Organizational Development and Big Data: Factors that impact successful Big Data Implementations

A dissertation submitted

by
Talha Ashraf

to
Benedictine University

in partial fulfillment, of
the requirements for the
degree of

Doctor of Philosophy
in
Organization Development

This dissertation has been
accepted for the faculty of
Benedictine University.

Ramkrishnan V Tenkasi

Ramkrishnan (Ram) V. Tenkasi, PhD May 17, 2017
Chair

Robert Verner, PhD May 17, 2017
Committee member

William B. Brock, PhD, CPA May 17, 2017
Committee member

Abstract

Big data is a term/concept that is often misunderstood in the field due to rapid evolution of its definition, confusion among executives of what it is, how it works and challenges in working with current statistical modeling for datasets comprising of structured and unstructured data (Gandomi & Haider 2014). It is for those reasons that meaningful results around Big Data are not clearly established and realized. Furthermore, Barker et. al (2016) concluded, Big Data remains a fragmented, early-stage domain of research in terms of theoretical grounding, methodological diversity and empirically oriented work. Hence, we observe Big Data implementations taking place in specific organization functions, settings and business cases. This research provides empirical evidence through use of a mixed methods study that opinions on Organizational Development characteristics (change, collaboration), organizational competencies (technical skills, computing resources and support) and change management expertise (enabling change, initiatives for betterment of organization) are associated with higher perceptions as part of measuring successful Big Data Implementations via new products, services, competitive advantage, patents or furthering of business strategy. All, aside from technical skills and computing resources, very much align with the field of Organizational Development (O.D), whose mission it is to develop, create efficiencies and effective organizations by managing change, collaborating, involving business areas and focusing on betterment of the organization.

Keywords: *Organizational Development, Big Data, Big Data Implementations*

Dedication

To my parents and my sister.

Acknowledgments

I am thankful to the Almighty for everything and words cannot do justice to describe the passage from where I was to where I am today.

My sincere appreciation to my chair, Dr. Ramkrishnan Tenkasi, for his mentorship, leadership, direction and valuable insights. My committee members, Dr/s. William and Robert for their guidance throughout my dissertation journey. My cohort for their support and my wife for encouragement.

Here is to the next phase of learning and growth.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

"We're really just getting under way. But the march of quantification, made possible by enormous new sources of data, will sweep through academia, business and government. There is no area that is going to be untouched."

-Gary King, Director Harvard Institute for Quantitative Social Science

About 400 years ago, Galileo observed that "the book of nature is written in the language of mathematics". This is very much the case today with the enormous amount of data sources and actual volume of data being made available (McAfee and Brynjolfsson 2012; Jagadish et al 2014; Manyika et al., 2011; Kiron and Shockley, 2011). As the above scholars rightfully point out, it is not the abundance of data that is new but rather the myriad of many factors, which include:

1. Explosion in the formation of new data streams fueled by more rapid/real time data made available
2. The numerous devices (hardware) that are now able to connect to the Internet
3. The software and computing power that allows to us to mine and harness input from those devices where once not available, and finally
4. The cost effectiveness, speed and abundance of Internet connectivity present across the world for us to subscribe to.

This simple platform and technological advances have made additions and accessibility of that data a reality. This explosion of sorts benefits us with gleaning knowledge, insights and opportunities (Xu et al 2015; Chen et al 2012; Forrester 2012; Wamba et. al 2015). On the one hand, the collection, analysis, and amalgamation of this data is creating challenges and questioning current practices, ethics, procedures and processes (Mantelero and Vaciago, 2015; McAfee and Brynjolfsson, 2012; Gudivada et al., 2015; Punathambekar and Kavada 2015) while on the other hand, it is creating opportunities by creating new business streams (Wamba et. al, 2015). One such business stream, however, deals with organizations realizing their own value of data housed and sharing that to create information based products and services for transactional (profits/money) or strategic value of some kind (Wixom, 2014).

To put Big Data into perspective, roughly ~2.5 exabytes of data is being created every day and that number is doubling every 40 months (McAfee and Brynjolfsson, 2012). Similarly, other reports, such as from Halaweh and Massry (2015) estimate ~5 exabytes of data created every two days and a grand total of 8 Zettabytes by 2015 (equivalent of 18 million Libraries of Congress), which is consistent with McAfee and Brynjolfsson's findings. Furthermore, the sheer variety that is presented with this volume of data, per blog.qmee.com, an online shopping companion that can be installed on your browser, is noteworthy. In a mere 60 seconds:

1. 5 million videos are viewed on YouTube,

2. 293k statuses are updated on Facebook,
3. 2.66 million searches are conducted on Google, and
4. 138.8 million emails are sent (of course, spam included).

There are more stats available from various other internet applications such as Amazon.com, Snapchat, Skype, iTunes, Twitter and Pinterest (see Figure 1) that further highlight the variety of this voluminous data. Keep in mind that this time-box data capture is not restricted only to Internet-ready applications or specific industries. Virtually all industries have their own variations and mechanisms of data collection, use and value creation of products/services. Industries such as technology, education, healthcare, insurance, finance/banking, commerce and even retail are investigating the amount of data that they have, can collect and use. However, what is not clear is the focus and what that data can be useful for.

Big Data in Organizations across industries

Kuketz (2013) showed 13 new business models created by big data analytics. Here we examine a sample of industries that shows the value that Big Data currently has and is bringing to the table:

Technology

Netflix analyzes millions of real time data points that its viewers create, thus helping the firm determine if a pilot will become a successful show (Xu et al., 2015). Facebook hosts over 500 terabytes of data everyday – including uploaded photos, likes and users' posts (Provost and Fawcett, 2013). Google alone contributed roughly $54 billion to the US economy in 2009 (Labrinidis and Jagadish, 2012). Akamai Technologies Inc, a leading global Content Delivery Network provider collects and analyzes petabytes of data every day to help its customer base with cloud performance and security initiatives. Amazon, another ecommerce/technology company, utilizes its various data points to ensure personalized experiences for its client base.

Healthcare

Burg (2014) argued that Big Data can enable a better and transparent healthcare system. Allouche (2014) identified cost saving and unnecessary procedure reducing capabilities from Big Data. Tormay (2015), identifies pharmaceutical R&D as the engine that fuels the pharmaceutical industry. He claims this engine has been declining in productivity over the last 20 years with increasing costs, demands for better standard care, and concomitant productivity challenges. He believes that data, specifically the fast and voluminous nature along with technological advances will help revitalize this engine. Furthermore, Groves et al. (2013) document the innovations identified because of Big Data Implementations. Another organization, Intel, announced its Collaborative Cancel Cloud in August of 2015 to enable diagnosing of cancer patients based on their specific genome sequencing and tailor a precision treatment plan for them all based on the concept of Big Data.

Education

Erwin (2015) argues for students to be more literate in their abilities to use data. He argues that there is a growing call for students to develop data literacy (Gunter 2007; Vahey et. Al., 2012). His theory is that a more project based learning where students solve real world problems with data that is provided to them will enable them to build skills and be able to meet the current demands of business. Similarly, Rijmenam (2014) reasoned changes in the education systems by using Big Data to change the way that students and teachers interact. A more practical example, Gwinnett, in suburban Atlanta, Georgia, is the 14th largest school system in the United States, has 23,000 employees and transports more people every school day than locally based carrier, Delta Air Lines. All that activity generates information, more and more of it captured digitally and in 2002, as the school system's leaders continued seeking fresh educational solutions, they began to explore how analytics could help how all that information could be investigated for patterns, relationships, dependencies, and predictors.

Public Sector (Government)

Gamage (2016), in his article examines the opportunities presented by effectively harnessing big data in the public-sector context. He talks about the impact of Big Data and how it will play an important role in the future. Furthermore, he also outlines key challenges to be addressed to adopt and realize the benefits of Big Data in the public sector. Similarly, another article, stemming from SAP's partnership within the Middle East governments, documents high level of Big Data production, consumption and the need to train public sector to be successful at these opportunities (Arabia 2000, 2015).

Miscellaneous

The Big Data Strategy framework in Servitization as proposed by Opresnik and Taisch (2014) is focusing on new revenue streams and decreasing product-service costs in manufacturing. An optimization model for green supply chain management based on Big Data proposed by Zhao et. al (2015) is a scheme that minimizes the inherent risk of hazardous materials, associated carbon emissions and economic cost. Finally, Nike, as Galbraith (2014) notes, has used Big Data to create a completely new business unit and summarizes that impact Big Data has on the organization design by using the Star Model ™ Framework with Big Data. The CEBR (2012) has anticipated that the benefits of big data innovation opportunities would contribute £24 billion to the UK economy between 2012 and 2017. These opportunities are described to be identifying hidden patterns, better decision making, improving business processes and developing new business models (Halaweh & Massry 2015). There are many more examples of such initiatives and values across industries that organizations are realizing and to go back to Gary Kings' quote, at the start of the paper, the accumulation of data is reaching out to every industry and organization across geographies.

In summary, we clearly see the various opportunities being explored, examined and extracted for the betterment and effectiveness of organizations across the different industries that have successful Big Data Implementations. This is the very objective behind the reason this research. The research favors utilizing OD characteristics and engaging OD in a leadership role to lead successful Big Data Implementations. The research employs interviews, observations and empirical data to test this opinion. Alternatively, this will also afford us the opportunity to look at relationships that can impact successful Big Data Implementations.

Statement of Research Purpose and Question

Big Data is here but how much do we know about it in concept? It is also very clear from literature and practitioners of the field that Big Data is here to play a role in our future (Gamage 2014; Burg 2014; Allouche 2014; Halaweh and Massry 2015; Wamba et. al 2015; Wixom 2014; Xu et al 2015; Chen et al 2012; Forrester 2012). For that purpose alone, it is imperative we learn, take advantage and realize its potential to transform entire business processes (Wamba et al., 2015). The main purpose of this research is to build on the current diverse literature around Big Data by contributing discussion and data that allow common agreement on definition, characteristics and factors that influence successful Big Data Implementations. The research question being investigated is based on the argument establishing Big Data be used as a tool for the organization by which to develop and create efficiencies enterprise wide. Furthermore, the researcher hopes to prove that Big Data, especially during the current infancy stage (Barker et. al 2016; Halaweh and Massry 2015), should be primarily led by organizational theorists, specifically the field of Organizational Development (O.D), whose mission it is to develop and create effective organizations. The researcher explores the following question with this research, “What impacts successful Big Data Implementations?” As part of the research question the following hypotheses will be examined:

H1: Given Big Data characteristics, Organizational Development characteristics positively impact successful Big Data Implementations.

H2: Given Big Data characteristics, technical competence positively impacts successful Big Data Implementations.

H3: Given Big Data characteristics, addressing privacy concerns positively impacts successful Big Data Implementations.

H4: Given Big Data characteristics, cross functional collaboration positively impacts successful Big Data implementations.

H5: Given Big Data characteristics, understanding of Behavioral knowledge positively impacts successful Big Data Implementations

H6: Given Big Data characteristics, organizational change management expertise positively impact successful Big Data Implementations.

H7: Given Big Data characteristics, established Data ownership roles positively impacts successful Big Data Implementations.

H8: Given Big Data characteristics, type of industry does not significantly impact successful Big Data Implementation.

H9: Given Big Data characteristics, age does not significantly impact successful Big Data Implementations.

H10: Given Big Data characteristics, profession does not significantly impact successful Big Data Implementation.

H11: Given Big Data characteristics, geographic region does not significantly impact successful Big Data Implementations.

H12: Given Big Data characteristics, technical infrastructure positively impacts successful Big Data Implementations.

H13: Given Big Data characteristics, technology intensive organizations positively impact successful Big Data Implementations.

H14: Given Big Data characteristics, organizations with an information governance policy positively impacts Big Data Implementations.

H15: Big Data Characteristics (volume, variety, velocity, veracity, variability, value and computing resources) positively impact successful Big Data Implementations.

H16: Leadership and employee engagement positively impact successful Big Data Implementations.

H17: Given Big Data characteristics, organizations with a data strategy positively impacts successful Big Data Implementations.

H18: Given Big Data characteristics, business area involvement positively impacts successful Big Data Implementations.

H19: Given Big Data characteristics, funding positively impacts successful Big Data Implementations.

H20: Given Big Data characteristics, experience in working with many data driven projects positively impacts successful Big Data Implementations.

To accomplish this, the research will follow the below five step process:

1. Conduct a comprehensive review of literature.
    a. The research reviews Wamba et al., 2015 approach and Halaweh and Massry (2015) model as it relates to Big Data Implementations and the tie in with Organizational Development.
2. The research identifies the current gaps, definitions and existing variables from literature regarding Big Data Implementations and Organizational Development.
3. The research employs a two-part mix methods study based on grounded theory. The two-parts are: (a) Qualitative: interviewing industry experts on successful Big Data Implementations, performing qualitative analyses to identify factors and inducing a model to measure via an online survey and (b) Quantitative: performing quantitative analyses to test the model, answer our research question and proving/disproving our hypotheses.
    a. The interviews took place between April and October 2016. Following the interviews, survey distribution took place between January and March

2017 to the following groups on LinkedIn (of whom the researcher is a member) and the researcher's personal/professional contacts.

1. Big Data in Insurance – 444 members
2. IoT, Big Data in Canada – 65 members
3. Big Data in the Pharmaceutical Industry – 710
4. Big Data for Music Industry – 73 members
5. Big Data Challenges – 383 members
6. Big Data Analytics for Financial Institutions – 312 members
7. Predictive Analytics, Big Data, Business Analytics, Data Science in Oil and Gas and Energy – 904 members
8. Big Data and Analytics Group – 280,707 members
9. Business Intelligence Professionals (BI, Big Data, Analytics, IOT) – 192,761 members
10. Big Data, Analytics, Business Intelligence & Visualization Experts Community – 192,346 members
11. Personal/Professional contacts – 624 members

4. Next, the research compiles a list of factors that impact successful Big Data Implementations from our quantitative tests.
5. Finally, the research discusses the findings.

Summary

Using a two part mix methods study (qualitative first followed by quantitative analysis) and grounded theory the research found the following factors as significant in determining success of Big Data Implementations: opinions on Organizational Development characteristics (change, collaboration), competencies (technical skills, computing resources and support) and change management (enabling change and initiatives for betterment of organization) are associated with higher perceptions as part of measuring successful Big Data Implementations via new products, services, competitive advantage, patents or furthering of business strategy. Furthermore, from the opinions gathered, the perception that Big Data is just a technical challenge is not true but rather a challenge for the organization.

A few points regarding implications for the study to keep in mind are: (a) there were only a small number of experts who were interviewed and they were all found via professional and personal contacts of the researcher, (b) while the survey was distributed across world the qualification and response rate for the survey was very low, (c) the interview coding, categorizing and inspection of themes was conducted solely by the researcher and subject to interpretation, (d) the survey questions were formed by the researcher based mostly from experience and expert opinions, (e) the survey questions had multiple questions measuring similar characteristics and that may have distributed the impact of some of the factors and (f) anonymity was a very important factor to the interviewee's. Many didn't want to be interviewed nor did they want to be recorded. The researcher provided as much leeway as possible in answering questions, opting out of study and minimizing use of competitive knowledge.

The following chapters provide the details on this research, discussion and findings. Chapter 2 provides a comprehensive review of the literature including search criteria, definitions and gap the research is investigating. The methodologies used to study the research are provided in Chapter 3 with the results and analysis of that data presented in Chapter 4. Chapter 5 discusses those findings as well as implications to theory and practice and Chapter 6 presents a summary of the study as well as areas for future research.

Chapter 2 Literature Review

For a comprehensive literature review, this research utilized a custom structure borrowed from Wamba et. al.'s, (2015) approach (which included approaches by Niagi and Wat 2002 in electronic commerce, Niagi et. al., 2009 in CRM and data mining, and Niagi et al., 2008; Wamba et. al, 2013 and Lim et. al, 2013 regarding RFID related topics), Halaweh and Massry's (2015) on Big Data and the researcher's own education and experience in the field of Big Data.

Wamba et. al., (2015) approach entails three characteristics: (i) the development of a classification framework, (ii) conduct the literature review and (iii) realize classification of relevant journal articles. Halaweh and Massry's (2015) approach relied on concentrating on specific opportunities and challenges for Big Data to help organizations realize value.

The reason for using both as part of the custom structure for review was three-fold: (a) Wamba et al.'s (2015) use of the 5 Big Data enabled dimensions based on 5 key dimensions from a McKinsey Classification Framework, (b) utilization of an existing conceptual model proposed by Halaweh and Massry (2015) and (c) focus on organizational level attributes as the researcher is in a doctoral program in Organizational Development (O.D) and believes that O.D can enable organizations to maximize success regarding Big Data implementations, especially in its infancy.

The 5 dimensions from Wamba et. al., (2015) are: (i) data policies, (ii) technology and techniques, (iii) organizational change and management, (iv) access to data and (v) industry structure. These techniques allow us to squarely focus on utilizing the field of Organizational Development as we explore the relationship with successful Big Data Implementations.

The conceptual model by Halaweh and Massry (2015) focuses on: (i) top management support, (ii) organizational change, (iii) data availability and quality, (iv) infrastructure, (v) required skill set, and (vi) challenges such as security and privacy.

Literature Search

In existing literature, Big Data appears ~ 44,615 times across 37 main databases (list is shown in Appendix A). At this juncture, the research utilized only peer reviewed/scholarly/academic journals that were to be most commonly used by academics and practitioners alike for acquiring information and disseminating new findings and represent the highest level of research (Wamba et. al., 2015 quoted from Niagi and Wat 2002). That bought the search down to 6,480 articles without any year limitations.

Since the focus was on non-technical articles but still considering the IS (information systems) side of things, the researcher wanted to further confine this result set with a term that incorporated many of Big Data findings and relevance to the organizations today. The research utilized various searches using terms such as management, organizations, marketing, analytics and information technology to name a few. The one term that gave the best results and covered a large breadth of the Big Data landscape was, business. Thus, limiting the search with the business subject term dropped the number down to 493 main articles spanning between the years of 2010 to 2016. Any further chopping or restriction removed certain articles and as such I decided to use the 493 articles as it covered a large surface area regarding Big Data.

Breaking this by year showed that the focus has just recently started to catch on and there is more research being added. See figure 1 below for a breakdown by year.

Figure 1. Number of Articles by year

Year	**Total for year (“Big Data” & “business)**	**Total Articles (cumulative)**
2010	1	1
2011	1	2
2012	14	16
2013	104	120
2014	118	238
2015	159	397
2016	96	493

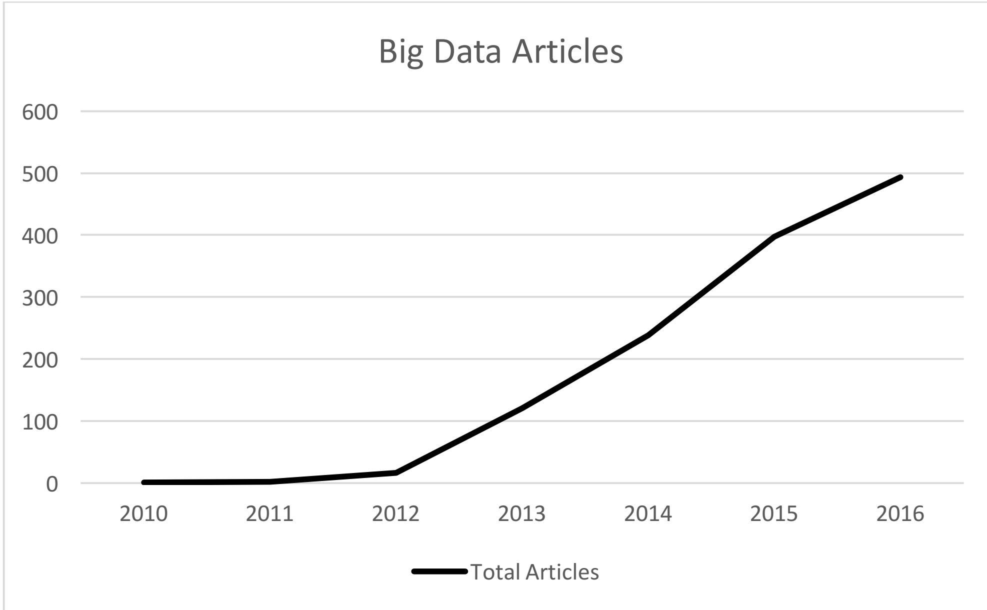


Figure 2: Chart – Big Data Articles

Going through the literature, there were various articles on implementations, best practices, case studies, management/organization theories, complementing technologies, big data challenges, big data analytics and other multiple variations with each either providing proof (by theory) or via identifying challenges regarding Big Data. What was clear was that there was consensus that Big Data was deemed as the future, the real deal and central in creating big impacts (Halaweh and Massry 2015; Wamba et. al 2015; Wixom 2014; Xu et al 2015; Chen et al 2012; Forrester 2012; Church and Dutta 2013, McAfee and Brynjolfossn 2013; Manyika et al. 2013). Furthermore, there was no consensus on the definition of the term, Big Data (Hartmann et al. 2014; Young 2014; George et al. 2014; Church and Dutta 2013; Manyika 2013; McAfee and Brynjolfossn 2013, Wamba et. al 2015; Halaweh and Massry 2015), evidence of what influences successful Big Data implementations and if Organizational Development can lead this effort. As such, in this chapter we review definitions and examine literature on factors that can impact successful Big Data Implementations. To begin with, let's start with definitions for both Organizational Development and Big Data.

Definition: Organizational Development

Many pioneers and experts in the field have penned definitions for Organizational Development. Each definition has embodied traits that characterize this field and hope to be all inclusive of the current real world situation.

