indianapolis (2013)
42. Smeda, J.: Benefits, business considerations and risks of big data. Masters. University of Pretoria (2015)
43. Soares, S.: A framework that focuses on the data in big data governance. *IBM Data Management* (2012)
44. Stander, J.: The modern asset: big data and Information Valuation. Stellenbosch University (2015). <https://scholar.sun.ac.za/handle/10019.1/97824>. Accessed 13 June 2018
45. Tesfaye, B.: What is the influence of big data and analytics on management control systems? Masters. Radbound University (2017)
46. Tykheev, D.: Big data in marketing. Undergraduate. Saimaa University of Applied Sciences (2018)
47. Tait, A.: 5 big data security challenges—learning tree blog. *Learning Tree Blog* (2016). <https://blog.learningtree.com/five-big-data-security-challenges/>. Accessed 5 Jun. 2018
48. Wamba, S.F., Akter S.: Big data analytics in E-commerce: a systematic review and agenda for future research (2016)
49. Wu, S.: Big Data Processing with Hadoop. Turku University of Applied Sciences (2015)
50. Zicari, R.V.: Big Data: Challenges and Opportunities. This is big data (2012). <http://www.odtms.org/wp-content/uploads/2013/07/Big-Data.Zicari.pdf>
51. De Coninck, N.: The relationship between Big Data and operations research. Ghent University (2017)
52. Cukier, K., Mayer-Schönberger, V.: The rise of big data. *foreign affairs*, pp. 28–40, May/June (2013)