An OD pioneer, Richard Beckhard, in his 1969 article, defines OD as follows: "Organization development is an effort (1) planned, (2) organization-wide, and (3) managed from the top, to (4) increase organization effectiveness and health through (5) planned interventions in the organization's "processes," using behavioral-science knowledge" (Gallos, 2006).

A second definition authored by Thomas Cummings and Christopher Worley in the Addison-Wesley 2009 10th edition book states: "Organization development is a system wide application of behavioral science knowledge to the planned development, improvement, and reinforcements of the strategies, structures and processes that lead to organizational effectiveness" (Cummings and Worley, 2014).

A third definition, also found in Thomas Cummings and Christopher Worley's 10th edition, was formulated by Wendell French, in his 1969 article: "Organization development refers to a long-range effort to improve an organization's problem-solving capabilities and its ability to cope with changes in its external environment with the help of external or internal behavioral-scientist consultants or change agents, as they are sometimes called" (Cummings and Worley, 2014).

Finally, a fourth definition, also found in Thomas Cummings and Christopher Worley's 10th edition, was taken from Michael Beer's 1980 book: "Organization development is a system-wide process of data collection, diagnoses, action planning, intervention and evaluation aimed at (1) enhancing congruence among organizational structures, process, strategy, people and culture, (2) developing new and creative organizational solutions; and (3) developing the organization's self-renewing capacity. It occurs through the collaboration of organizational members working with a change agent using behavioral science theory, research and technology" (Cummings and Worley, 2014).

As you read those definitions, you will see that there are similar characteristics that embody each definition of OD and how each characteristic ties in with Big Data, namely:

1. Organizational effectiveness: We understand from the definitions that the goal of OD is marching towards improving organizational

effectiveness. There is something the organization must do differently or change to ensure it is being more effective and/or competitive. Big Data will only allow this goal to be met more feverishly, simply because the sole act of measuring the effectiveness focuses on data. It can exist in the form of unstructured (non-categorical text), structured (relational) or a combination of both but nevertheless is data itself. It can be measured via financial, economics, humanistic values or in simplistic forms, such as time saved with a new process that shows positive gain overall for increasing organizational effectiveness.

2. Planned change: we can infer that for organizations to embark on their goals and do things differently, change will occur. This change needs to be planned. The planned change is not just based on accepted best practices of change but rather answering how change should be approached. For instance, action research, which is one of the planned change models by Kurt Lewin (Appendix B), is based on compiled data from research conducted to present a diagnosis and plan of action. As such, data is observed, gathered, extracted and mined for decision making to plan change.

3. Use of Behavioral Science/Knowledge: One of organizations' main assets are people and any planned changed will likely impact those people who work in teams or groups. As such, the use of behavioral science that deals with group dynamics, individuals, culture, strategy,

leadership, etc allows focus on tools, actions and methodologies to be used by OD experts to ensure that change is carried out successfully across the organization. This is important as the environment we live and work in is impacted by what we can collect (structured vs non-structured data), how much we can access (available and ready for use), how to harness and use to improve.

4. Long term: there is no start and end date like a project. These interventions and changes carried out are for the long term and are cycled (repeat). We don't stop once we complete one aspect of change; it is continuous cycle to ensure that effectiveness is a never-ending process for the organization that deals with changes in its external and internal climate. It also becomes part of organizational strategy as you revisit it with changing conditions, new data and growth.
5. Organization-wide and teams: the effects of this are very rarely limited to one individual. The effects and changes usually start with one team but in most cases, include multiple teams, the whole organization and in some cases multiple business units to achieve a common goal.

In summary, we can see from the above characteristics that define OD that data has been utilized throughout the core characteristics to help with defining OD and is very much a main stay in guiding organizations to improve its effectiveness and value. Armed with the above, we can examine Big Data's impact by the characteristics that define OD (Planned Change, Use of behavioral science/knowledge, organizational effectiveness and impact and collaboration of teams).

Definitions: Big Data

Compared to Organizational Development characteristics, there is discrepancy in coming up with defining what Big Data stands for (Hartmann et al. 2014). Young (2014), Church and Dutta (2013), and George, Haas and Pentland (2014) have put together the various definitions of Big Data that are present across literature and the confusion that exists in defining Big Data. Below is a sample:

1. "Big Data" is a science of fielding algorithms that enable machines to recognize complex patterns in data. It fuses machine learning with a very deep understanding of computer science and algorithms and that, of course, is key to being able to take machine learning and deploy it in a very scalable way (Paredes, 2012).
2. "Big Data" exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it (Dumbill, 2013).
3. "Big Data" is the ability to mine and integrate data, extracting new knowledge from it to inform and change the way providers, even patients, think about healthcare (Roney, 2012).
4. "Big Data" is not a precise term; rather, it's a characterization of the never-

ending accumulation of all kinds of data, most of it unstructured. It describes data sets that are growing exponentially and that are too large, too raw, or too unstructured for analysis using relational database techniques. Whether terabytes or petabytes, the precise amount is less the issue than where the data ends up and how it is used (EMC2, 2012).

5. "Big Data" is the ability to collect, process, and interpret massive amounts of information. One of the biggest potential areas of application for society is healthcare (Rooney, 2012).
6. "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (Gartner IT Glossary, n.d)
7. "Big Data" are datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze (Manyika et al., 2011).
8. "Big data is a term that describes 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's Federal Big Data Commission, 2012)
9. "Big Data" is techniques and technologies that make handling data at extreme scale affordable (Hopkins & Evelson, 2012).
10. "Big Data" is more data than our current systems and resources can handle (Fogarty 2012).
11. "Big data" is an explosion of available information, a byproduct of the digital

revolution (I. Thomas, 2013).

12. “Big data” does not exist in healthcare settings (Bollier & Firestone, 2010).
13. “Big Data” n: the belief that any sufficiently large pile of s--- contains a pony (Arbesman, 2013).
14. Big Data "high volume, velocity, and/or variety information assets that demand new, innovative forms of processing for enhanced decision making, business insights or process optimization." (Gartner 2012)
15. At a basic level, Big Data is a concept, approach, or way of thinking about massive amounts of information and the outcomes that can be achieved by integrating that data. (Church and Dutta 2013)
16. Big Data has been defined by one group of technology consultants as “information that can’t be processed or analyzed using traditional processes or tools” (Zikopoulos, et al., 2012).
17. ‘Big data’ is a general term referring to the massive amounts of data collected from many sources, including the web and the cloud (Manzoor, 2015).
18. “Large pools of data that can be captured, communicated, aggregated, stored and analyzed” (Manyika et al. 2011)
19. “big data is the new generation of technologies and architectures which are designed to economically extract value from very large volumes of a wide variety of data, by enabling high velocity capture, discovery and/or analysis” Woo et al. (2011)

Given the above variance in defining the term, “Big Data” and no agreement across academia and professionals on what the term means, the paper further explores the various characteristics of Big Data available today.

Gandomi & Haider (2015), IDC, IBM, Gartner, and many others have contributed with an excellent summary regarding Big Data characteristics. Clearly, size is the first characteristic that comes to mind considering the question “what is big data?” (Gandomi & Haider 2014). Following that, the Three V’s have emerged as a common framework to describe big data (Chen, Chiang, & Storey, 2012; Kwon, Lee, & Shin, 2014): Volume, Variety and Velocity. There have been more additions: IBM, White (2012) introduced Veracity – the fourth V, SAS introduced Variability and Complexity, the fifth V and Oracle introduced Value, the sixth V. While these are existing today there are possibilities with further enhancements more maybe added, or defined further contextually. For instance, definitions of big data volumes are relative and vary by factors, such as time and the type of data. What may be deemed big data today may not meet the threshold in the future because storage capacities will increase, allowing even bigger data sets to be captured. There is even the possibility of having “smarts” added to this volume of data as well. There are questions around the usefulness and life of the data as well. Furthermore, that new insights gleaned from such data-value extraction can meaningfully complement official statistics, surveys, and archival data sources that remain largely static, adding depth and insight from collective experiences—and doing so in real time, thereby narrowing both information and time gaps.

In hopes to define Big Data, we look at the definitions for each of the 6 Vs below as they seem to characterize Big Data broadly:

1. Volume – the quantity of data that is generated is very important in this context. It is the size of the data, which determines the value and potential of the data under consideration and whether it can be considered as Big Data or not. The name 'Big Data' itself contains a term which is related to size and hence the characteristic. A survey conducted by IBM in mid-2012 revealed that just over half of the 1144 respondents considered datasets over one terabyte to be big data (Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012). Beaver, Kumar, Li, Sobel, and Vajgel (2010) report that Facebook processes up to one million photographs per second. One petabyte equals 1024 terabytes. Earlier estimates suggest that Facebook stored 260 billion photos using storage space of over 20 petabytes.
2. Variety - the next aspect of Big Data is its variety. This represents the different types of data available, such as text, numbers, images, videos, documents, spreadsheets, etc. This signifies the category or type of data something belongs to. This can be very essential fact that needs to be known. This helps the people, who are closely analyzing the data and are associated with it, to effectively use the data to their advantage and thus upholding the importance of the Big Data.

3. Velocity - the term ‘velocity’ in this context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges, which lie ahead in the path of growth and development.
4. Veracity - which represents the unreliability inherent in some sources of data. For example, customer sentiments in social media are uncertain in nature, since they entail human judgment. Yet they contain valuable information. Thus, the need to deal with imprecise and uncertain data is another facet of big data, which is addressed using tools and analytics developed for management and mining of uncertain data.
5. Variability- this is a factor, which can be a problem for those who are analyzing the data. This refers to the inconsistency, which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively. Data management can become a very complex process, especially when large volumes of data come from multiple sources, especially when they are not linked or consistent in nature. For instance, data being gathered from multiple sensors or servers that is not consistent in reporting and providing the same data. Therefore, the data needs to be linked, connected and correlated to be able to grasp the information that is supposed to be conveyed. This situation, is therefore, termed as the ‘complexity’ within Big Data.
6. Value - big data is often characterized by relatively “low value density”. That is, the data received in the original form usually has a low value relative to its

volume. However, a high value can be obtained by analyzing large volumes of such data.

Gandomi and Haider (2015) highlight that it is important to remember that universal benchmarks do not exist for volume, variety, and velocity that define big data. The definition limits depend upon the size, sector, and location of the firm and these limits evolve over time. Also important is the fact that these dimensions are not independent of each other. As one dimension changes, the likelihood increases that another dimension will also change. What this means is that there is a point (using the Three Vs, which Gandomi & Haider 2015 call the "Three-V tipping point") where traditional data management and analysis technologies become inadequate for deriving timely intelligence. The Three-V tipping point is the threshold beyond which firms start dealing with big data. This also holds true for the other added Vs. The opportunity cost, ambiguity and collection ability pay a role in authenticity/reliability of the data, the inconsistencies behind gathering and gaining the data and the value derived and implementation costs from the data.

In summary, having gone through the definitions that exist in literature today and having looked at characteristics to date we are still not closer to agreeing on the definition of the term, Big Data. The next step, therefore, was to go out in the field, speak with the experts to understand what was being done, perform analyses and create a survey that will allow us to gain agreement on the characteristics that form Big Data. This will help us with providing a baseline for what Big Data is. Moreover, it will afford us the understanding of the term and allow agreement upon the characteristics (6 Vs). As an OD scholar-practitioner the interest is in all the moving parts that contribute to the effectiveness and profitability of the organization. Data, is focal to OD tools and methods to derive those valuable recommendations. To take this next step, it was first important to search the literature for pre-existing factors that characterize successful Big Data Implementations, examine literature for arguments in favor of OD and review gaps in current literature.

Evidence of Big Data and OD in Literature

Miller (2014) argues that data is quickly becoming a strategic business asset. He further argues that data will impact the full business spectrum and is not just about IT and technology. "Job spanning the entire business spectrum, including legal, sales, marketing, finance, product development, manufacturing, and operations, will be impacted by the big data phenomenon" (Miller 2014).

Other scholars such as Gobble (2013) and Manyika et. al., (2011) identified big data as the next big thing in innovation and the next frontier for innovation, competition and productivity, respectively. Strawn (2012) called it the fourth paradigm of science. Furthermore, McAfee and Brynjolfsson appropriately categorized their article on Big Data as a management revolution similar to what Ann Keller et al., (2012) termed Big Data as bringing a revolution in science and technology. Miller 2010; Parry 2010; Vaidhyanathan, 2010, all argue that every profession, whether business or technical, will be impacted by Big Data. Galbraith (2014) also maintains the legitimacy of Big Data by citing reports from the World Economic Forum, the McKinsey Global Institute, and *The Economist* Intelligence Unit Furthermore, Gartner Inc. through its research mentions "Through 2015, 85% of Fortune 500 organizations will be unable to exploit big data for competitive advantage."

Big data, McAfee and Brynjolfsson (2012) write, is far more powerful than the analytics of the past. Executives can measure and therefore manage more precisely than ever before. They can enable better predictions and smarter decisions. They can allow for more targeted and more-effective interventions in areas that so far have been dominated by gut and intuition rather than by data and rigor. Wamba et al., (2015), similarly, quoting from other scholars, characterized such rationale behind Big Data based on the reason that Big Data is "capable of changing competition by "transforming processes, altering corporate ecosystems, and facilitating innovation" (Brown et al., 2011); unlocking organization business value by unleashing new organizational capabilities and value (Davenport et al., 2012); and facilitating firms to tackle key of their business challenges (Gehrke, 2012).

A study by Perrey et al., (2013) whose research in academia and industry, showed that retailers can achieve up to 15–20% increase in ROI by putting big data into analytics. Moreover, Manyika et al (2011) found that collecting, storing, and mining big data for insights can create significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector and creating a substantial economic surplus for consumers. Finally, McAfee and Brynjolfsson (2012) depicted Big Data's capability of transforming the decision-making process by allowing enhanced visibility of firm operations and improved performance measurement mechanisms. They all rely on the fact that more data now crosses the Internet every second than was stored in its entirety 20 years ago (McAfee and Brynjolfsson 2012; Halaweh and Massery 2013). As such, near real-time information makes it possible for a company to be much more dynamic and responsive than its competitors.

Obsession with and creation of data is nothing new – in fact, data has existed over the last few decades. Organizations have had the capability to work with, record and use them to develop insights and make decisions. They have had the ability to work with large datasets as well – think the Census Bureau. The question beckons - why the sudden hype and focus on Big Data? Well, for a few reasons. First, and as Galbraith (2014) notes, this data is unstructured and is different from the usual structured datasets that we have shared across and within organizations. It is different than the columns and rows we are used to visualizing and sharing. Second, and as Galbraith (2014) notes, this data is available in real time. Third, the first two aspects of Big Data open business opportunities that did not exist before. Lastly, while current technology is expanding there is not a single defined architecture that can manage and provide insights with the variety and the volume of data gathered. Essentially, we are sitting on a gold mine of information, which we are struggling to mine and gain insights from.

As Wamba et al., 2015 note, despite the excitement and recent interest in 'big data', little is known about what encompasses the concept. Current adopters are trying to better understand the concept and therefore capture the business value from 'big data'. Also, it's worth nothing that there are very few empirical studies that have been conducted to assess the real potential of big data (Wamba et al., 2015). As more data crosses the Internet every second nearly real-time information makes it possible for a company to be much more dynamic and responsive than its competitors. The famous adage in management literature, "you can't manage what you don't measure", attributed to both Edward Demming and Peter Drucker rings in the reason why the recent explosion of this data is important.

Data, as it exists today, comes in three forms: (i) structured – research, transactional, product data and relationship oriented, (ii) non-structured – social media posts (Facebook, Twitter, etc), texts, voice recordings and (iii) machine/web/software related – mapping, call records, application logging, web browser and personalization, etc. It can be used to provide value and betterment to organizations or individuals one and all. This multidimensional existence of data is allowing business to be re-imagined as businesses look to realize value. We refer to this data as the industry defined concept of Big Data. The fact of the matter is that organizations have been slow to get on the bandwagon and start utilizing Big Data. There are some good reasons for the slow start, namely: (a) this is going to change how organizations conduct current business and operations – their current value proposition may come at stake, (b) the resources and expertise are not readily available to embrace this change and (c) there is no framework, governance or method to go about implementing such a program.

The literature does not provide any empirical studies or publications that have tied Organizational Development with Big Data or somehow illustrated their correlation. OD as a field lacks a model to accompany and complement its process strengths in the usability and advantages from Big Data. Additionally, there is very little empirical evidence on agreement of the definition or on characteristics and elements. There is, however, one conceptual model that looks at Big Data Implementations. Halaweh and Massry (2015) presented a conceptual model for research.

Figure 3: Halaweh and Massry (2015) conceptual model

The model above focuses on the 3 Vs for and utilizes challenges, failure and success criteria and obstacles needed for successful implementation of Big Data that exist in literature. The model is completely based on the review of literature, is generic and conceptual in nature. The two future directions listed by both authors (a) use quantitative research methods to test model and verify validity of the assumptions and (b) evaluate model by applying qualitative methods to interview experts who work in different sectors to develop or extend the model that affect Big Data implementation. This research study uses both methods (a) and (b) to develop and test statistical relationships of the model.

Examining Literature: Argument in favor of O.D

This is where I believe OD has a role to play. OD, foundationally, is concerned with organizational effectiveness/value. We are trying to get the organization from one instance to another through research, data collection, diagnosis and action due to changes as they occur in the world. Our goal is always with improving effectiveness or value of the organization. The arguments in our favor include:

(a) Our methods, tools and models require us collect data both structured and non-structured from multiple sources and vantage points. This is how Big Data is formed from multiple sources. Since the beginning, dating back to Fredrick Taylor's work with organizations, the effort revolved around collecting quantitative (structured) and qualitative (non-structured) data. Weisbord (2012) talks about how Frederick Taylor in the early 1900s did consulting engagements with organizations helping to provide efficiency by studying not only the solid quantitative evidence for improvement but also the human interaction. Church and Dutta (2013) capture this piece perfectly noting the singular role that data (and feedback) play in creating energy for change. Similarly, David Nadler (1976) indicates that the process of collecting, analyzing, and using data for feedback is important as a determinant of the nature and extent of feedback effects for organizational change. This use of data, (whether quantitative or qualitative) as a catalyst for action is one of the most unique contributions that OD makes to organizational transformation. Furthermore, the evidences to partake in a change was derived in early OD was as much based on structured data, such as error rates, cost, profit as unstructured data that referenced human interaction, behavioral knowledge and organizational culture (qualitative). Similarly, Pentland (2014) mentions "the use of large-scale data to predict human behavior is gaining currency in business and government policy practice, as well as in scientific domains where the physical and social sciences converge".

(b) The second argument in OD's favor deals with the insights, analytics and meaning that is derived from such data that exists and/or can change in near real-time when working on a change effort from the current state to a future ideal state. The current issue where tools and expertise are not available to utilize meaningful data insights presents OD scholar-practitioners with path to lead laying down the foundation for organizations to embrace and to build on from.

(c) Finally, the last argument in favor, as Church and Dutta (2013) also mention, deals with the action that is tied once data and realization is made. The action is derived from what we gathered, analyzed and given as feedback.

Organizations are now increasingly being known for their digital property, footprint and presence. This has caused a shift in how organizations tackle the value proposition provided by them. The organizations don't need to be left behind when new technology or data, in our case, is available to be "opportunized". It was a similar case when e-commerce came calling and left many brick and mortars behind who did not feel obliged to take notice. As Styrk (2015) notes, In the past 15 years, many organizations have seen an increase in the amount of data, named Big Data, in the order of terabytes (10^{12}) which has been acquired and stored, but not prepared, cleaned, managed, and analyzed enough to provide competitive value for the organization (LaValle et al., 2011). The organizations that analyze Big Data have profit performance that is about twice the amount of the organizations, which do not analyze Big Data (LaValle et al., 2011). The decisions include providing enhanced customer service and making improved or new products using updated business processes (Beath et al., 2012). Malik (2013) compared Big Data to oil. The costly extraction of oil in a crude form from the earth must be managed and then it should be prepared to provide a cleaned substance, which can be used in a variety of ways to provide value (Malik, 2013). Big Data may come from many different sources, and the process to prepare and clean Big Data to extract value may also be costly (Malik, 2013).

Given the above arguments, there is change brewing and OD should take the leadership position and come forward and align business, strategy and Big Data. Twenty-six years ago, Jelinek and Litterer (1988) wrote an article titled, "Why OD Must Become More Strategic" made a case for OD being needed by businesses ever more so than before. The caveat was that OD had to align with business goals and take the leadership position in the business world by (a) leading the way of enabling the acceptance and amalgamation of change, (b) acknowledging more integration across business units and (c) using their skills to help leaders across the organization implement and shift strategy/mission as needed. The basis for such an argument from Jelinek and Litterer stems from facts such as: (a) competitive pressures, (b) an ever changing technological and (c) economic landscape and organizational demands. It was clear then and should be now that OD is needed by business to traverse the various changes on the horizon and be strategic about it. One such new area we need to align and take a leadership position is with focus on "Big Data".

Head (2009) argues that true OD is strategic. It is not that the focus or definition of strategy/OD has changed but the comprehensiveness of what strategy entails. Furthermore, Yaeger and Sorensen (2009) point to Cummings and Worley's 2009 text that highlights a broad encompassing of activities that show the strategic nature of OD. In a more practical use case, if our aim/strategic goal was to acquire market share to become more profitable, we may look to form alliances, acquire companies, or look at the various global opportunities present. These options would be considered OD interventions (Cummings & Worley 2014) but are strategic in nature. All the above scenarios will impact, by way of change, the current organization, its processes, structure, people and strategy.

It is important to define what the term means within realm of OD. My understanding of strategy is of an assessment of the organizations' mission and how to utilize its scarce resources to effectively achieve that mission. This may contain business strategy; change strategy, team strategy and the like to effectively allow the organization, team or department to best utilize its resources. Richard Beckhard's article (1969), talks about how OD efforts are required in various conditions and one of them is the adaptation of the organization to a new environment. Again, this is now commonly referred to as the organizations' strategy. This new environment could be culture change due to a merger or acquisition, or an alliance, or even use of a new technological innovation (hint: Big Data) that affects the organizations' effectiveness and induces change. In all cases, the importance of giving OD a strategic view extends its boundaries to look at various other factors that can influence and cause disruptions in achieving organizational effectiveness. The way organizations allow their teams to adapt, communicate and function gives way to how a challenge is approached. With various team focuses, such as marketing, finance, operations, marketing, accounting and the like team building, interactions and communications at various levels becomes strategic especially with the advent of vendors, contractual employees and virtual members that while contributing may not be strategically aligned to the organizations' mission.

Weisbord (2012) calls Frederick Taylor as the "father of scientific management" who worked as a traveling consultant going to organizations and working to create efficiencies based on research and data. Cummings and Worley (2014) give the title of "Father of OD" to Kurt Lewin where his initial work in the1940s with T-Groups led to the coining of the term itself. They talk about how his focus on more of the unstructured data (behavioral, environmental, etc.) allowed him to do what he did in the 40s-50s. Even the more recent scholars of OD: McGregor, Weisbord, Likert, Schein and others used data, some unstructured through observations, to assess the current state, diagnose and put an action plan together for change.

The reality of the 21st century, according to Freidman (2006) and Kerber and Buono (2010), that companies are faced with constant change. It is ongoing, everlasting and just part of everyday life. The "pace" of this change, along with its flavors, is what is hard to keep up with. Burke (2014), Cummings and Worley (2014), Buono and Jamieson (2010), Sorenson and Yaeger (2009) and Galbraith et. al (2002) all discuss the various areas of change within OD. Buono (2010) discusses the challenges and taking account of organizations with respect to their change capacity and the strategies that go with it. The common theme among the scholars is how changes (advances in technology, global presence, leadership methods and various cultural dimensions) are more broad and complex. Furthermore, Cummings and Greiner (2009) further document the ease of eliminating competitive advantage by using the technology to enter a space and start competing in today's hyper connected world.

This pace of change keeps increasing and managing this change, therefore, requires a certain element of flexibility. Burke (2002) talks about change as transitions and goes on to mention that such change is an "ongoing, evolving process". Buono and Kerber (2010) categorize change through micro, meso and macro levels within the organization that contribute to the effective management and profitability of the organization. The other parts of the organization such as strategy making and the environment are being challenged to be nimble and dynamic. All these combined create challenges for the organization - the one common ground for such items to coexist. As such, an organization, with its limited resources, needs to be ever watchful of how, when, where and with the biggest impact it can create value. Below you will find a quick summary for each of the different early planned change models and how they tie in with the Big Data:

1. Lewin's Planned Change Model: This is one of the earliest of the change models. Lewin's basic idea of change was anything or any force that keeps the system from being stable. He proposed the following three

steps: (a) unfreezing: breaking down the current behavior/level or a group, department or team, (b) moving: shifting the behavior/level to the required place that will allow it to get to the new state and (c) refreezing: reinforcing the current behavior/level such that it becomes the norm or expected from the group, department or team.

Clearly, this change model does not use structured data for research, diagnosis and action implementation. There is a method of approaching the change that causes the instability but the data to move into a more stable form requires analysis, study and observation beyond simple structured data.

2. Action Research Model: this model treats the process of planned change as an ongoing cycle. There are 8 main steps: (a) problem identification, (b) consultation with a behavioral science expert, (c) data gathering and preliminary diagnosis, (d) feedback to key client or group, (e) joint diagnosis of the problem with members of the team, group or department where problem was identified, (f) joint action planning, (g) action and (h) data gather after action. After (h) you will go back to (d) and repeat the process till problem is rectified or a new diagnosis is identified that needs attention.

   In some respects, this is similar to #1 except that this involves a far more steps and treats the gather of data and its insights separately. The freedom to record the amount (volume), type (variety) and speed (velocity) of data and communication is evident in this model.

3. Positive Change Model: this model is different from the first two models as it focuses on what the organization is doing best. It takes what the organization or teams do effectively and builds on that to infuse change. This is applied through the process of using appreciative inquiry (AI). Its goal is to promote positive value and use that to analyze the organizations. It involves 5 phases: (a) initiate the inquiry, (b) inquire into the best practices, (c) discover the themes, (d) envision a preferred future and (e) design and deliver ways to create the future. You would repeat the process back to (b) in an ongoing cycle to implement that across the team, teams or the organization.

   In this model, we look at current "best" method of work or infusing change and utilize that to move forward with any new change that is causing the organization to be unstable or performing below its peak.

4. Another change model, such as Likert's method to identify and move towards ideal System 4 for participative management was to utilize survey feedback to identify differences to achieve the ideal System 4 conditions. It is clear the use of data, strategy and emphasis on using the internal environment (individuals and teams) in Likert's method to allow for organizational effectiveness and to institute change with minimal resistance.

5. Finally, the QWL (Quality of Work Life) model was all about connecting people and technology and utilizing data with coining of the term "socio-technical" systems. The other phase of the QWL movement was bought together by global competition having relatively low cost and high quality foreign products. It pressed home the value of low cost and high quality resulting from management practices. It bought about well-known programs such as "TQM", Total Quality Management, "EI", employee involvement, and Six Sigma. The goal of such is to recognize the growing emphasis on how teams together with the right data, which is made readily accessible and shared, can unite and contribute to an organization so that it can be more flexible, productive and competitive. It goes without saying that as it gets easier to collaborate, share information and conduct business without any boundaries, as Thomas Friedman (2006) mentions in his book, organizations will need to have efforts to address and adapt to change and Big Data stands as an option that can allow teams and organizations to come together for the bigger purpose.

Organizations exist to provide value to stakeholders by maximizing effectiveness and profitability of the organization. This is in the very definition of organizational development. An OD scholar-practitioner is interested in all the various moving parts that contribute to the effectiveness and profitability of the organization.

What an organization does daily ranges from simple mundane tasks to complex and time-consuming activities. There are likely to be multiple business teams, partners and vendors involved in accomplishment of those varied activities. All organizations exist for providing something – a service, a product, value or even combinations of the others – with the goal to become (and remain) effective or profitable while adding value for the stakeholders. Stringer (2002), in his book, leadership and climate, argues that there exists an organizational climate that provides the motivation necessary for the accomplishment of activities.

Lawrence and Lorsch (1967) talk about organizations and the environment they operate in. The need to understand climate, the environment, and the organization all fall under the scope of Organizational Development. For instance, an example of looking at the organization from the inside to highlight OD's scope, Lorsch and Mathias (2001) talk about how professional firms need ability to manage internally as much as they do for the professional services they provide to the clients. The emergence of new technologies, new processes, threats, regulations and thought leadership all affect the organization more than ever. It is important that OD, whether internal or external, plays a role in allowing the organizations to continue to be effective and profitable.

Unique challenges and opportunities arise as business expands into new countries with new governments, new technology and differing cultures. The values that lay the foundation of OD are being challenged as change and technology collides with business, management and the global economy. According to Hofstede et. al (2010), each individual comes with their own programming in terms of understanding, thinking, feelings and approach of problems and solutions that be had. Similarly, Stringer (2002), points out that the focus went from research validity to managerial validity – they looked to answer the “so what” question. The organization requires the understanding of such differences to ensure that threats and opportunities alike are handled with the utmost care to finding a solution.

This creates an opportunity for OD and its consultants to use its principles and marry the humanistic and business together across the globe. In pursuit of doing so, it does raise certain questions. As Yaeger (2002) asks, “What becomes of OD, as it was founded as a value-laden, primarily American, approach to change?” and “Further, what are these values and do they apply universally?” As organizations grow so that the interaction between differentiation and integration are both global and local practices and are incorporated into an organizational unit that may reflect the true state of affairs (Brewster, 2002; Khilji 2002).

It is a known fact that Fredrick Taylor himself, the father of scientific management, was an external consultant (Weisbord 2012) in the early 1900s and went from organization to organization implementing his meticulous data centric methods. His method of looking at the process and building efficiencies within organizations are well documented (Weisbord 2012). Even the other famous OD personalities, Taylor, Lewin and Likert did what most external consultants do today as they approach problems within organizations they are not part of. They studied the issues at hand; looked to collect data, diagnose the issue and attempted scientific solutions based on observations. Taylor did this with the Steel Mill, Lewin with Hardwood and Religious Affiliates and Likert while heading the Institute of Social Research. As the solutions worked they created a template, in one way or another, that they utilized and approached different organizations to help them become more effective. Such application of OD, with the goal of serving the client - by helping them diagnosing organizational issues, identifying root causes and developing/implementing interventions/solutions that enhance organizational capability and effectiveness – were what defined the birth phase of OD application in the real practical world.

Regardless of being external or internal – consultants need to be able to drive business results. This requires assessment, diagnosis, and confronting the client/team with reality to act and move forward. What made them stand out was that they knew what their strengths were. They did not try to solve every problem but started by observing and collecting data to support their hypothesis. Moreover, Schein, in this book, process consultation talks about process vs content and paints the picture of how there are different methods of consulting regarding the various issues and problems that arise. He goes on to list his various principles (10) in helping this cause for helping, irrespective of internal or external consultant, with number 9 being of importance for this research, all principles listed below:

1. Always try to be helpful
2. Always stay in touch with current reality
3. Access your ignorance
4. Everything you do is an intervention
5. It is the client who owns the problem and the solution
6. Go with the flow
7. Timing is crucial
8. Be constructively opportunistic with “confrontive” interventions
9. Everything is a source of data; errors are inevitable – learn from them
10. When in doubt share the problem.

Similarly, in talking about High Commitment and High Performance (HCHP) organizations, Dr. Beer, breaks it into 5 parts with a focus on data collection to identify key levers for change, multilevel and multiunit systems to transform organizations. These are key to the 3 HCHP pillars – performance alignment, psychological alignment and capacity for learning and change.

In summary, we examined change models and each requires structured, unstructured or semi structured data coming from individuals who contribute to improve the organization. There is leadership (individual, leadership team) who usually initiates/identifies that a change is needed, followed by diagnosis by collection of data, planning and implementation of change and finally the evaluation of the result.

The emphasis is on utilizing various strengths of the data that can be mined, the people involved, and the action to take for organization effectiveness. This is, in its purest form, OD's overall objective in providing value for organizations.

Literature: Current Gap

The various implementations, theories and proof of successes has been singular (tied to single instance within the organization), has only dealt with show casing an area of focus (trying to solve one problem and not organizationally prevalent) and has not been replicable or able to pass on success/learnings to other aspects of the organization. How can I say this with? Well, the current literature provides such evidence:

1. There is very little published management scholarship that tackles the challenges of using Big Data tools—or, better yet, that explores the promise and opportunities for new theories and practices that big data might bring about (George, Haas, Pentland 2014).
2. Church and Dutta (2013) note, "Despite the potential inherent in Big Data driven OD applications to deliver entirely new types of insights for organizations, there are some potential barriers to using this approach as well. These consist of the three different issues: capability, mindset, and ethics."

3. The realm of big data-sharing agreements remains informal, poorly structured, manually enforced, and linked to isolated transactions (Koutroumpis & Leiponen, 2013).
4. The phenomenon of Big Data or Big Data analytics – is one that has been less written about by academics. Moreover, the distinctiveness of Big Data is debatable, with perhaps no totally robust definition available (Calvard 2016).
5. Such transactions, manual methods and isolated incidents act as a significant barrier to the market in data—especially for social science and management research (George, Haas, Pentland 2014).
6. Heterogeneity, complexity and structural variability characterize the 'Big data' concept (Motta et al., 2016)
7. Church and Dutta (2013) very clearly state: "there is no singular Big Data theory, methodology, or value structure" and "just about anyone can enter the consulting space and engage in Big Data mining activity".
8. Big Data is a phase where data are so big and complex it requires more advanced processing technologies to handle it, and can potentially offer organizational insights and sources of value that smaller scales of data processing cannot (Mayer-Schönberger and Cukier, 2013)
9. There is a lot of discrepancy in coming up with defining what Big Data means. Young (2014), Church and Dutta (2013), George, Haas and Pentland (2014) have put together the various definitions of Big Data that

are present across literature and the confusion that exists in defining Big Data

10. There are a lot of challenges that still exist with Big Data. There is the issue with dealing with heterogeneity, inconsistency and incompleteness, varying scale, timelessness, privacy and data ownership and visualization and collaboration. (Jagadish et al., 2014)

Summary

Since the beginning, Organizational Development, has utilized various forms of data to conduct research, formulate hypotheses and contribute value to organizations. Frederick Taylor utilized his scientific method that was modeled in collecting quantitative data. Kurt Lewin, widely regarded as the Father of Organizational Development, developed methods that relied on qualitative (unstructured) and quantitative (structured) collection of data. Rensis Likert and his various systems followed a similar theme in collecting data. The basic point here is that data collection in the field of Organizational Development has been varied, flexible, unstructured and focused on answering the question to ensure effectiveness of the organization. These are the exact reasons why I believe that OD should be at the forefront of Big Data. Big Data as a technology exists to be utilized for the betterment. Granted there are ways that it can be exploited but OD's concern would be to take the leadership position to guide organizations through the betterment of this change. It can be used for better engagement opportunities, better healthcare services, better financial controls, better business regulations and governance and even for enhancing safety and security. For example, as McAfee and Brynjolfsson (2012) note, the United Nations' Global Pulse is an initiative that uses new digital data sources, such as mobile calls or mobile payments, with real-time data analytics and data mining to assist in development efforts and understanding emerging vulnerabilities across developing countries. The gaps identified in literature provide an opening on investigating of how OD can create the bridge between what the organization is looking for and how Big Data with its potential can help them achieve that. At the end, we examined various characteristics that define OD, did the same for Big Data and found elements for further review.

The gap, characteristics and the limited amount of existing data, framework and variables exist concerning successful Big Data Implementations we reviewed further lends to our initial research question, "What impacts successful Big Data Implementations?". Furthermore, the researcher reasons that OD can lead the way for successful Big Data Implementations. Next to investigate this research question, the researcher will utilize mix methods study utilizing grounded theory by interviewing industry experts on successful Big Data Implementations. Then perform qualitative analyses to identify variables to measure via an online survey and finally perform quantitative analyses to answer our research question and proving/disproving our hypotheses.

Chapter 3 Research Methods

In this chapter, we go over the research purpose, method and design. This chapter will also cover participant identification, process, and data gathering.

From the previous chapter, we concluded the following: (a) little existence of quantitative empirical study on Big Data Implementations in the management literature, (b) very little research on theories, makeup, frameworks or definition around variables investigating Big Data Implementations with Organizational Development and (c) due to its infancy, emergence of new challenges that have been bought to light primarily due to lack of standardizations and accountability (Halaweh and Massry 2015). Challenges such as privacy and security have been clearly documented in literature (Wamba et. al 2015; Halaweh and Massry 2015).

Given the above and the early nature of this concept, the main purpose of this research is to investigate characteristics in successful Big Data implementations driven primarily around premise of the OD scholars' arguments regarding OD, as a field of study, leading such innovations and being strategic (Galbraith 2014, Church and Dutta 2013, McAfee and Brynjolfossn 2013, Wamba et. al 2015, Halaweh and Massry 2015). There have been calls suggesting that OD is strategic (Jelinek and Litterer 1988; Head 2009) and experience with many methods that tie in with Big Data historically as evidenced by Weisbord (2012) analysis of OD history.

Unfortunately, a very limited amount of existing data, framework and variables exist concerning successful Big Data Implementations. It was therefore important to formulate a method that would allow us collect data, review, analyze, deduce a model, formulate theory and finally test the phenomenon statistically.

Approach

The best approach for such a study was mix methods utilizing Grounded Theory. Mixed methods allow for the integration of qualitative and quantitative data within a study to provide a more complete analysis of the research problem being investigated (Creswell and Plano Clark, 2011). It allows for, especially for an early concept, data to be built and further explored using a secondary method. Grounded theory allows the researcher to begin with the question, collect data, examine ideas and concepts, extract and categorize that data to use it to form the basis of a new theory. This new theory can then be applied and tested statistically. To successfully accomplish this the approach for the study was fragmented into a two-part mix methods study. A qualitative section utilizing interviews with individuals and organizations in the field followed by a quantitative section to test relationships between core concepts derived from the qualitative section. The qualitative portion of the study was done first, which allowed relationships to be tested later in a quantitative manner using statistical techniques. The knowledge gained through such a process allowed the quantitative section to be further insightful, concentrated and exploratory in nature.

Furthermore, it is important to note that this is a correlational study (Kerlinger & Lee, 2000) as it is being conducted to determine the relationships, as is, for successful Big Data Implementations. Standard strength and direction of relationships between variables is examined and predications provided given the strength and conclusive nature of the variables within the study. The step by step process to investigate the research problem is as follows:

1. The first step was to be formally educated on both these topics. As the researcher was already on the journey to obtain a PhD in Organizational Development it was vital to start learning about Big Data. From a professional standpoint, the researcher works as a consultant and encountering work on Big Data projects is a likelihood scenario with fortune 100/500 organizations. Even so, education was needed to familiarize with various tools and techniques that professionals use in this trade every day. The researcher started by speaking to multiple global, startup and mid-size organizations, signed up for Big Data technical and management classes, attended conferences and seminars, joined LinkedIn professional discussions groups (listed in chapter one) and took up reading the latest on Big Data. All this was done to increase the knowledge and skill level with the goal of being able to conduct interviews and have conversations with professionals. This was an evolving process started in August 2013.
2. The next step was to start conducting interviews with professionals who have worked on and implemented Big Data projects, programs and solutions. This

was the qualitative (Interview) phase. The researcher utilized their personal and professional network to find individuals/professionals and organizations who (a) had implemented Big Data initiatives, solutions, projects and/or programs and (b) who were willing to speak about their experiences which is commonly referred to as the "snowball" technique in qualitative research. The purpose of the interview was to talk about success factors, challenges, how Big Data initiatives were defined and their successful implementations. The initial goal was to speak with roughly 10-30 individuals regardless of industry, profession or location. The interview would be targeted for 60 min with a total of 8 total questions.

3. After conducting the interviews, the researcher would look for common codes that can be grouped into concepts and further into categories of characteristics that can be measured as variables using an online survey.
4. The final step was the creation of the survey. This was the quantitative (survey) portion of the mix methods study. This would then allow the researcher to run statistical procedures to determine various relationships to successful Big Data Implementations.

The integration of the qualitative and quantitative design for this research allowed the researcher to help better understand, compile and relate Big Data Implementations with Organizational Development. This integration, as Creswell and Plano (2011) elude to allows for a single study to provide a more complete analysis of the research question being explored. In other words, we take one set of data, perform analysis and apply our learnings to build the other data set. This helps to further expand on the learnings gleaned from just the primary method. As such, this two-part design allowed the researcher to holistically look at factors impacting successful Big Data Implementations. As we will review here, the qualitative research was conducted prior to the quantitative study. The learnings gathered from the initial qualitative analysis allowed the researcher to create an online survey to statistically analyze the hypotheses and the research question.

Qualitative

The researcher reached out to various individuals across retail, commerce, travel, manufacturing, education, communications, healthcare, finance, high-tech, insurance, and real estate industries using their personal and professional network. Furthermore, the researcher did not limit the interview to only a specific geographic region instead reaching out to individuals from North America (US and Canada), South America (Costa Rica and Argentina), Europe (Ireland, UK and Turkey), Middle East (UAE and Saudi Arabia), Asia (India, Pakistan and China) and Oceania (New Zealand and Australia). The interviews were conducted via phone (3 out of 15) and in most cases, face to face (12 out of 15) due to the sensitivity and confidential competitive advantage information regarding Big Data. There were absolutely no recorded conversations to maintain the confidentiality of the interviews (only researcher and research committee are aware of specifics). Some even went as far as to ensure legal compliance as well. The researcher provided each with reason and nature of the study (template in Appendix C) and ensured strict confidentiality. The researcher also provided the ability for the interviewee to (a) opt-out of the interview at any time and (b) not share specific details if they did not feel they wanted to. The researcher did take notes around contextual and characteristic specific findings (such as technical skills required or involvement of teams, etc) to create a survey to statistically establish and research other relationships.

Research Timeline

The timeline for this qualitative research was from April 2016 to October 2016. In early March, interview questions were shared with the committee to fine tune them. In April, the researcher reached out to various contacts via email, phone and in person (at conferences, seminars and training). Interviews were lined up late April through late October 2016. The researcher, for in person interviews, started travels across North America beginning April 20th through August 17th, 2016. Locations traveled to in North America included Canada (Ontario) and the US (TN, TX, MO, CA and IL). The researcher traveled internationally to Ireland, Turkey, and Saudi Arabia from August 28th to September 17th, 2016.

Sample and Data Gathering

The researcher had reached out to more than 30 individuals and not everyone responded. Once a total of 15 individuals were identified the researcher stopped recruiting and focused on setting logistics for the interviews. Different scholars showed that data saturation was reached between 10-30 interviews and in some cases main themes and concepts were identified within 6 interviews and 100% of codes were identified by the thirteenth interview (Marshall et. al 2013 reference Glaser and Strauss 1967 and Thompson 2010; Guest et. al 2006).

The interviews were 60 min in length comprising of 8 total questions. The questions were broad in nature where the interviewee had to elaborate and explain their responses. As we have previously established, Big Data is still in its infancy, so the researcher included questions comparing Big Data with other new technology or related initiatives that the organization was undertaking and how they were measured. The goal was to understand the full spectrum of what the organizations were trying to do holistically and establish a baseline to compare a new concept (Big Data) with other more established initiatives (new technology rollouts or process engineering).

The following questions were compiled for the interview:

1. What are some of the initiatives that your organization is currently undertaking? Please elaborate on the variety of initiatives.
2. Of the initiatives you mentioned above, what is the organization trying to achieve with these initiatives? Do you categorize these initiatives and if so, can you give examples (ex: technology, business, marketing, etc)?

3. What are some success criteria that the organization uses to measure these initiatives? Furthermore, in your previous experience are these characteristics similar, different or a combination?
4. How do you define Big Data initiatives?
5. What characteristics would you identify with Big Data initiatives in your organizations that you believe will help with the success of these initiatives? Please elaborate on what success criteria would be for the organization and for you individually.
6. What would you do differently knowing what you know today about Big Data initiatives?
7. Can you describe examples, from your experiences or from what you have heard/know, of any unsuccessful Big Data or technology initiatives that were undertaken?
8. Do you have any closing thoughts or comments that you would like to share?

The table below lists a summary of the interviews. The full interview template can be found in Appendix D:

Table 1. Qualitative Sample results from Interviews

Industry	**Region**	**Focus**	**Reason**	**Interview Type**
Insurance (3)	North America	Organization and Technical	Innovation and business	Phone and In Person

			need	
HealthCare (1)	North America	Organization and Technical	Innovation	In Person
Retail, Commerce (2)	North America	Technical	Innovation	In Person
High Tech (2)	Middle East, Asia,	Technical/Security	Innovation	In Person
Travel (1)	North America	Organizational	Business need	In Person
Hospitality (1)	North America	Organizational	Business need	In Person
Communications (1)	Middle East	Technical/Security	Business need	In Person
Manufacturing (1)	North America	Organizational	Innovation	Phone
Education (1)	North America	Technical	Business need	Phone/Virtual Chat
Finance (2)	North America, Europe	Organizational and Technical/Security	Business need	In Person

The table above has information summarized into 5 columns. Industry, region and interview type are self-explanatory. The numbers in parenthesis in the industry column are the number of different organizations the researcher spoke with. The two columns on the right, focus and reason, were created to understand the driver, success and need for Big Data implementations. They were grouped into the representative categories.

The goal of the qualitative portion was to discover factors (i.e. criteria, variables, elements and characteristics) that influence successful Big Data Implementations. Furthermore, this discovery would lead to a model being induced to allow us to test and establish statistical relationships between the variables. To accomplish the above goal, the extraction of variables from interviews was completed in a total of five rounds, as described below, to induce a model for testing. In the first two rounds, the researcher specifically extracted and categorized success criteria. In the next three rounds, all items that can impact successful Big Data Implementations were extracted, categorized and examined for patterns. Details around keywords extracted are all listed in Chapter 4 as part of findings and analysis. The method followed for the extraction is as follows:

1. Round 1: Raw data review to define "what does successful Big Data Implementation appear as".
    - This is what the result is at the end of an implementation. The research only focused on the successful implementations. In other words, to find the characteristics that impact successful Big Data Implementations, which is the research objective, we need to identify

what success means for Big Data Implementations. This identifiable success will be what we measure in our model. This would be our dependent variable or outcome in the model. As this outcome is dependent on other characteristics for it to come to fruition. For instance, talk about a new resulting product due to a Big Data initiative would be classified as a success item to measure for Big Data Implementations. The researcher identified these as codes.

2. Round 2: Revision of codes to fit into categories.
   - In this round, the researcher reviewed all the identified codes to see if they could be grouped into categories. Categories represent similar themes. For instance, in the interviews, the interviewee's spoke about a product they developed, a new tool or a new service that was provided for their members. Through these separate examples, they represent a new offering that resulted in a successful Big Data Implementation.

3. Round 3: Revision of the categories to see patterns and themes for measurement.
   - Finally, in this third round, the researcher reviewed for patterns or major themes that emerged from raw data, codes and categories.

Moreover, everything else that impacts the success item would be deemed an independent variable. An independent variable is a variable on which its variation or result is not dependent on another item. It stands alone and is not impacted by other variables. For instance, age is an independent variable which does not depend on anything for its result. Similarly, after the above two rounds of understanding what successful Big Data Implementation means, we inspect the interviews for what will create this success. In other words, we want to find what causes successful Big Data Implementation. For instance, if new products are identified because of a successful Big Data Implementation, we investigate causality that allows us to get to that result.

After having completed rounds one and two to identify success criteria (what we are measuring), the next step was to find what such success actual depends on.

4. Round 4: Raw data review to find keywords that call out characteristics that can impact the success of Big Data Implementations.
    - Here the researcher did a full review to find any characteristic that can impact successful Big Data Implementations. For instance, the interviewee's spoke about focus on information, data and leadership abilities in terms of impact. All these were quotes were extracted and identified as codes, just like in round 1. This first pass was to capture everything possible that would help answer the research question.

5. Round 5: Review of the codes to fit into categories.

   - Major themes would act like broad categories where codes can fit in and can be grouped together to represent the initial set of codes. Similar in nature to round 2 but this time focusing on the codes developed from round 3, above. The only difference here being that context was important to consider and extract as well. For instance, when interviewee's talk about focus on information and data it might be easier to group them together but keeping context in mind resulted in different categories being identified for each. One being, where focus on information gathering was important versus ownership of data. As you can see both are different concepts and thus put in different categories.

6. Round 6: review for patterns or themes

   - Finally, in this last round, the researcher reviewed for patterns or major themes that emerged from raw data, codes and categories. For instance, the interviewee's spoke about skills, tools and platform from a technology standpoint. As such the major theme, here, could be classified as technical competence of the organization.

Again, Chapter 4, contains all the findings and analysis for each of the above rounds with the goal of identifying variables to test in our model for successful Big Data Implementations.

Summary

We covered the research design, sample, timeline and data gathering techniques in the first part of this mix methods study. This portion allowed the researched to formulate hypotheses, derive communalities (discussed in the findings chapter – Chapter 4) and gather first hand insights through the interviews. We discuss the same for the second portion of the study, the online survey.

Quantitative

For the quantitative portion, the method of research used was survey research. The intent was to measure characteristics representing the population using statistical techniques (Kerlinger & Lee, 2000; Creswell & Plano Clark, 2013). Survey research is a method of understanding aspects of behavior through statistical analysis from a sample of the population. The statistical analysis allows comparisons and strength of relationships between variables to the hypotheses being tested.

Process and Compilation

The use of an online survey tool, SurveyMonkey®, enabled the researcher to gather information globally. The survey comprising of a total of 50 questions using a Likert scale was shared electronically via email and professional networking messaging and groups (LinkedIn). While confidentiality was maintained to not tie responses that would identify individuals there were, however, separate links used to track submissions from the various sources. For instance, separate links for email submissions were used as compared to LinkedIn.

Of the 50 questions; 6 of were demographics related, 2 were targeted conditional questions used as a check to ensure the individual can contribute to the survey and the remaining 42 were questions surrounding variables captured from our interviews. Of the 42 variables, they were divided into (a) Big Data Characteristics (8 questions), (b) Technical Competence (8 questions), (c) Successful Big Data Implementation (4 questions), (d) Organizational Characteristics (18 questions) and (e) Miscellaneous (4 questions) that were used to get general agreement on Big Data.

The variety of perspectives shared in the interviews made clear that to gather meaningful data and use it for research in successful Big Data implementations the researcher had to make sure to ask two conditional questions before a responder can take further part in the survey. The reasoning behind the 2 targeted conditional questions was to maximize data collection from organizations/individuals who (a) knew what Big Data was and (b) had the experience with Big Data implementations. Only implementing part one of the above seemed incomplete given that the research's objective was to study successful Big Data Implementations. It was important, however, to understand Big Data before the act of implementation. While the researcher understood that the understanding for Big Data can be subjective the implementation of such was less so and hence the addition of the second question to ensure quality of responses captured this aspect for this research.

Given the above condition, the online survey was divided into 3 pages with the first page mostly capturing required demographics data and asking the two conditional yes/no questions. If the response to one of the conditional questions was no, the survey stopped there and thanked the participant for their time. There was no other way to go forward or back. You could not re-click the link or redo the survey as that session would be expired.

Timeline and Sample

The survey was live from Jan 3, 2017 to March 10, 2017. Pilot and testing was conducted from Jan 3rd to Jan 20th, 2017. The survey was shared with a total of 669,361 unique individuals across the globe beginning on Jan 21st, 2017. The online survey distribution is listed below:

Table 2. Online Survey Sample Distribution

Group/Membership	Total Individuals
Big Data in Insurance	444

IoT, Big Data in Canada	131
Big Data in the Pharmaceutical Industry	710
Big Data for Music Industry	73
Big Data Challenges	383
Big Data Analytics for Financial Institutions	278
Predictive Analytics, Big Data Business Analytics, Data Science in Oil and Gas and Energy	904
Big Data and Analytics Group	280,707
Business Intelligence Professionals (BI, Big Data, Analytics, IOT)	192,761
Big Data, Analytics, Business Intelligence & Visualization Experts Community	192,346
Email	504
Professional Contacts (Targeted via Survey Monkey)	120
	669,361

Survey Pilot and Testing

There were 2 main goals for piloting the survey: (a) clear understanding of the survey content/questions and (b) overall survey functionality (ease of, total time, format and flow of survey). During pilot and testing (Jan 3rd to Jan 20th, 2017), it was observed that the average time to complete the survey was less than 5 min. The survey initially had a total of 36 questions of which 6 were demographics and 2 were conditional questions designed to ensure reliability and accuracy of the data gathered to measure successful Big Data Implementations.

Upon feedback, 10 survey questions had wording changed for better focus on single constructs. Fourteen questions, some specifically to help with defining Big Data, were added to get to a total of 50. Other edits, including, randomization of the order of questions, required answering of certain questions and addition of the "other" choice free flow textbox right below the dropdown questions were added. The survey was also divided in to 3 pages for ease of navigation.

Instrumentation

The survey was distributed via SurveyMonkey®. An upgraded plan was purchased by the researcher to obtain additional features for custom distribution, design, confidentiality, support, data gathering and analysis. Aside from demographic and conditional questions, the survey question asked each respondent to reply with "strongly disagree," "disagree," "neutral," "agree," and "strongly agree." Participants indicated their response by clicking the circle next to their desired response. The core construct related questions were randomized in order each time a unique participant took the survey. Although IP, computer, and session tracking capture was disabled (only enabled for manual additions) the survey was strictly allowed to be completed once by a unique participant. The online survey started by capturing basic demographic data with participants selecting responses from a pre-defined drop down. On the same page, the two conditional yes/no questions were added before clicking next. Following the responses gathered the participant was either thanked for their time or moved onto page 2 comprising of the core questions. Data was captured through the tool itself and then exported to SPSS version 24.0 for analysis.

Demographic items

For both a control perspective and to also capture general information for further analysis, several demographic items were captured at the start of the survey. These include profession (Experienced, Professional/Technical, Managerial, Administrative), functional area (business, IT, PMO, Operations, etc), age (21-29, 30-39, 40-49, 50-59, 60+), geographic region (North America, South America, Europe, Asia, Middle East, Africa and Oceania), gender (Male, Female) and industry (HealthCare, Finance, etc).

Data Collection

Individuals completed surveys online collected securely using TLS encryption. Data collected through the SurveyMonkey® tool was securely stored on their servers. Data access was only available to the primary research and committee members through secure log on using a password. Data was exported to Microsoft Excel and SPSS for analysis and kept confidential on a password protected computer. The statistical package chosen for the current study was SPSS version 24.0. Use of SPSS provides for a variety of statistical analyses including descriptive and inferential statistical procedures.

Source data was further transcribed onto a removable storage device and transmitted to a Benedictine University faculty member for security purposes and ultimate disposal after a period of not less than 7 years. The primary researcher and committee were the only ones to have access to any data and only from respondents that agreed to sharing.

Summary

This chapter described the study design and data collection process for this mixed methods research. The research methods chosen for this study guided the researcher in exploring the hypotheses in multiple ways. First, the qualitative research helped to identify the relationship. Informed by the results of the first portion of the study, the quantitative researched helped to take a closer look at the relationship between those variables as well as gain additional insights through first-hand accounts during interviews. The next chapter will provide the details of the results of the study and analysis of the findings.

Chapter 4 Analysis and Findings

The purpose of this chapter following the methods chapter is to present the results and analysis of the data. As indicated earlier, this is a two-part mixed methods study. In part one, we discuss the results of the qualitative section followed by part two, where we discuss the quantitative results.

Qualitative Examination of Research Questions

The purpose of the qualitative research was to explore factors, criteria and characteristic of Big Data Implementations. It was conducted to get a holistic picture of the organization, its strategy, environment and other factors that may influence the success of Big Data Implementations. The literature provided initial exploratory discussion regarding Big Data. It was important to further explore these perception and discussions within the field to document reliability and other concepts.

Findings

Interviews

A total of 15 interviews were conducted. While the interviews were capped at 60 min each only 2 interviews finished early or on time. The others all lasted longer than 60 minutes and a few even extended longer than 120 minutes. The interviews conducted were either on the phone or in person. A total of three interviews were conducted on the phone while the other twelve were conducted in person.

Coding Results

As stated earlier, given the nature, topic and to explore all possibilities that can be found within the data, a grounded theory approach to coding and subsequent theory creation was used (Cresswell, 2011) comprising of multiple reviews of the raw data. There are certain themes that we did see from literature (Wamba et al 2015; Miyanka 2011, and Halaweh and Massry 2015) and the researcher did look to tag those themes after coding process was completed. Open coding was done based on ideas and constructs discussed by question. Given the length of some interviews and interviewee's hesitation in being recorded, this method for open coding was utilized. The basic workflow followed was:

1. Part 1: Multiple rounds (6) to extract variables and induce a model, as described in Chapter 3
2. Part 2: Review of current literature for Big Data to compare reliability, findings and emergence of any new themes.
3. Part 3: Creation of a test model and survey using core constructs
4. Part 4: Finally, forming hypotheses for testing of relationships of model.

Part 1: Multiple Rounds to extract variables for testable Model

As discussed in Chapter 3, variables were extracted through a 6-round process. Recap of each round and results are discussed below:

Round 1: Raw data review to define "what does successful Big Data Implementation appear as". The goal here was to identify what to measure and what the result looked like. This identifiable success will be what we measure in our model. This would be our dependent variable or outcome in the model. These would be identified as codes. When interviewee's spoke about Big Data enabling them to create new patents or guide them towards insights that led to new innovative products the researcher would label these as "Patents" or "innovation" codes to identify the result of what was enabled by successful Big Data implementation. The table below lists the other identification of codes from the various excerpts taken from the interviews. Here is the result after the first round of raw data review:

Table 3. Qualitative Review – Round 1 Examination (codes)

Quotes from Interviews:	Codes:
...leads to creation of proprietary systems to help organization with its processes and delivery of services	Processes Improvement… New Services
...newly added services that add value provided for our members...	New value provided for systems New service provided with value
...new patents and innovation…	Patents, innovation
…more services to capture market share as strategy…	New service, strategy
…better visualization tools developed to help the company with initiatives, process and change	New tools, initiatives focus, processes and help with change
… allow for organization to provide better care services, with more accuracy, flexibility and be prescriptive	New services that are better
...give more value with the implementation	Added value (for product or service)
...gain competitive advantage as a result of our Big Data Initiatives	Competitive advantage
…build better diagnostic products…	New Products that are better
We may potentially pursue a different strategy	Different strategy

Round 2: revision of the codes to identify categories they can fit in. Looking at the list of codes above it is plausible to group items into categories for measurement. To recap, the goal of the first two rounds, as discussed in Chapter 3, was to identify what success looks like for Big Data implementations. In reviewing the above codes most fit into categories. For instance, there were multiple instances of using Big Data to create new products, offer new services or create new tools to either help the organization or its customers. The above instances are focused on the result of offering new products, services and tools. As such, these can be used to define what success in Big Data Implementations appear as. These were labelled as "Success (Categories)" and added right after the Codes column. The table below shows where we end up with after the second round:

Table 4. Qualitative Review – Round 2 Examination (Category)

<table>
<tr><th>Quotes</th><th>Codes - Round 1</th><th>Success (Categories) - Round 2</th></tr>
<tr><td>...leads to creation of proprietary systems to help organization with its processes and delivery of services</td><td>Processes improvement… New Services</td><td rowspan="10">New Services
New Products
Patents
Process Improvement
Strategy
Competitive Advantage</td></tr>
<tr><td>...newly added services that add value provided for our members...</td><td>New value provided for systems
New service provided with value</td></tr>
<tr><td>...new patents and innovation…</td><td>Patents, innovation</td></tr>
<tr><td>…more services to capture market share as strategy…</td><td>New service, strategy</td></tr>
<tr><td>…better visualization tools developed to help the company with initiatives, process and change</td><td>New tools, initiatives focus, processes and help with change</td></tr>
<tr><td>… allow for organization to provide better care services, with more accuracy, flexibility and be prescriptive</td><td>New services that are better</td></tr>
<tr><td>...give more value with the implementation</td><td>Added value (for product or service)</td></tr>
<tr><td>...gain competitive advantage as a result of our Big Data Initiatives</td><td>Competitive advantage</td></tr>
<tr><td>…build better diagnostic products…</td><td>New Products that are better</td></tr>
<tr><td>We may potentially pursue a different strategy</td><td>Different strategy</td></tr>
</table>

At the end of two rounds, we have sense of what we can use to test for success in Big Data Implementations and they are: new services, products, patents, process improvement, strategy and competitive advantage. All a result of successful Big Data Implementations from our set of interviews.

Round 3: Here the categories were examined to see if there were any themes or patterns. We see that new products and services can be grouped together as they are considered new value add to the organization. Patents and competitive advantage can be grouped together as both provide organizations added advantage over competitors and in the market place. Finally, process improvement and strategy were kept separately. At the end of three rounds for us to induce a model for statistical testing relationships, we have our set of dependent variables as the figure below shows.

Figure 4: Dependent Variables after Round 3 of Qualitative Review

Dependent Variable
Successful Big Data Implementations
Competitive advantage/Patents
New Products/Services
Processes Improvement
Strategy

Round 4: Review raw data to find keywords that call out characteristics that can impact the success of Big Data Implementations. Now that an understanding of what success looks like for Big Data Implementations, in this round a review of the raw data was done to find characteristics that can influence success for Big Data

Implementations. Again, as in the first round, after review and end of round 4, the raw data items will be labeled as codes for further review. The table below lists results at the end of round 4:

Table 5. Qualitative Review – Round 4 Examination (codes)

Quotes	Codes
…there is a need to understand what actually Big Data is. We define it a certain way to include volume of data, speed (velocity), use of computing resources, variety and purpose it served. We measure against that.	Define Big Data
We started to experiment with a lot of the open source available. Just being able to run our data using R or Python to slice and dice the data was helpful to get started	Open Source programming, Coding techniques, Technical skills and know-how
Hadoop has been our choice of Big Data file system to work with but have found that involvement from the business really helps us map what we are looking for.	Tools – software package, collaboration between business teams and their involvement
look, clearly, it is impossible to do this on our own, we need to make sure that leadership understand that this development will need some development and nurturing before churning out products	Leadership and executive support around such programs
A much more developed focus on security and privacy has led the way in most of our Big Data initiatives. We rely on a lot of ingestion of data to be able to pick fraudulent activities, suspicious behavior and attacks. Even so, privacy of user information, especially what can identify them is of the utmost importance.	Focus on a specific matter to help other areas, privacy of data
At one point, we didn't know what data we had till we embarked on a trail of creating visual dashboards to help visualize what data is available to us. A technology driven platform has allowed us to really look at possibilities that it can afford us in the future.	Focus on building a technology platform, focus on information
There are lot of places where we get our data from and what has helped is pushing ownership of various types of data to the stakeholders. This has allowed creation of governance and the other buzz words you hear in the industry around managing content.	Stakeholder ownerships of data, how to govern data and its subsequent ownership
the… team owns that data and they will clean and maintain it for us. This is just good practice and maximizes what we are able to do with such.	data ownership and accountability held outside of information technology or services
…much confusion on what Big Data means…	Big Data definition confusion

I believe we engage in Big Data due to the sheer volume and type of data we interact with. The key here is having a formal methodology. It allows for a much higher success rate. The reason being is that there is structure around inquiries of Big Data Initiatives without a lot of upfront investment in skills, technology or setup.	Following of a methodology (formal), discretionary funds
We always captured and created a lot of information about our assets. It's the way we have been successful at some of our Big Data projects.	Focus on information gathering
Even before Big Data, we used to comb through our data to look at fraud and threats. Now with advances in technology like cloud computing, non-relationship stores, and explosion of mobile and internet connected devices we are able to capture more data to be more accurate in predicting outcomes, behaviors and false positives.	focus on information gathering, technology advances, technology platform to include cloud and Internet of Things devices

As you can see from the above table, there are a mix of items as discussed by the interviewee's. There was confusion about defining Big Data, the focus on certain organizational contexts, technical skill importance and a few new concepts and challenges identified, such as privacy.

Round 5: review codes to fit into categories. Similar to what was done in round 2 and as explained in Chapter 3, the goal here is to fine tune and identify if there are general categories that these codes fit in with. See table below with added categories for each row:

Table 6. Qualitative Review – Round 5 Examination (category)

Quotes	Codes	Categories
…there is a need to understand what actually Big Data is. We define it a certain way to include volume of data, speed (velocity), use of computing resources, variety and purpose it served. We measure against that.	Define Big Data	Definition Big Data

We started to experiment with a lot of the open source available. Just being able to run our data using R or Python to slice and dice the data was helpful to get started	Open Source programming, Coding techniques, Technical skills and know-how	Technical Skills
Hadoop has been our choice of Big Data file system to work with but have found that involvement from the business really helps us map what we are looking for.	Tools – software package, collaboration between business teams and their involvement	Technical Tools
look, clearly, it is impossible to this on our own, we need to make sure that leadership understand that this development will need some development and nurturing before churning out products	Leadership and executive support around such programs	Leadership Support
A much more developed focus on security has led the way in most of our Big Data initiatives. We rely on a lot of ingestion of data to be able to pick fraudulent activities, suspicious behavior and attacks.	Focus on a specific matter to help other areas, privacy of data	Business involvement, cross functional collaboration, Security, privacy
At one point, we didn't know what data we had till we embarked on a trail of creating visual dashboards to help visualize what data is available to us. A technology driven platform has allowed us to really look at possibilities that it can afford us in the future.	Focus on building a technology platform, focus on information	Technical Platform
There are lot of places where we get our data from and what has helped is pushing ownership of various types of data to the stakeholders. This has allowed creation of governance, taxonomy and the other buzz words you hear in the industry around managing content.	Stakeholder ownerships of data, how to govern data and its subsequent ownership	Data Ownership, governance
the… team owns that data and they will clean and maintain it for us. This is just good practice and maximizes what we are able to do with such.	data ownership and accountability held outside of information technology or services	Data Ownership, governance
…much confusion on what Big Data means…	Big Data definition confusion	Definition Big Data
I believe we engage in Big Data due to the sheer volume and type of data we interact with. The key here is having a formal methodology. It allows for a much higher success rate. The reason being is that there is structure around inquiries of Big Data Initiatives without a lot of upfront investment in skills, technology or setup.	Following of a methodology (formal), discretionary funds	Methodology
We always captured and created a lot of information about our assets. It's the way we have been successful at some of our Big Data projects.	Focus on information gathering	Information focus

Even before Big Data, we used to comb through our data to look at fraud and threats. Now with advances in technology like cloud computing, non-relationship stores, and explosion of mobile and internet connected devices we are able to capture more data to be more accurate in predicting outcomes, behaviors and false positives.	focus on information gathering, technology advances, technology platform to include cloud and, Internet of Things devices	information focus, IoT devices, Technology Platform

Round 6: Upon reviewing the categories, it was very easy to see some emerging themes. In fact, after the completion of the five rounds there appears to be a total of 3 major themes that emerged. The themes are: organizational characteristics, labelled (a) in table below (examples of key words include: methodology, management support, processes, reliance on information or information intensity, size of the organization, cross functional collaboration, governance and funding), Technical characteristics, labelled (b) in table below (examples of key words include: technology platform, types of data, software and tools available to use and technology management), and Big Data definitions, labelled (c) in table below (definition would include volume, variety, velocity, etc. but there was no agreement across the board around what Big Data was). Furthermore, there were some outliers that included broad topics that I added to a Miscellaneous section, labelled (d) in table below, which were mostly out of scope of the research but nevertheless bought up in discussion. These included: privacy and security. At the end of 6 rounds, this is what our table looks like:

Table 7. Qualitative Review – Round 6 Examination (themes)

Quotes	Codes	Categories	Themes
…there is a need to understand what actually Big Data is. We define it a certain way to include volume of data, speed (velocity), use of computing resources, variety and purpose it served. We measure against that.	Define Big Data	Definition Big Data (c)	(a) **Organizational Characteristics** Leadership Collaboration Methodology Size of Organization Business involvement Data ownership Governance Information Intensity (b) **Technical Competence** Platform Skills (c) **Big Data define** Volume Variety Velocity Veracity Value Variability (d) **Miscellaneous** Privacy Security
We started to experiment with a lot of the open source available. Just being able to run our data using R or Python to slice and dice the data was helpful to get started	Open Source programming, Coding techniques, Technical skills and know-how	Technical Skills (b)	
Hadoop has been our choice of Big Data file system to work with but have found that involvement from the business really helps us map what we are looking for.	Tools – software package, collaboration between business teams and their involvement	Technical Tools (b)	
look, clearly, it is impossible to this on our own, we need to make sure that leadership understand that this development will need some development and nurturing before churning out products	Leadership and executive support around such programs	Leadership Support (a)	
A much more developed focus on security has led the way in most of our Big Data initiatives. We rely on a lot of ingestion of data to be able to pick fraudulent activities, suspicious behavior and attacks.	Focus on a specific matter to help other areas, privacy of data	Business involvement, cross functional collaboration (a), Security, privacy (d)	
At one point, we didn't know what data we had till we embarked on a trail of creating visual dashboards to help visualize what data is available to us. A technology driven platform has allowed us to really look at possibilities that it can afford us in the future.	Focus on building a technology platform, focus on information	Technical Platform (b)	
There are lot of places where we get our data from and what has helped is pushing ownership of various types of data to the stakeholders. This has allowed creation of governance, taxonomy and the other buzz words you hear in the industry around managing content.	Stakeholder ownerships of data, how to govern data and its subsequent ownership	Data Ownership, governance (a)	
the… team owns that data and they will clean and maintain it for us. This is just good practice and maximizes what we are able to do with such.	data ownership and accountability held outside of information technology or services	Data Ownership, governance (a)	
…much confusion on what Big Data means…	Big Data definition confusion	Definition Big Data (c)	
I believe we engage in Big Data due to the sheer volume and type of data we interact with. The key here is having a formal methodology. It allows for a much higher success rate. The reason being is that there is structure around inquiries of Big Data Initiatives without a lot of upfront investment in skills, technology or setup.	Following of a methodology (formal), discretionary funds	Methodology (a)	
We always captured and created a lot of information about our assets. It's the way	Focus on information gathering	Information focus	

we have been successful at some of our Big Data projects.			
Even before Big Data, we used to comb through our data to look at fraud and threats. Now with advances in technology like cloud computing, non-relationship stores, and explosion of mobile and internet connected devices we are able to capture more data to be more accurate in predicting outcomes, behaviors and false positives.	focus on information gathering, technology advances, technology platform to include cloud and, Internet of Things devices	information focus, IoT devices (a), Technology Platform (b)	

At the end of the six rounds, we can conclusively identify success criteria for Big Data Implementations and we have criteria that can influence that success. These criteria are known as our independent variables, which are not impacted by variation but instead impact the dependent variables.

Figure 5: Model after Qualitative review (6 rounds)

Independent Variables
Organizational Characteristics
Leadership Support
Collaboration
Methodology
Size of Organization
Business involvement
Data ownership
Governance
Information Intensity
Technical Competence
Skills
Platform
Tools
Miscellaneous
Privacy
Security

Dependent Variables
Successful Big Data Implementations
Competitive advantage/Patents
New Products/Services
Processes Improvement
Strategy

It is important to note that the above were the ones that were straight forward and were called out specifically by the interviewees. There did exist other interesting patterns that needed to be confirmed as well, they are:

First, the real interesting part of the conversation was the different ways everyone spoke about their Big Data projects. Some defined it purely in terms of characteristics, such as velocity and volume of data coming in (Fraud detection) or using data type (Risk analysis services, Security products) that will allow business to provide new services or create new products and others referred to it from an analytics perspective (visual dashboards) to make sense of the data was about. The importance of using computing resources in Big Data initiatives was stressed as well. It was apparent that without such computing power it would be near impossible to gather data, process and examine it to create successful implementations. It is also important to note that only one interview described actual use cases for more than three characteristics of Big Data. The main goal behind that was how they are using such characteristics in developing efficiencies (route optimizations) and offering new products (road safety). The reason this is interesting is because this means that even before working on Big Data projects or inquiring about success and challenges of such, clarification is needed for what we were using as a baseline to define Big Data and further if any evolution would take place. This meant that in our model, as setup above, to engage in Big Data initiatives, we need to ensure Big Data definitions were met. In other words, they did not really impact the success of Big Data Implementations but characterized what Big Data Implementation means. This was an important discovery and was consistent with how Halaweh and Massry (2015) deduced their theoretical model. Albeit, the characteristics have since evolved and it was necessary to measure all items called out to get a baseline of the Big Data definition to engage in such initiatives. As such, all six Vs were listed to measure.

Second, industry seemed to be key when discussing Big Data. Everyone spoke how implementations within their industry and attributed certain characteristics such as, maturity, to other industries. They used them to look at various models and come up with other ideas. For instance, the manufacturing industry spoke about certain innovations they were working on but referred to the insurance industry as a more mature setup for Big Data initiatives and programs. The key thing here was that although organizations frequently looked outside their industry to look at trends these did not appear to impact the success of Big Data Implementations. It is important to note that geography did not play much of a role here, meaning, across the world not one interviewee spoke about how being in different places of the world would impact successful Big Data implementations.

Third, another item that was mentioned over multiple times was the importance of technical platform. There was a time, cost and availability aspect tied to this. In other words, short time to bring up multiple environments, work on developing custom or creating improvements, software and tools available for use and testing in a variety of formats for different times, audiences and business segments. It was apparent that this was not just a list of wants from the information technology group. These were key pieces that would allow business ideas to "fail fast", "experiment quickly" and "make decisions on data in real time".

Fourth, the difference when speaking with technical and non-technical individuals was also very apparent. Non-technical individuals spoke more about strategy, vision and usefulness of data. Comparatively, technical individuals focused more on tactical approaches such as "operations", "flexibility in performing" actions and kept "focus on building scalability for other projects". Profession seemed to play a role in how success was viewed in Big Data Implementations. Non-technical users were happier when able to use data to answer questions as compared to technical folks who were more tactical in approaching and considered implementation of technology more of a success factor.

Finally, one last aspect that caught my attention was the topic of having data owners and stewards. While this subject is not new but the application of this term in Big Data projects was used more from a maintenance and security perspective where for once information technology was not solely responsible for it. Multiple interviewees mentioned "who owns the data", "you need to get access from the owner of that Data" and "… owns that data and they will clean and maintain it".

In summary, when all the raw data is referenced we end up with the following model available for us to test, study and resolve our curiosity surrounding successful Big Data Implementations:

Figure 6: Second Model after final inspection for Qualitative review

Big Data Characteristics
Volume
Variety
Velocity
Value
Veracity
Variability
Computing Resources
Evolution

Independent Variable
Organizational Characteristics
Leadership Support
Collaboration
Methodology
Size of Organization
Business involvement
Data ownership
Governance
Information Intensity
Technical Competence
Skills
Platform
Tools
Miscellaneous
Privacy
Security

Dependent Variable
Successful Big Data Implementations
Competitive advantage/Patents
New Products/Services
Processes Improvement
Strategy

Part 2: Literature Relationship and Contribution

Big Data Implementation has several factors at play to ensure success. Qualitative analysis through interviews revealed a lot of factors that have played a role in Big Data Implementation. As Halaweh & Massry (2015) suggest, factors need to be handled properly so that they do not become challenges in Big Data Implementations. Given the above results, it was important to review the literature to (a) look at any consistency of findings (if available) and (b) document new findings so that hypotheses and relationships can be examined in part two (quantitative portion) of this study. The table below summarizes what we find in literature comparatively to what is gathered from part one of this mix methods study:

Table 8. Qualitative Review Comparison w/Literature

Literature	Source	Part One Findings (Categories)
Data Policies	LaValle et. al (2011); Manyika et. al (2011)	Governance
Organizational Change & Talent	McAfee & Brynjolfssn (2012); Manyika et. al (2011)	Change and Technical Competence
Access to Data	Lavalle et. al (2011); Manyika et. al (2011)	Data Ownership & Governance
Industry Structure	LaValle et. al (2011); Manyika et. al (2011)	Industry
Data Availability & Quality	Halaweh & Massry (2015)	Data Ownership, Governance
Top Management Support	Halaweh & Massry (2015)	Leadership Support
Infrastructure	Halaweh & Massry (2015); Manyika et. al (2011)	Technical Competence
Required Skillset	Halaweh & Massry (2015); Manyika et. al (2011)	Technical Competence
Privacy	Halaweh & Massry (2015)	Privacy
Security	Halaweh & Massry (2015)	Security
Competitive Advantage	Halaweh & Massry (2015)	Competitive Advantage
Innovating new business models,	McAfee & Brynjolfssn	New Products, Services,

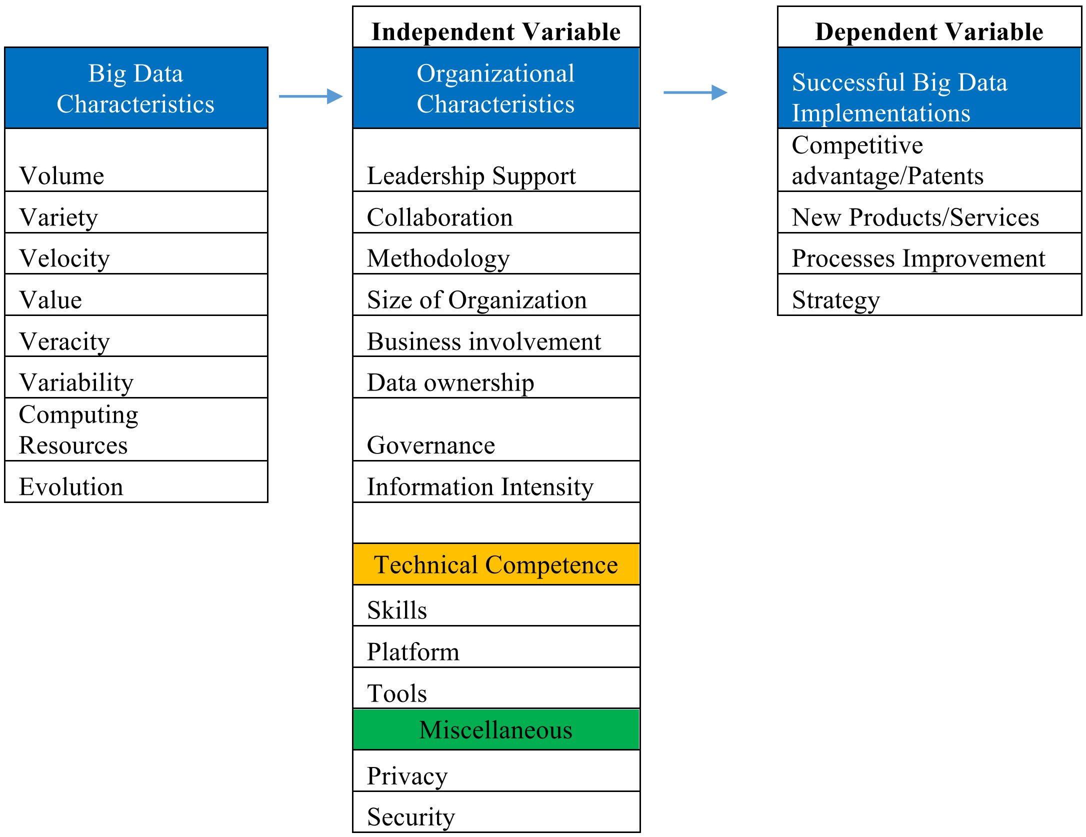
products and services	(2012); LaValle et. al (2011); Manyika et. al (2011)	Processes
Automating Algorithms	McAfee & Brynjolfssn (2012); LaValle et. al (2011); Manyika et. al (2011)	Process Improvement
Create transparency	McAfee & Brynjolfssn (2012); LaValle et. al (2011); Manyika et. al (2011)	Process Improvement
enable experimentation for improved performance, discover needs and expose variability	Lavalle et. al (2011); Manyika et. al (2011)	New/Improved Processes and Big Data definition: Variability

The table above does validate some of our findings as it relates to successful Big Data Implementations. It is important to note that the above scholars, as Wamba et al (2015) point out, were more focused on value creation from Big Data and more specifically around analytics fueled innovation from Big Data. In this study, from the conducted interviews, there are a few other categories that are added to the above list as impacting/influencing successful Big Data Implementations, namely: Methodology, Strategy, Business involvement, Cross functional collaboration, Information intensity, Technical platform to include data, cloud and infrastructure and Technical skills to include analytics, programming, mining and maintenance. We also find the emergence of new Patents, competitive advantage and strategy as a qualifier of success of Big Data Implementations. Furthermore, there are three more additions to the Big Data definition, to include: Value, Veracity and use of Computing Resources.

Part 3: Testable Model and Core Constructs

Given the above, we finalize our model as follows for statistical testing in our study:

Figure 7: Final Model for testing successful Big Data Implementations



Survey Questions

The table below lists all the variables gathered from part 1 along with the core construct questions in the survey. One limitation for using Survey Monkey to distribute surveys to targeted audiences around the world was not being able have more than 50 questions.

Table 9. Qualitative Review – Online Survey

Category	Question
BigDataGeneral_Under	Big Data is a concept that is currently misunderstood?
BigDataCompute	Use of computing resources is an absolute necessity when engaging in Big Data initiatives
BigDataOD1	The field of Organizational Development (O.D), who study organizations to create organizational design and strategy are in the best position to lead new innovations
BigData_md_Industry	Use of Big Data is more prevalent in certain industries than others
BigDataVolume_negative	High volume of data alone does not qualify engaging in Big Data initiatives
BIgDataContribute_IoT	Devices, sensors, and various electronics that connect to internet have all contributed to the popularity of Big Data
BigDataTechPlatform	Having the technical infrastructure/platform to support Big Data implementations is a key enabler for success
BigDataMethodology	Organizations that follow a formal methodology (e.g: Agile/DevOps/DataOps/custom) are more likely to succeed at Big Data implementations
BigDataOD2	I believe Organizational change experts, those of whom are concerned with organizational effectiveness overall, should lead Big Data implementations across the organizations
BigDataInformationIntensity	I believe that an organization's focus on information (data) sharing is an indicator of success in Big Data implementations
BigDataCollaboration	I believe Big Data implementation success if more dependent on cross functional collaboration than just Information Technology implementation.
BigDataCollaboration2	Organizations where business areas are involved lead to more successful Big Data initiatives

BigDataSizeOrg	Size of the organization does not have anything to do with Big Data Implementations
BigDataOD3	Big Data initiatives are undertaken for the betterment of the organization
BigDataMethodologyChange	Failure to change is usually unrelated to the actual technical implementation.
BigDataCollaboration3	Leadership and employees engaged together lead to success in implementations of ambiguous or new initiatives.
BigDataInformation2	Data owners and stewards with clear established roles allow for a more successful Big Data Implementation
BigDataInformation3	Organizations who have a information governance policy allow for a better Big Data implementation and strategy
BigDataMethodology3	To stay competitive businesses must continuously change and adapt to new innovations, such as Big Data.
BigDataChange	Big Data can help enable change
Dependent Variable	Big Data can help further business strategy
BigDataInformation4	It is important that organizations form an overarching data strategy
BigDataCollaboration4	Big Data implementation is a business challenge and not a technical one
BigDataHelp_Expertise	My organization has hired consultants to help with Big Data initiatives
BigDataOD4	Understanding of behavioral sciences leads to success in Big Data implementation.

BigDataInformation5	There is an increase in Data driven projects in my organization
BigDataOrgSize2	Size of organization successfully impacts Big Data Implementation
BigDataTechPlatform2	Technology intensive organizations are more successful at implementing Big Data Initiatives
BigDataPrivacy_New	Privacy is a cause of great concern in Big Data initiatives
Dependent Variable	Successful Big Data Implementations can result in the offering of new products or services
BigDataChar_Volume	Volume of data is a characteristic of Big Data
BigDataChar_Velocity	Velocity of data is a characteristic of Big Data
BigDataChar_Variety	Variety of data is a characteristic of Big Data
BigDataChar_Value	Value of data is a characteristic of Big Data
BigDataChar_Veracity	Veracity (which represents unreliability or uncertainty) of data is a characteristic of Big Data
BigDataChar_Variability	Variability (which represents inconsistency) of data is a characteristic of Big Data
BigDataConsensus_Bench001	There are no universal benchmarks available to define Big Data
BigDataConsensus_Bench002	Characteristics of what constitutes Big Data will evolve

Dependent Variable	Successful Big Data Implementations will result in new processes developed
Dependent Variable	Successful Big Data Implementations will result in new patents or source of competitive advantage for the organization
BigDataTechplatform003	Technical skills are necessary for successful Big Data Implementations
BigDataFunding	Funding is key to successful Big Data implementations

Part 4: Forming Hypotheses

In the above three parts, we have reviewed the raw data to look for criteria for successful Big Data Implementations, reviewed what impacts said criteria and deduced a testable model for further statistical testing. This research study started with a simple question around successful Big Data Implementations. Armed with the knowledge gained from our first part of our mix methods study the researcher makes the following hypotheses to be tested:

H1: Given Big Data characteristics, Organizational Development characteristics positively impact successful Big Data Implementations.

H2: Given Big Data characteristics, technical competence positively impacts successful Big Data Implementations.

H3: Given Big Data characteristics, addressing privacy concerns positively impacts successful Big Data Implementations.

H4: Given Big Data characteristics, cross functional collaboration positively impacts successful Big Data implementations.

H5: Given Big Data characteristics, understanding of Behavioral knowledge positively impacts successful Big Data Implementations

H6: Given Big Data characteristics, organizational change management expertise positively impact successful Big Data Implementations.

H7: Given Big Data characteristics, established Data ownership roles positively impacts successful Big Data Implementations.

H8: Given Big Data characteristics, type of industry does not significantly impact successful Big Data Implementation.

H9: Given Big Data characteristics, age does not significantly impact successful Big Data Implementations.

H10: Given Big Data characteristics, profession does not significantly impact successful Big Data Implementation.

H11: Given Big Data characteristics, geographic region does not significantly impact successful Big Data Implementations.

H12: Given Big Data characteristics, technical infrastructure positively impacts successful Big Data Implementations.

H13: Given Big Data characteristics, technology intensive organizations positively impact successful Big Data Implementations.

H14: Given Big Data characteristics, organizations with an information governance policy positively impacts Big Data Implementations.

H15: Big Data Characteristics (volume, variety, velocity, veracity, variability, value and computing resources) positively impact successful Big Data Implementations.

H16: Leadership and employee engagement positively impact successful Big Data Implementations.

H17: Given Big Data characteristics, organizations with a data strategy positively impacts successful Big Data Implementations.

H18: Given Big Data characteristics, business area involvement positively impacts successful Big Data Implementations.

H19: Given Big Data characteristics, funding positively impacts successful Big Data Implementations.

H20: Given Big Data characteristics, experience in working with many data driven projects positively impacts successful Big Data Implementations.

Quantitative Examination of Research Question

The study had a total of 170 responses out of which 95 were of those who had either not implemented Big Data projects or had not heard of Big Data. Those were removed leaving us with a total of 75 responses. There were a few responses where respondents had skipped a few questions. For those with no responses, they were eliminated from the tests being run. See the table below for descriptive statistics. The standard deviation is very much within the range of the responses. The first 6 rows are displayed to show the total number of responses and the other rows display the actual completed responses where conditional logic for the survey (understanding of Big Data and Implementation of Big Data) were met.

Table 10. Quantitative Review – Descriptive Statistics

Descriptive Statistics

	N Statistic	Min Statistic	Max Statistic	Mean Statistic	Std. Dev Statistic
I understand what Big Data is?	170	1.00	2.00	1.2706	0.44558
I work in a company where Big Data is/has been implemented	170	1.00	2.00	1.5176	0.50116
Which functional area/business unit do you represent	170	1.00	9.00	5.1588	2.67214
Which region are you based in?	165	1.00	7.00	3.6485	2.34990
Which Industry do you work in?	170	1	14	8.56	4.193
Please state your profession level	170	0.00	4.00	2.4647	1.21209
Use of Big Data is more prevalent in certain industries than others	75	1.00	5.00	3.7600	0.92765
Big Data can help enable change	75	1.00	5.00	3.8667	1.04407
Value of data is a characteristic of Big Data	74	1.00	5.00	3.9054	1.07485
Variability (which represents inconsistency) of data is a characteristic of Big Data	74	1.00	5.00	3.3784	1.11899
Variety of data is a characteristic of Big Data	73	1.00	5.00	3.9178	0.89370
Velocity of data is a characteristic of Big Data	74	1.00	5.00	3.7297	0.89592
Veracity (which represents unreliability	74	1.00	5.00	3.5676	0.99424

or uncertainty) of data is a characteristic of Big Data					
Volume of data is a characteristic of Big Data	74	1.00	5.00	3.8108	0.94616
I believe Big Data implementation success if more dependent on cross functional collaboration than just Information Technology implementation.	75	1.00	5.00	4.0667	0.90544
Organizations where business areas are involved lead to more successful Big Data initiatives	74	2.00	5.00	3.9730	0.82716
Leadership and employees engaged together lead to success in implementations of ambiguous or new initiatives.	75	1.00	5.00	3.9200	1.02351
Big Data implementation is a business challenge and not a technical one	74	1.00	5.00	3.1757	1.07726
Use of computing resources is an absolute necessity when engaging in Big Data initiatives	75	1.00	5.00	3.8533	0.98218
There are no universal benchmarks available to define Big Data	74	1.00	5.00	3.5811	0.99322
Characteristics of what constitutes Big Data will evolve	73	2.00	5.00	3.7945	0.83265
Devices, sensors, and various electronics that connect to internet have all contributed to the popularity of Big Data	74	1.00	5.00	3.9459	0.97772
Funding is key to successful Big Data implementations	74	2.00	5.00	3.6892	0.89022
Big Data is a concept that is currently misunderstood?	75	1.00	5.00	3.4400	0.94783
My organization has hired consultants to help with Big Data initiatives	74	1.00	5.00	3.2838	1.05363
Successful Big Data Implementations can result in the offering of new products and services	74	1.00	5.00	3.9189	1.05670
Successful Big Data Implementations will result in new processes developed	74	1.00	5.00	3.8784	0.90588
Successful Big Data Implementations will result in new patents or source of competitive advantage for the organization	73	1.00	5.00	3.9315	0.96218
Data owners and stewards with clear established roles allow for a more successful Big Data Implementation	74	1.00	5.00	3.9054	0.96754
Organizations who have a information governance policy allow for a better Big Data implementation and strategy	75	1.00	5.00	3.7067	0.96944
It is important that organizations form an overarching data strategy	75	1.00	5.00	3.9333	0.96329

There is an increase in Data driven projects in my organization	74	1.00	5.00	3.7703	0.91483
I believe that an organization's focus on information (data) sharing is an indicator of success in Big Data implementations	75	1.00	5.00	3.4400	1.09347
Organizations that follow a formal methodology (e.g: Agile/DevOps/DataOps/custom) are more likely to succeed at Big Data implementations	74	1.00	5.00	3.8378	0.95124
To stay competitive businesses must continuously change and adapt to new innovations, such as Big Data.	74	1.00	5.00	4.0135	1.02694
Failure to change is usually unrelated to the actual technical implementation.	73	2.00	5.00	3.5616	0.78149
The field of Organizational Development (O.D), who study organizations to create organizational design and strategy are in the best position to lead new innovations	75	1.00	5.00	3.5067	1.17833
I believe Organizational change experts, those of whom are concerned with organizational effectiveness overall, should lead Big Data implementations across the organizations	75	1.00	5.00	3.5867	1.15189
Big Data initiatives are undertaken for the betterment of the organization	75	1.00	5.00	3.6533	0.99313
Understanding of behavioral sciences leads to success in Big Data implementation.	73	1.00	5.00	3.3151	1.06558
Size of organization successfully impacts Big Data Implementation	74	1.00	5.00	3.2297	1.09228
Privacy is a cause of great concern in Big Data initiatives	74	1.00	5.00	3.7027	1.16724
Size of the organization does not have anything to do with Big Data Implementations	74	1.00	5.00	3.4324	1.09895
Big Data can help further business strategy	75	1.00	5.00	4.1067	0.79820
Technical skills are necessary for successful Big Data Implementations	73	1.00	5.00	3.9315	0.93287
Having the technical infrastructure/platform to support Big Data implementations is a key enabler for success	75	1.00	5.00	3.9600	0.82920
Technology intensive organizations are more successful at implementing Big Data Initiatives	74	1.00	5.00	3.5676	0.92279
High volume of data alone does not qualify engaging in Big Data initiatives	75	1.00	5.00	3.7200	0.99404

KMO & Bartlett's Test

As Tabachnick and Fidell (2001) cite Comrey and Lee's (1992) advice, as a rule of thumb, a bare minimum of 10 observations per variable is necessary to avoid computational difficulties. In our case, we exceed that observation but given the overall response rate, it was important to conduct a Kaiser-Meyer-Olkin (KMO) test of sampling adequacy. This was done to ensure that the constructs exceeded a value of .5 or higher, indicating that the data was suitable for further testing (Kaiser, 1974). Furthermore, Bartlett's test of Sphericity test to see if variables are unrelated and therefore unsuitable for structure detection. A significance level of less than 0.05 would indicate that a factor analysis would be useful. The researcher ran 2 tests in total for KMO and Bartlett.

1. First, the full data set of all the variables with completed responses with an acceptable 0.657 at a significance level of 0.000 alerting that data is useful for factor analysis.
2. Second, the full data set of all variables except this time removing the questions asking about the characteristics of Big Data (volume, variety, etc) and came up with a higher rating of 0.722 at a significance of 0.000 alerting that data, again, is useful for factor analysis. As shown below in Table 11.

Table 11. Quantitative Review – KMO and Bartlett's Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0.722

Bartlett's Test of Sphericity	Approx. Chi-Square	1477.328
	df	666
	Sig.	0.000

As this is a new area of academic research with no prior survey implemented or empirical data present the researcher wanted to minimize assumptions and it was important to see if a structure did exist with the data collected and if further analysis would be beneficial. Thus, the reason for running two tests to see how concentration of variables affected the result. The results indicated moving forward with factor analysis and the researcher ran communalities, total variance and rotated component matrix to further examine the data.

The reasons behind communalities is to estimate variance in each variable accounted by all the available factors/components. The initial is always 1, which is the minimum eigen value in terms of variance explained by a factor. The extraction communalities are estimates the variance in each variable is accounted for by the components. As you can see in the table below, the extracted components are high with the lowest of 0.598. This means that components represent the variables well.

Table 12. Quantitative Review – Communalities

Communalities

	Initial	Extraction
	Initial	**Extraction**
I believe Organizational change experts, those of whom are concerned with organizational effectiveness overall, should lead Big Data implementations across the organizations	1	0.742

Big Data initiatives are undertaken for the betterment of the organization	1	0.83
Understanding of behavioral sciences leads to success in Big Data implementation.	1	0.724
Having the technical infrastructure/platform to support Big Data implementations is a key enabler for success	1	0.598
Technology intensive organizations are more successful at implementing Big Data Initiatives	1	0.792
Technical skills are necessary for successful Big Data Implementations	1	0.766
Organizations that follow a formal methodology (e.g: Agile/DevOps/DataOps/custom) are more likely to succeed at Big Data implementations	1	0.709
I believe that an organization's focus on information (data) sharing is an indicator of success in Big Data implementations	1	0.709
I believe Big Data implementation success if more dependent on cross functional collaboration than just Information Technology implementation.	1	0.79
Organizations where business areas are involved lead to more successful Big Data initiatives	1	0.813
Leadership and employees engaged together lead to success in implementations of ambiguous or new initiatives.	1	0.721
Big Data implementation is a business challenge and not a technical one	1	0.751
Failure to change is usually unrelated to the actual technical implementation.	1	0.83
Big Data can help enable change	1	0.642

Data owners and stewards with clear established roles allow for a more successful Big Data Implementation	1	0.645
Organizations who have a information governance policy allow for a better Big Data implementation and strategy	1	0.783
It is important that organizations form an overarching data strategy	1	0.805
There is an increase in Data driven projects in my organization	1	0.653
To stay competitive businesses must continuously change and adapt to new innovations, such as Big Data.	1	0.666
Use of Big Data is more prevalent in certain industries than others	1	0.587
Devices, sensors, and various electronics that connect to internet have all contributed to the popularity of Big Data	1	0.698
My organization has hired consultants to help with Big Data initiatives	1	0.764
Size of the organization does not have anything to do with Big Data Implementations	1	0.719
Size of organization successfully impacts Big Data Implementation	1	0.78
Privacy is a cause of great concern in Big Data initiatives	1	0.836
Funding is key to successful Big Data implementations	1	0.715
Use of computing resources is an absolute necessity when engaging in Big Data initiatives	1	0.668
High volume of data alone does not qualify engaging in Big Data initiatives	1	0.846
Volume of data is a characteristic of Big Data	1	0.783
Velocity of data is a characteristic of Big Data	1	0.799
Variety of data is a characteristic of Big Data	1	0.734
Value of data is a characteristic of Big Data	1	0.788

Veracity (which represents unreliability or uncertainty) of data is a characteristic of Big Data	1	0.85
Variability (which represents inconsistency) of data is a characteristic of Big Data	1	0.839
There are no universal benchmarks available to define Big Data	1	0.758
Characteristics of what constitutes Big Data will evolve	1	0.599
Big Data is a concept that is currently misunderstood?	1	0.822
I believe Organizational change experts, those of whom are concerned with organizational effectiveness overall, should lead Big Data implementations across the organizations	1	0.742
Big Data initiatives are undertaken for the betterment of the organization	1	0.83
Understanding of behavioral sciences leads to success in Big Data implementation.	1	0.724
Having the technical infrastructure/platform to support Big Data implementations is a key enabler for success	1	0.598

Next, we look at the total variance in the variables by looking at initial, extracted and rotated components. As we can see in the table below, there are a total of 37 components. The "Initial" values show the full variance across the components. For instance, Component 1 explains 31% of the variance. The "Extraction" values show that with a loss of ~25% the first eleven components can explain ~75% of the variability. This means we can reduce our data set by from 37 individual components to a combined 11 components and only lose ~25% of the information. Finally, the "Rotation" values show a large change from the extraction section and that variation is much more evenly spread across the 11 components. Simply stated, all this means is that it will be easier to comprehend the rotated component matrix. The chart that follows the table shows the same picture visually.

Table 13. Quantitative Review – Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.504	31.093	31.093	11.504	31.093	31.093	5.407	14.612	14.612
2	2.867	7.747	38.84	2.867	7.747	38.84	2.803	7.575	22.187
3	2.133	5.765	44.605	2.133	5.765	44.605	2.786	7.529	29.716
4	1.834	4.956	49.56	1.834	4.956	49.56	2.5	6.757	36.472
5	1.626	4.394	53.954	1.626	4.394	53.954	2.265	6.121	42.593
6	1.534	4.147	58.101	1.534	4.147	58.101	2.24	6.054	48.647
7	1.422	3.843	61.944	1.422	3.843	61.944	2.132	5.762	54.409
8	1.377	3.723	65.666	1.377	3.723	65.666	1.943	5.253	59.662
9	1.179	3.188	68.854	1.179	3.188	68.854	1.915	5.175	64.837
10	1.071	2.894	71.748	1.071	2.894	71.748	1.789	4.835	69.672
11	1.007	2.722	74.47	1.007	2.722	74.47	1.775	4.798	74.47
12	0.954	2.578	77.048						
13	0.837	2.263	79.312						
14	0.761	2.058	81.369						
15	0.747	2.02	83.389						
16	0.672	1.816	85.204						
17	0.569	1.538	86.743						
18	0.552	1.492	88.234						
19	0.488	1.318	89.552						
20	0.465	1.256	90.808						
21	0.429	1.159	91.967						
22	0.381	1.029	92.995						
23	0.369	0.998	93.993						
24	0.339	0.916	94.909						
25	0.278	0.75	95.659						
26	0.228	0.617	96.276						
27	0.213	0.577	96.853						
28	0.203	0.549	97.401						
29	0.17	0.46	97.861						
30	0.164	0.443	98.305						
31	0.155	0.418	98.723						
32	0.117	0.315	99.038						
33	0.101	0.274	99.312						
34	0.08	0.217	99.529						
35	0.07	0.189	99.718						
36	0.064	0.174	99.892						
37	0.04	0.108	100						

Next, we look at the rotated component matrix, which helps to determine what the 11 components are comprised of. The reason this is helpful is this will help us focus on other analyses (reliability, correlation, and regression) on the variables that really create an impact. DeStefano et. al (2009) talk about factor and component scoring. They go on to mention that an easy way to consider an item's relationship to the factor when creating a factor score is to include only items with loading values above a cut-off value in the computations. By doing so, researchers are only using "marker" variables in the computation. However, the cut-off value to use is an arbitrary decision. As such, in our table below we consider an item loaded if it is above 0.4. You will see in the table below that all items are loaded in combinations relating to a component except for the variable "size of the organization does not have anything to do with Big Data Implementations". Other items such as, "Failure to change is unrelated to the actual technical implementation" is highly correlated to component 11. It is less correlated across the other components. What this tells us is that we can focus on the identified related for further analysis. All items contribute to most of the findings. Furthermore, we investigate the variables that form each of the 11 components.

Table 14. Quantitative Review – Rotated Component Matrix

	Component										
	1	2	3	4	5	6	7	8	9	10	11
I believe Organizational change experts, those of whom are concerned with organizational effectiveness overall, should lead Big Data implementations across the organizations	0.179	0.119	0.326	0.592	-0.232	0.085	0.226	0.172	0.231	0.195	0.075
Big Data initiatives are undertaken for the betterment of the organization	0.223	0.103	0.108	0.097	0.075	0.579	-0.057	0.104	0.568	-0.066	0.256
Understanding of behavioral sciences leads to success in Big Data implementation.	0.257	0.187	-0.02	0.509	0.223	0.092	0.061	0.042	0.238	0.326	0.37
Having the technical infrastructure/platform to support Big Data implementations is a key enabler for success	0.457	0.192	0.457	0.118	0.211	0.222	0.065	0.082	0.108	-0.059	-0.096
Technology intensive organizations are more successful at implementing Big Data Initiatives	0.463	-0.193	0.273	0.43	-0.029	0.187	0.057	0.012	-0.309	0.057	0.379
Technical skills are necessary for successful Big Data Implementations	0.209	0.77	0.208	0.043	0.204	0.156	0.08	0.078	0.051	0.043	0.042
Organizations that follow a formal methodology (e.g: Agile/DevOps/DataOps/custom) are more likely to succeed at Big Data implementations	0.752	0.034	0.224	0.082	-0.098	-0.123	-0.067	0.039	0.133	-0.156	0.11
I believe that an organization's focus on information (data) sharing is an indicator of success in Big Data implementations	0.334	0.576	0.178	0.301	-0.032	0.121	-0.027	0.075	0.309	0.16	-0.014
I believe Big Data implementation success if more dependent on cross functional collaboration than just Information Technology implementation.	0.166	0.309	0.134	0.079	0.039	0.176	0.247	0.633	0.036	0.366	0.114
Organizations where business areas are involved lead to more successful Big Data initiatives	0.523	0.25	0.124	0.086	0.069	0.099	0.049	0.57	0.278	0.084	0.168
Leadership and employees engaged together lead to success in implementations of ambiguous or new initiatives.	0.74	0.254	-0.006	0.047	0.063	0.071	0.164	0.001	0.22	0.034	0.149
Big Data implementation is a business challenge and not a technical one	0.206	0.055	-0.071	0.316	-0.006	-0.106	0.573	-0.017	0.217	0.233	-0.399
Failure to change is usually unrelated to the actual technical implementation.	0.164	0.127	-0.008	0.07	0.142	0.059	0.162	0.055	0.107	-0.005	0.847
Big Data can help enable change	0.152	0.179	0.045	0.169	0.031	0.729	-0.046	0.129	-0.038	0.061	-0.01
Data owners and stewards with clear established roles allow for a more successful Big Data Implementation	0.528	0.177	0.147	-0.138	0.216	0.331	0.181	0.174	-0.019	0.215	0.17
Organizations who have a information governance policy allow for a better Big Data implementation and strategy	0.274	0.748	0.088	0.191	0.092	0.128	0.138	0.162	-0.01	0.162	0.091
It is important that organizations form an overarching data strategy	0.528	0.358	0.176	0.251	0.22	0.242	-0.049	0.387	0.139	-0.024	0.157
There is an increase in Data driven projects in my organization	0.677	0.13	-0.025	0.132	0.101	0.275	-0.02	0.111	-0.045	0.244	0.012
To stay competitive businesses must continuously change and adapt to new innovations, such as Big Data.	0.532	0.425	0.142	-0.101	0.097	0.269	0.035	-0.236	0.053	0.053	0.168
Use of Big Data is more prevalent in certain industries than others	0.277	0.134	0.114	0.53	0.281	0.197	-0.015	-0.225	0.136	-0.094	0.043
Devices, sensors, and various electronics that connect to internet have all contributed to the popularity of Big Data	0.617	0.218	0.246	0.119	0.155	0.098	0.189	0.026	-0.216	0.274	-0.06
My organization has hired consultants to help with Big Data initiatives	0.065	0.177	0.1	0.2	0.055	0.09	-0.129	0.051	0.174	0.786	-0.009
Size of the organization does not have anything to do with Big Data Implementations	0.302	0.107	0.056	-0.239	-0.292	0.019	0.322	-0.557	0.153	0.051	0.175
Size of organization successfully impacts Big Data Implementation	-0.129	0.128	0.054	0.815	0.006	0.069	0.059	0.226	0.002	0.129	-0.053
Privacy is a cause of great concern in Big Data initiatives	0.378	0.326	0.545	-0.004	0.017	-0.102	-0.139	0.299	0.012	0.24	0.338
Funding is key to successful Big Data implementations	0.154	0.111	0.148	0.221	0.252	0.014	0.116	0.041	0.678	0.263	0.01
Use of computing resources is an absolute necessity when engaging in Big Data initiatives	0.436	0.074	0.074	0.099	0.03	0.372	0.059	0.499	0.166	-0.16	0.109
High volume of data alone does not qualify engaging in Big Data initiatives	0.321	-0.187	0.226	-0.289	-0.282	-0.108	0.396	0.119	0.425	0.088	0.351
Volume of data is a characteristic of Big Data	0.333	-0.036	0.767	0.091	-0.047	0.131	0.023	-0.062	0.184	0.096	0.088
Velocity of data is a characteristic of Big Data	0.029	0.254	0.829	0.106	0.09	0.009	0.096	0.115	0.008	-0.033	-0.057
Variety of data is a characteristic of Big Data	0.132	0.182	0.501	0.074	0.212	0.551	0.172	-0.106	0.038	0.191	0.025
Value of data is a characteristic of Big Data	0.34	0.001	-0.033	0.004	0.486	0.481	0.014	0.165	0.046	0.416	0.034
Veracity (which represents unreliability or uncertainty) of data is a characteristic of Big Data	0.046	0.094	0.056	-0.012	0.888	0.139	-0.057	0.101	0.025	-0.018	0.115
Variability (which represents inconsistency) of data is a characteristic of Big Data	0.065	0.329	0.147	0.135	0.66	-0.06	0.105	0.03	0.41	0.253	0.063
There are no universal benchmarks available to define Big Data	0.313	0.178	-0.067	0.198	0.175	-0.013	0.713	-0.083	0.092	0.005	0.173
Characteristics of what constitutes Big Data will evolve	0.64	0.151	0.207	-0.038	-0.048	0.152	0.218	0.037	0.199	0.095	0.004
Big Data is a concept that is currently misunderstood?	-0.109	0.02	0.271	-0.062	-0.14	0.075	0.768	0.124	-0.084	-0.269	0.148

Components:

The table below lists what we find as variables in component 1 that contributes a total of ~31% of the variance. The data below suggests that component 1 is very much representative of organizational factors. Formal methodology, leadership and employee engagement, to stay competitive in business, data strategy, involvement of business areas, and the like are all organizational factors. These are all important realizations needed by an organization to embark on Big Data implementations. We will label component 1 as "organizational contributors".

Table 15. Quantitative Review – Component 1 of 11

	Component 1 – Organizational Contributors	
1	Organizations that follow a formal methodology (e.g: Agile/DevOps/DataOps/custom) are more likely to succeed at Big Data implementations	0.752

2	Leadership and employees engaged together lead to success in implementations of ambiguous or new initiatives.	0.740
3	There is an increase in Data driven projects in my organization	0.677
4	Characteristics of what constitutes Big Data will evolve	0.640
5	Devices, sensors, and various electronics that connect to internet have all contributed to the popularity of Big Data	0.617
6	To stay competitive businesses must continuously change and adapt to new innovations, such as Big Data.	0.532
7	Data owners and stewards with clear established roles allow for a more successful Big Data Implementation	0.528
8	It is important that organizations form an overarching data strategy	0.528
9	Organizations where business areas are involved lead to more successful Big Data initiatives	0.523
10	Technology intensive organizations are more successful at implementing Big Data Initiatives	0.463
11	Having the technical infrastructure/platform to support Big Data implementations is a key enabler for success	0.457
12	Use of computing resources is an absolute necessity when engaging in Big Data initiatives	0.436

These variables that make up component 2 represent a set of knowledge, skills and abilities that will calibrate successful Big Data Implementations at the organization. The focal points being around information governance policy, technical skills, focus on information sharing and continuous evolution. They all represent developmental competencies, which makes up knowledge, skills and abilities of an organization for successful Big Data Implementations. We will label our component as, "Organizational Development Competencies".

Table 16. Quantitative Review – Component 2 of 11

	Component 2 – Organizational Development Competencies	
1	Technical skills are necessary for successful Big Data	0.770

	Implementations	
2	Organizations who have an information governance policy allow for a better Big Data implementation and strategy	0.748
3	I believe that an organization's focus on information (data) sharing is an indicator of success in Big Data implementations	0.576
4	To stay competitive businesses must continuously change and adapt to new innovations, such as Big Data.	0.425

Inspection of the variables that make up the 3rd component reveal characteristics that made up the initial perception of Big Data. Factors such as, volume, velocity and variety of data, privacy being a concern, and having a technical platform were all characteristics that made up initial perceptions of Big Data. Halaweh and Massry (2015) quote these five that represent their understanding of Big Data.

Table 17. Quantitative Review – Component 3 of 11

	Component 3 – Initial Big Data Focus	
1	Velocity of data is a characteristic of Big Data	0.829
2	Volume of data is a characteristic of Big Data	0.767
3	Privacy is a cause of great concern in Big Data initiatives	0.545
4	Variety of data is a characteristic of Big Data	0.501
5	Having the technical infrastructure/platform to support Big Data implementations is a key enabler for success	0.457

The variables that make up this component are focused on Organizational Development Characteristics. Variables such as size of organization, organizational change experts, understanding of behavioral knowledge, and technology intensive organizations are all characteristics representing Organizational Development.

Change and behavioral knowledge are core concepts from OD. Size, technology intensive and more prevalence in certain industries represents characteristics of Organizations.

Table 18. Quantitative Review – Component 4 of 11

	Component 4 – Organizational Development Characteristics	
1	Size of organization successfully impacts Big Data Implementation	0.815
2	I believe Organizational change experts, those of whom are concerned with organizational effectiveness overall, should lead Big Data implementations across the organizations	0.592
3	Use of Big Data is more prevalent in certain industries than others	0.530
4	Understanding of behavioral sciences leads to success in Big Data implementation.	0.509
5	Technology intensive organizations are more successful at implementing Big Data Initiatives	0.430

The 5th component focuses on the new characteristics identified of Big Data. Characteristics that we found in the literature as well quoted by Wamba et. al (2015). The variables represented include veracity, variability and value. As such, we call this component, "New Big Data Characteristics".

Table 19. Quantitative Review – Component 5 of 11

	Component 5 – New Big Data Characteristics	
1	Veracity (which represents unreliability or uncertainty) of data is a characteristic of Big Data	0.888
2	Variability (which represents inconsistency) of data is a characteristic of Big Data	0.660
3	Value of data is a characteristic of Big Data	0.486

This 6th component represents change. Variables such as enabling change and undertaking initiatives for betterment of organization both point to organizational change management. Furthermore, variety and value of change can also be attributed to change management from the standpoint of dealing with constant variety of data and using that to derive value for the organization. Therefore, we name our component, “Organizational Change Management”.

Table 20. Quantitative Review – Component 6 of 11

	Component 6 – Organizational Change Management	
1	Big Data can help enable change	0.729
2	Big Data initiatives are undertaken for the betterment of the organization	0.579
3	Variety of data is a characteristic of Big Data	0.551
4	Value of data is a characteristic of Big Data	0.481

Components 7 had an unusual setup of variables. It looked at affirming that Big Data is a concept that is misunderstood, there are no universal benchmarks available for Big Data and it is more of a business challenge and not a technical one. In other words, it was challenging what we perceived as our initial understanding of Big Data where we used specific characteristics (as in component 3) to benchmark Big Data initiatives. Here we were challenging the notion of what Big Data means, acknowledging no universality, and its players for formation.

Table 21. Quantitative Review – Component 7 of 11

	Component 7 – Challenge Status Quo	
1	Big Data is a concept that is currently misunderstood	0.768
2	There are no universal benchmarks available to define Big Data	0.713
3	Big Data implementation is a business challenge and not a technical one	0.573

Two out of the three component 8 variables focus on collaboration between teams and business areas. The last component identifies a need of using computing resources for engaging in Big Data Implementations. It was identified as such.

Table 22. Quantitative Review – Component 8 of 11

	Component 8 – Collaboration and Advanced Computing	
1	I believe Big Data implementation success if more dependent on cross functional collaboration than just Information Technology implementation.	0.633
2	Organizations where business areas are involved lead to more successful Big Data initiatives	0.57
3	Use of computing resources is an absolute necessity when engaging in Big Data initiatives	0.499

Component 9 variables look at key enablers for successful Big Data implementations. Funding, initiatives taken for betterment for the organization and not just relying on high volume of data alone are all key enablers for successful Big Data Implementations.

Table 23. Quantitative Review – Component 9 of 11

	Component 9 – Key enablers	

1	Funding is key to successful Big Data implementations	0.678
2	Big Data initiatives are undertaken for the betterment of the organization	0.568
3	High volume of data alone does not qualify engaging in Big Data initiatives	0.425
4	Variability (which represents inconsistency) of data is a characteristic of Big Data	0.41

Component 10 has two different variables. One is a little stronger than the other one but they both contribute to this component. As such, getting outside help, in form of consultants, is looking for guidance and value is something that organizations are looking to derive from data they have. As such, this component is named separately with both constructs, Guidance and Value.

Table 24. Quantitative Review – Component 10 of 11

	Component 10 – Guidance and Value	
1	My organization has hired consultants to help with Big Data initiatives	0.786
2	Value of data is a characteristic of Big Data	0.416

Component 11 has only one strongly related variable. It deals with calling out that a failure to change is usually not related to the actual technical implementation but other organizational factors. As such, the component was named Organizational Development change competency to highlight that change competence at an organization needs to be focal within Big Data Implementations.

Table 25. Quantitative Review – Component 11 of 11

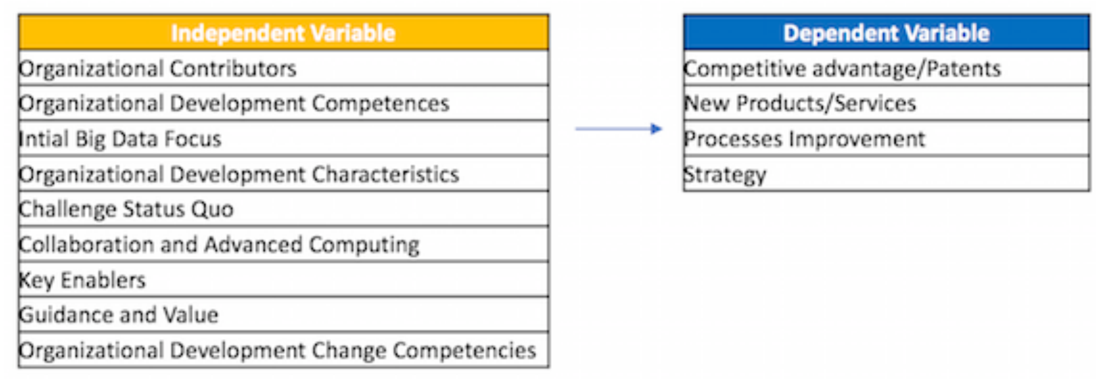
	Component 11 – Organizational Development Change Competency	

1	Failure to change is usually unrelated to the actual technical implementation.	0.847

In summary, we can conclude that our data is suitable for Factor Analysis using KMO and Bartlett's tests. We can also conclude after running communalities that extracted components represent the variables well. Furthermore, we can reduce our data file comprising of a total of 37 variables to be represented within 11 components by Factor Analysis using principal components extraction with a loss of ~25% of data. Further analyses revealed that there was not a single variable that represented a component strongly. As interviews have revealed, successful Big Data Implementations are a combination of factors as described by multiple interviewees that lead to the deductive model for quantitative study. The factor analysis conducted provided further evidence with the various combination of variables representing the individual components. and it was evident within the variables representing the 11 components. The only variable that did not feature in the 11 components was "size of the organization does not successfully impact Big Data Implementation". At the end of this section, we are now able to continue forward to investigate correlation and regression relationships for our model with all the variables well represented.

At the end of this section, we can, therefore, update our model to reflect what our factor analysis has provided to us. We update the independent variables with the components listed above. The new model used for further testing is as below:

Figure 8: Updated model after Factor Analysis



Given the updated model the hypotheses had to be updated to reflect the relationships from the model in figure 8.

Below are the 7 hypotheses that this research will test:

H1: Opinions of Organizational Change Management will be positively associated with higher perceptions of successful Big Data implementations as generator of new products and services.

H2: Opinions of Collaboration and advanced computing will be positively associated with higher perceptions of successful Big Data implementations as generator of new products and services.

H3: Opinions of organizational development competencies will be positively associated with higher perceptions of successful Big Data implementations as developer of new processes.

H4: Opinions of Collaboration and advanced computing will be positively associated with higher perceptions of successful Big Data implementations as developer of new processes

H5: Opinions of initial Big Data characteristics will be positively associated with higher perceptions of successful Big Data Implementations as generator of new patents and source of competitive advantage

H6: Opinions of Collaboration and advanced computing will be positively associated with higher perceptions of successful Big Data Implementations as furthering of business strategy

H7: Opinions of challenging status quo will be positively associated with higher perceptions of furthering business strategy as part of successful Big Data implementations

Correlation

Correlation is a statistical technique to test the strength of relationship between pairs of variables. The correlation coefficient ("r") ranges from -1 to +1. If the coefficient is close to 0, it signifies no relationship, positive means as one variable gets larger so does the other and negative means as one variable gets larger the other gets smaller.

Running correlations against all the variables we see a positive relationship at the 0.01 and 0.05 significance level for many of the variables. The table below shows the correlation between the independent variables and the dependent variables. The scale used to look at relationships was > +/- 0.5 to 1.0 is a strong relationship, +/- 0.25 to 0.499 is a moderate relationship and anything below that is a small correlation. The yellow highlights show the variables that are not correlated at all. All others are correlated at moderation or strong correlations.

Looking at the table we can quickly see that our major themes are carried across the 4 dependent variables: (a) Majority of the Big Data characteristics and (b) majority of the Organizational Development characteristics (collaboration, change management, and competencies that include technical skills and abilities) are all strongly correlated (at 0.01 significance level) to successful Big Data Implementations. The dual asterisks (**) is a sign of significance at 0.01 confidence level while the single asterisk (*) is a sign of significance at 0.05 confidence level.

Table 26. Quantitative Review – Correlations to Dependent Variables

	Big Data can help further business strategy	Successful Big Data Implementations can result in the offering of new products and services	Successful Big Data Implementations will result in new processes developed	Successful Big Data Implementations will result in new patents or source of competitive advantage for the organization
Factor 1 = Organizational Contributors	.690**	.812**	.611**	.694**
Factor 2 = Organizational Development Competencies	.569**	.677**	.664**	.649**
Factor 3 = Initial Big Data Focus	.621**	.698**	.493**	.691**
Factor 4 = Organizational Development Characteristics	.551**	.640**	.480**	.540**
Factor 5 = New Big Data Characteristics	.523**	.504**	.480**	.507**
Factor 6 = Organizational Change Management	.638**	.688**	.502**	.547**
Factor 7 = Challenge Status Quo	0.129	.298**	.378**	.404**
Factor 8 = Collaboration and Advanced Computing	.659**	.768**	.604**	.545**
Factor 9 = Key enablers	.414**	.495**	.440**	.505**
Factor 10 = Guidance and Value	.515**	.548**	.461**	.481**
Factor 11 = Organizational Development	.263*	.260*	.250*	.242*

Change Competencies				

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Reliability Statistics

Reliability is a measure of similar results for similar inputs by an instrument. In our case, given survey variables, we need to measure the reliability of the scale that measures the variables. We will measure the reliability of our scale represented by the 11 components from our factor analyses. Our goal is to identify characteristics that result in successful Big Data Implementation and we used a two-part study to answer that questions. In the first part, we used qualitative analysis to come up with the variables that measure and influence the success of Big Data Implementations. In the second part, we created a new survey from scratch that would allow us to sample the population and provide a way to generalize our findings. To ensure that our measure reliable (scale used for our survey) we run Cronbach's Alpha for reliability. The result is a very strong 0.914.

Table 27. Quantitative Review – Reliability Statistics (Cronbach's Alpha)

Reliability Statistics

Cronbach's Alpha	N of Items
0.914	11

Furthermore, below you will find a table which lists out Cronbach's Alpha for each variable if deleted from the survey. For the listed items below Cronbach's Alpha is > 0.85. While deleting factors 7 and 11 would result in a higher Cronbach's Alpha it was not done for two reasons: (a) the change was a very small number <=0.005 and (b) we would want to see if there was any significance between these factors and successful Big Data Implementation.

Table 28. Quantitative Review – Cronbach's Alpha (if item is deleted)

	Cronbach's Alpha if Item Deleted
Factor 1 = Organizational Contributors	0.899
Factor 2 = Organizational Development Competencies	0.901
Factor 3 = Initial Big Data Focus	0.905
Factor 4 = Organizational Development Characteristics	0.905
Factor 5 = New Big Data Characteristics	0.906
Factor 6 = Organizational Change Management	0.899
Factor 7 = Challenge Status Quo	0.919
Factor 8 = Collaboration and Advanced Computing	0.904
Factor 9 = Key enablers	0.905
Factor 10 = Guidance and Value	0.906
Factor 11 = Organizational Development Change Competencies	0.918

Regression

To test our hypotheses and relationships between our variables to successful Big Data Implementations we run regression tests. Basically, regression is a statistical process that provides us with an estimate of the relationships among variables. We will run regression in two parts: (a) run regression to see if there is any evidence of Multicollinearity, and (b) run regression to test our full model and hypotheses.

Part (a) Regression for Multicollinearity:

In our first run, we run regression for our 11 factors with each dependent variable. There are a total of 4 tables, one for each dependent variable. What we find is that Factor 1 (representing Organizational Contributors) had strong case of multicollinearity. The cutoff used was Tolerance < 0.2 and VIF (variance) > 5. There was no evidence of any other factor where multicollinearity was an issue. See the table below for data for factor 1:

Table 29. Quantitative Review – MultiCollinearity Testing

DV#1: New Products and Services		
	Collinearity Statistics	
	Tolerance	VIF
Factor 1	0.181	5.512

DV#2: New Processes		
	Collinearity Statistics	
	Tolerance	VIF

Factor 1	0.181	5.512

DV#3: New Patents or Source of Competitive Advantage		
	Collinearity Statistics	
	Tolerance	VIF
Factor 1	0.18	5.547

DV#4: Further Business Strategy		
	Collinearity Statistics	
	Tolerance	VIF
Factor 1	0.181	5.512

Next, we look at the regression tables to test strength and relationship of the variables in our model.

Part (b) Full Regression

Running full regression of the model with our controls (industry, age, profession and geography) and each of the dependent variables gives us two tables. As the test for multicollinearity was conducted earlier and as factor 1 was the only one susceptible to it, it has been removed for the remaining regression tests. Below you will find 3 models for each dependent variable: (a) the regression model listing all the controls, (b) regression model listing controls and factor relationships and (c) only significant factors and controls from model 2 along with beta, collinearity and confidence statistics, R square and F value for each.

Table 30. Quantitative Review – Dependent Variable #1 (New products and services) Full regression

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics		Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF	R2	F
1	(Constant)	4.157	0.433		9.611	0.000	3.293	5.021				
	Which Industry do you work in?	0.032	0.026	0.139	1.26	0.212	-0.019	0.083	0.892	1.121		
	Please state your profession level	0.064	0.095	0.072	0.671	0.505	-0.126	0.254	0.935	1.07		
	Which region are you based in?	-0.255	0.051	-0.527	-4.975	0.000	-0.357	-0.153	0.968	1.033		
	What is your age group?	0.065	0.108	0.064	0.603	0.549	-0.151	0.281	0.971	1.03		
											0.282**	6.495**
2	(Constant)	0.245	0.646		0.379	0.706	-1.049	1.538				
	Which Industry do you work in?	0.011	0.019	0.046	0.566	0.574	-0.027	0.048	0.789	1.267		
	Please state your profession level	0.045	0.068	0.051	0.657	0.514	-0.091	0.18	0.877	1.14		
	Which region are you based in?	-0.128	0.043	-0.265	-2.992	0.004	-0.214	-0.042	0.664	1.507		
	What is your age group?	0.029	0.079	0.028	0.367	0.715	-0.13	0.188	0.866	1.154		
	Factor2	0.094	0.143	0.077	0.662	0.511	-0.191	0.38	0.384	2.604		
	Factor3	0.189	0.151	0.138	1.249	0.217	-0.114	0.493	0.428	2.335		
	Factor4	0.262	0.135	0.2	1.945	0.057	-0.008	0.532	0.489	2.043		
	Factor5	-0.047	0.131	-0.043	-0.357	0.723	-0.31	0.216	0.351	2.852		
	Factor6	0.357	0.176	0.276	2.032	0.047	0.005	0.71	0.281	3.555		
	Factor7	-0.1	0.108	-0.079	-0.93	0.356	-0.316	0.115	0.72	1.388		
	Factor8	0.398	0.149	0.307	2.672	0.010	0.1	0.697	0.394	2.539		
	Factor9	-0.102	0.158	-0.072	-0.648	0.520	-0.418	0.213	0.423	2.366		
	Factor10	0.011	0.135	0.01	0.08	0.936	-0.26	0.281	0.338	2.957		
	Factor11	-0.084	0.103	-0.07	-0.819	0.416	-0.291	0.122	0.718	1.393		
											0.709**	9.751**
3	(Constant)	0.28	0.468		0.599	0.551	-0.654	1.215				
	Which region are you based in?	-0.122	0.039	-0.236	-3.1	0.003	-0.2	-0.043	0.787	1.271		
	Factor6	0.528	0.116	0.41	4.554	0.000	0.297	0.759	0.563	1.775		
	Factor8	0.502	0.13	0.381	3.869	0.000	0.243	0.76	0.469	2.13		
											0.686**	50.133**

a Dependent Variable: Successful Big Data implementations can result in the offering of new products and services

Test for 1st Dependent Variable we find that Change Management, Collaboration and Advanced Computing are significant factors.

Table 31. Quantitative Review – Dependent Variable #2 (New processes) Full Regression

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics		Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF	R2	F
1	(Constant)	3.336	0.422		7.899	0.000	2.493	4.179				
	Which Industry do you work in?	0.023	0.025	0.105	0.91	0.366	-0.027	0.073	0.892	1.121		
	Please state your profession level	0.04	0.093	0.049	0.433	0.666	-0.145	0.226	0.935	1.07		
	Which region are you based in?	-0.136	0.05	-0.301	-2.713	0.008	-0.236	-0.036	0.968	1.033		
	What is your age group?	0.308	0.106	0.324	2.92	0.005	0.098	0.519	0.971	1.03		
											0.211**	4.414**
2	(Constant)	-0.461	0.683		-0.675	0.502	-1.829	0.906				
	Which Industry do you work in?	-0.01	0.02	-0.046	-0.502	0.617	-0.05	0.03	0.789	1.267		
	Please state your profession level	0.055	0.072	0.067	0.767	0.447	-0.089	0.199	0.877	1.14		
	Which region are you based in?	0.012	0.045	0.026	0.258	0.797	-0.079	0.102	0.664	1.507		
	What is your age group?	0.338	0.084	0.355	4.034	0.000	0.17	0.505	0.866	1.154		
	Factor2	0.593	0.151	0.52	3.938	0.000	0.291	0.895	0.384	2.604		
	Factor3	-0.03	0.16	-0.023	-0.188	0.852	-0.351	0.291	0.428	2.335		
	Factor4	0.153	0.142	0.126	1.073	0.288	-0.132	0.438	0.489	2.043		
	Factor5	0.175	0.139	0.175	1.264	0.211	-0.103	0.454	0.351	2.852		
	Factor6	-0.111	0.186	-0.092	-0.595	0.554	-0.483	0.262	0.281	3.555		
	Factor7	0.095	0.114	0.081	0.838	0.406	-0.133	0.323	0.72	1.388		
	Factor8	0.319	0.158	0.264	2.023	0.048	0.003	0.634	0.394	2.539		
	Factor9	-0.23	0.167	-0.174	-1.379	0.173	-0.563	0.104	0.423	2.366		
	Factor10	-0.086	0.143	-0.085	-0.603	0.549	-0.372	0.2	0.338	2.957		
	Factor11	0.048	0.109	0.042	0.438	0.663	-0.171	0.266	0.718	1.393		
											0.625**	6.669**
3	(Constant)	0.021	0.4		0.053	0.958	-0.776	0.819				
	What is your age group?	0.323	0.078	0.327	4.167	0.000	0.168	0.478	0.956	1.046		
	Factor2	0.596	0.116	0.54	5.154	0.000	0.366	0.827	0.535	1.869		
	Factor8	0.226	0.122	0.196	1.848	0.069	-0.018	0.47	0.525	1.905		
											0.588**	33.356**

Test for 2nd Dependent Variable we find Organizational Development Competencies is a significant factor while Collaboration and advanced computing is not significant.

Table 32. Quantitative Review – Dependent Variable #3 (New Patents or Source of Competitive Advantage) Full Regression

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics		Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF	R2	F
1	(Constant)	4.278	0.443		9.653	0.000	3.393	5.163				
	Which Industry do you work in?	0.003	0.027	0.015	0.124	0.902	-0.05	0.056	0.861	1.161		
	Please state your profession level	-0.046	0.1	-0.053	-0.463	0.645	-0.245	0.153	0.9	1.111		
	Which region are you based in?	-0.203	0.052	-0.435	-3.917	0.000	-0.307	-0.1	0.975	1.026		
	What is your age group?	0.183	0.114	0.181	1.605	0.113	-0.045	0.41	0.943	1.06		
											0.219**	4.550**
2	(Constant)	0.645	0.732		0.881	0.382	-0.822	2.113				
	Which Industry do you work in?	-0.017	0.022	-0.076	-0.777	0.440	-0.061	0.027	0.754	1.327		
	Please state your profession level	-0.104	0.079	-0.121	-1.317	0.193	-0.263	0.054	0.856	1.169		
	Which region are you based in?	-0.101	0.048	-0.216	-2.082	0.042	-0.198	-0.004	0.672	1.487		
	What is your age group?	0.2	0.094	0.198	2.138	0.037	0.012	0.388	0.842	1.188		
	Factor2	0.259	0.162	0.22	1.602	0.115	-0.065	0.584	0.382	2.614		
	Factor3	0.467	0.172	0.353	2.717	0.009	0.122	0.811	0.43	2.327		
	Factor4	0.133	0.155	0.104	0.857	0.395	-0.178	0.445	0.487	2.053		
	Factor5	0.129	0.149	0.124	0.866	0.390	-0.17	0.429	0.351	2.85		
	Factor6	-0.14	0.204	-0.112	-0.686	0.496	-0.548	0.269	0.269	3.718		
	Factor7	0.168	0.122	0.137	1.373	0.175	-0.077	0.412	0.725	1.38		
	Factor8	-0.135	0.17	-0.107	-0.796	0.430	-0.476	0.205	0.4	2.499		
	Factor9	0.039	0.179	0.028	0.218	0.828	-0.319	0.397	0.423	2.365		
	Factor10	0.113	0.154	0.107	0.735	0.465	-0.195	0.421	0.34	2.937		
	Factor11	-0.042	0.117	-0.036	-0.356	0.723	-0.276	0.193	0.72	1.388		
											0.602**	5.395**
3	(Constant)	1.055	0.519		2.033	0.046	0.019	2.092				
	Which region are you based in?	-0.113	0.043	-0.233	-2.637	0.010	-0.198	-0.027	0.902	1.109		
	What is your age group?	0.142	0.088	0.136	1.615	0.111	-0.033	0.317	0.994	1.006		
	Factor3	0.762	0.114	0.592	6.673	0.000	0.534	0.99	0.898	1.114		
											0.520**	24.515**
a Dependent Variable: Successful Big Data Implementations will result in new patents or source of competitive advantage for the organization												

Test for 3rd Dependent Variable we find Initial Big Data Characteristics were significant along with Age and Geographic Region controls.

Table 33. Quantitative Review – Dependent Variable #4 (Further Business Strategy) Full Regression

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics		Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF	R2	F
1	(Constant)	3.862	0.351		11.013	0.000	3.162	4.563				
	Which Industry do you work in?	0.004	0.021	0.024	0.191	0.849	-0.038	0.045	0.892	1.121		
	Please state your profession level	0.1	0.077	0.161	1.299	0.199	-0.054	0.254	0.935	1.07		
	Which region are you based in?	-0.049	0.042	-0.143	-1.175	0.244	-0.132	0.034	0.968	1.033		
	What is your age group?	0.065	0.088	0.09	0.742	0.461	-0.11	0.24	0.971	1.03		
2	(Constant)	0.798	0.539		1.482	0.144	-0.281	1.878			0.056	0.972
	Which Industry do you work in?	-0.01	0.016	-0.059	-0.619	0.538	-0.041	0.022	0.789	1.267		
	Please state your profession level	0.086	0.057	0.138	1.517	0.135	-0.028	0.199	0.877	1.14		
	Which region are you based in?	0.055	0.036	0.162	1.555	0.126	-0.016	0.127	0.664	1.507		
	What is your age group?	0.049	0.066	0.068	0.74	0.463	-0.083	0.181	0.866	1.154		
	Factor2	0.175	0.119	0.202	1.469	0.147	-0.063	0.413	0.384	2.604		
	Factor3	0.076	0.126	0.078	0.6	0.551	-0.177	0.329	0.428	2.335		
	Factor4	0.133	0.112	0.143	1.18	0.243	-0.093	0.358	0.489	2.043		
	Factor5	0.098	0.11	0.129	0.897	0.373	-0.121	0.318	0.351	2.852		
	Factor6	0.28	0.147	0.306	1.907	0.062	-0.014	0.574	0.281	3.555		
	Factor7	-0.186	0.09	-0.207	-2.07	0.043	-0.366	-0.006	0.72	1.388		
	Factor8	0.308	0.124	0.336	2.477	0.016	0.059	0.557	0.394	2.539		
	Factor9	-0.017	0.131	-0.017	-0.128	0.898	-0.28	0.246	0.423	2.366		
	Factor10	-0.111	0.113	-0.144	-0.983	0.330	-0.337	0.115	0.338	2.957		
	Factor11	-0.018	0.086	-0.021	-0.208	0.836	-0.19	0.155	0.718	1.393		
											0.595**	5.868**
3	(Constant)	1.682	0.412		4.078	0.000	0.86	2.504				
	Factor7	-0.124	0.098	-0.119	-1.273	0.207	-0.319	0.07	0.875	1.143		
	Factor8	0.719	0.096	0.701	7.484	0.000	0.527	0.91	0.875	1.143		
											0.447**	29.089**
a Dependent Variable: Big Data can help further business strategy												

Test for 4th Dependent Variable we find Challenging Status Quo (negative), Collaboration and Advanced Computing (positive) were significant factors.

Using the above models to test our hypotheses, we have the following results:

H1: Opinions of Organizational Change Management will be positively associated with higher perceptions of successful Big Data implementations as generator of new products and services. (TRUE)

H2: Opinions of Collaboration and advanced computing will be positively associated with higher perceptions of successful Big Data implementations as generator of new products and services. (TRUE)

H3: Opinions of organizational development competencies will be positively associated with higher perceptions of successful Big Data implementations as developer of new processes. (TRUE)

H4: Opinions of Collaboration and advanced computing will be positively associated with higher perceptions of successful Big Data implementations as developer of new processes (FALSE; positive but no significance)

H5: Opinions of initial Big Data characteristics will be positively associated with higher perceptions of successful Big Data Implementations as generator of new patents and source of competitive advantage (TRUE)

H6: Opinions of Collaboration and advanced computing will be positively associated with higher perceptions of successful Big Data Implementations as furthering of business strategy (TRUE)

H7: Opinions of challenging status quo will be positively associated with higher perceptions of furthering business strategy as part of successful Big Data Implementations (FALSE; Negative and no significance)

In summary, in this chapter we focused on 5 items: (a) we examined the raw data via six rounds of data examination, (b) we induced a model from our examination and cross referenced literature for consistency of findings, (c) we created an online survey from our model to use for statistical testing, (d) statistically tested relationships of our variables via data collected from the online survey and (e) found our hypotheses to stand true except for two (H4 and H7). In the next chapter, we discuss the findings.

Chapter 5 Discussion

The purpose of this chapter is to discuss the key findings of this research, the studies contribution to management literature and definition of Big Data given the data collected.

Key Findings - Qualitative

The research study started with a question around understanding factors that impact successful Big Data Implementations. The literature review provided us with details around Big Data as a concept in its infancy with room for study. The qualitative portion of this study allowed us to speak with practitioners in the field who gave us insights into how they have successfully implemented Big Data initiatives. It enabled exploration into criteria used in organizations for success relating to Big Data Implementations, such as, creation of new products and services, patents and competitive advantage, process improvements, and promoting strategy. First, through further qualitative analysis of the responses, four major patterns emerged: (a) organizational characteristics highlighting cross functional collaboration, leadership support, methodology, funding and data governance, (b) technical competence for organizations highlighted by technical skills, platform and tool/devices, (c) variance in defining Big Data that allowed categorizing and engaging in such initiatives in form of characteristics such as, variety, volume, velocity, value, veracity, computing resources and variability and evolution of such and (d) other challenges which are important to many as Big Data evolves as a concept, namely security and privacy. This led to the formation of a statistically testable model. Given previous research done in this area by Manyika et. al (2011), a comprehensive review of literature by Wamba et. al (2015) and conceptual model by Halaweh and Massry (2015) the findings were consistent in terms of identifying the major themes. There were a few extra factors found from the qualitative section that did not appear in prior literature around Big Data Implementations, such as: data governance, organizations' information intensity, evolution of Big Data characteristics, use of computing resources, definition of is Big Data, business area involvement and cross functional collaboration but overall this research study was able to validate those factors via the qualitative portion of this research study.

Key Findings - Quantitative

The quantitative portion resulted in testing of relationships between the variables. The quantitative process was done in 5 parts:

1. KMO and Bartlett's test to see if data can be used for analysis: we conducted KMO test to ensure that data can be used for further analysis. The cutoff used in this research was greater than 0.5. The result was a healthy 0.72 and we were good to proceed.
2. Run Factor Analysis to see if data can be reduced: factor analysis was conducted on the data. The goal of the factor analysis is to reduction of data. Factor analysis was done in three parts:

- Communalities: communalities estimate variance in each variable accounted by all the available factors/components. The extraction communalities are estimates the variance in each variable is accounted for by the components. The lowest was ~0.6 which means that the variables are well represented.
- Total variance: running the total variance tells us that 11 components explain ~75% of the variance in the data.
    - The cutoff point to look at items loaded was >0.4
    - Further examination of the 11 factors revealed that out of the 37 variables one was not part of the 11 factors. All 11 factors are listed in tables 15 – 25.
- After reducing our data to 11 factors, the model had to be updated to reflect the new independent variables and new hypotheses were

3. Run Correlations: Correlations of the independent variables to the dependent variables were, for the most part, strong and significant. Most of the correlations resulted in positive moderate to strong (> +0.25) relationships with significance (at 0.01 level). The real important aspect to consider here is that variables tested for Organizational Development characteristics and competences (Change Management, understanding of behavioral science/knowledge, cross team collaboration, business involvement, technical skills, and use of computing resources) were all positive, strong (significance

at 0.01 level) and successful indicators for Big Data Implementations as evidenced in chapter 4 results.

4. Run Reliability: reliability testing revealed strong Cronbach's alpha at 0.914. This tells us that the scale is reliable. This test was for measuring internal consistency. Factor 7 and 11 were identified as two that if removed would result in a higher alpha (by <=0.005). The reason for not removing them was two-fold: (a) the alpha would have increased by only 0.005 and (b) factor 7 was identified as significant in our regression analysis.
5. Run Regression on the full model: finally, regression results show positive and strong significance for 4 factors out of 11. It also proved 5 hypotheses out of 7.

In summary, we can conclude the following from the analysis conducted:

(a) Given the current data and analysis we can conclude that opinions on Organizational Development characteristics (change, collaboration), competencies (technical skills and support) and change management (enabling change and initiatives are for betterment for organization) are associated with higher perceptions as part of measuring successful Big Data Implementations via new products, services, competitive advantage, patents or furthering of business strategy.

(b) While the above is true and focuses on core Organizational Development constructs, we cannot conclusively state that Organizational Development leading (aside from collaborating or leading change specific areas) is key to understanding perceptions on successful Big Data Implementations. There is, however, a case to be made where certainly change management expertise, collaboration and specifically focusing on betterment of organization (all OD specific areas of focus) can be considered important aspects in successful Big Data Implementations.

Contribution to Literature:

Halaweh and Massry (2015) said it best, future contribution to this field would be to (a) applying qualitative methods to interview experts who work in different sectors to develop or extend the model that affect Big Data implementation and (b) use quantitative research methods to test model and verify validity of the assumptions. This is what this research study has accomplished. At the end, identification of characteristics relating to organizations and technical skills were identified after engaging with experts in the field to deduce a testable model. Having gone through the two-part study and gathered the data it is imperative to provide a definition that is more dependent on expert analysis and empirically tested data. This research, based on data gathered, defines Big Data as:

Definition: *Big Data is a characterization describing quick varied accumulation of information on which computing resources must be applied to derive new products/services, create efficiencies, form strategy, or gain competitive advantage.*

This above definition keeps all the characteristics of Big Data that the industry has discussed (volume, variety, veracity, velocity, variability and value) and is not limited by new additions as long as it is within the core areas of this definition. The core areas of this definition are: (a) that the term is used as a characterization, (b) is an accumulation defined by the 6Vs, (c) on which computing resources must be applied and (d) analysis are conducted to provide value to the organization in form of new products or services, new strategy, new processes or competitive advantage.

Implications and Study Limitations

There are some limits to the findings of this research that should be noted. While passing the KMO test of sampling adequacy for this study, the sample size of 75 is still a limitation by quantitative research standards given that the survey was circulated to more than 669,361 individuals.

Acknowledging that diversity in responses did exist but they were limited to a few specific regions and to specialized roles mostly due to personal and professional connections of the researcher.

Anonymity was a major issue and respondents did not want to be recorded. As such, there were items that were generalized to fit within the study. In the case of the online surveys, some participants would start and then stop at various points and the researcher has no insight into the cause. As those were incomplete responses, the data was not included in this study. Some people opted out entirely as the study was presented as voluntary.

Regarding the survey itself, the questions were formed mostly with experience of the researcher, expert opinions and overall short in nature (42 variables) as no prior versions of the instruments were utilized and limited by SurveyMonkey when utilizing the targeted audience feature. Furthermore, the survey measured constructs in multiple ways that would have limited impact if tested with one variable. Moreover, there was no intention of linking factors taken from interviews with the success of Big Data Implementation. In other words, if the interviewer only spoke in terms of new products or services, all factors that impacted such were not just linked to that success criteria. It is possible that this may have provided a stronger link if taken into consideration earlier.

Finally, the coding analysis were done by the researcher only. It was subject to interpretation. Furthermore, it is important to note that English is not the researcher's first language which may have caused the researcher to misinterpret what was communicated in the interviews.

Every study has its limitations and the research presented here is no exception. However, every attempt was made to collect an appropriate data set for the research and enough information to make informed conclusions.

Chapter 6 Conclusion

OD as a field has been around for many decades while Big Data is a new phenomenon that exists in today's complex, hyper connected world. Big Data in its most raw form contains the necessary ingredients that can help OD in maximizing value, improving efficiency and keeping its focus on behavioral knowledge for planned change. There is a case to be made for Big Data's use in every industry from finance to marketing to education. The use cases that exist covers for profit businesses and not for profit organizations alike. Some use cases such as:

- Data management—smarter data management to offer potential savings via storage, extraction and analysis.
- Personalization of services—by better understanding individuals or groups of individuals, agencies may be able to offer more tailored services and products.
- Real time data coupled with predictive analytics will advance problem-solving capabilities and offer better insights to support decision-making.
- Productivity and efficiency—analysis of big data can identify cost savings and efficiency opportunities.

OD, in a sense is all about challenging tradition. If it was done a certain way, we can do it better now. The driving force or motivation, it seems, is derived from better understanding and availability of knowledge, tools, technology and other resources. Traditionally, organizations have existed as teams of people, businesses, and even individually run but today data, software, hardware, artificial intelligence, and virtual reality all form part of the organization and its assets. These are due to advances in computing, real time data availability and mobility. Organizations are now increasingly being known for their digital property, footprint and presence. This has caused a shift in how organizations tackle the value proposition provided by them. The organizations don't need to be left behind when new technology or data, in our case, is available to be "opportunized". It was a similar case when e-commerce came calling and left many brick and mortars behind who did not feel obliged to take notice. As such, I believe that the current form and method of data gathering that exists will change with further technological advances. OD will need to continually respond and adapt to the many technological advances. It may even be required that OD examines the implementations of its models due to the growing complexities and the near real time decision making power that is being made available to us via Big Data. It will also require us, as scholar-practitioners, to be able to tell the story, understand the vision and guide the organization with the data that we can collect, analyze and use for feedback. Ethics, privacy, security and the basic governance of the data will also be a major source of contention. We would want to ensure cross collaboration across the organization to ensure the basic OD values are being realized and organization effectiveness is at the core. In short, the future is bright shining with opportunity. It is finally time for us, as OD scholars-practitioners to soak in the sun and do what is natural, allowing organizations, in the many forms they exist, to showcase their value effectively.

Future Research

There are many areas of future research. The survey tool and questions can be improved upon. There was some consistency that was observed between literature, Halaweh and Massry and this research's model. The study was limited to four classifications of what successful Big Data Implementation is, this can certainly be expanded and/or refined further. It would be beneficial to have a much larger sample size to allow for better statistical testing. Geographic region, which was found to be significant with certain classifications, can also be explored further in terms of value and success of Big Data Implementations. Specific factors can be examined at a more micro level to validate impact.

Finally, I leave you with this vignette from Church and Dutta (2013) that summarize my sentiments and passion regarding this subject:

"Big Data can help OD practitioners look above and beyond their traditional organizational perspective to infer insights that are much more proactive than existing methods. The synthesis of variety, volume, velocity, and veracity of people data and beyond if applied appropriately to inform OD strategies could prove to be a very powerful tool for the future. The addition of a Big Data mindset may also finally provide a compelling argument for enhancing the data analytic skills and storytelling capabilities of OD practitioners going forward. Based on this discussion the future of Big Data-driven OD for change has potential indeed".

References

A. Gandomi, M. Haider / International Journal of Information Management 35 (2015) 137–144 139

Allouche, G. (2014). How Big Data can Save Health Care. Retrieved from: http://www.innovationexcellence.com/blog/2014/11/14/how-big-data-can-save-health-care/

Ann Keller, S., Koonin, S.E., Shipp, S., 2012. Big data and city living – what can it do for us? Significance 9 (4), 4–7. http://dx.doi.org/10.1111/j.1740-9713.2012. 00583.x.

Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). Finding value in the information explosion. *MIT Sloan Management Review*, *53*(4), 18.

Beer, Michael, Russell A. Eisenstat, and Nathaniel Foote. (2009). *High commitment high performance: How to build a resilient organization for sustained advantage*. John Wiley & Sons.

Boyd, D. & Crawford, K. (2012). *"Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon."*Information, Communication, & Society 15:5, p. 662-679.

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, *6*(3), 188-205.

Brown, B., Chul, M., Manyika, J., 2011. Are you ready for the era of 'big data'? McKinsey Q. 4, 24–27þ30-35.

Buono, A., Jamieson, D. (ed.) (2010). *Consultation for Organizational Change.* Information Age Publishing, Inc. Charlotte, NC ISBN: 978-1-61735-086-3

Buono, A. F., & Kerber, K. W. (2010). Intervention and organizational change: Building organizational change capacity. *EBS Review*, *27*, 9-21.

Burg, N. (2014). How Big Data Will Help Save Healthcare. Retrieved from: [http://www.forbes.com/sites/castlight/2014/11/10/how-big-data-will-help-save-healt]

Burke, W. W. (2014). *Organizational change theory and practice*. California: Sage Publications.

C DiStefano, M Zhu, D Mindrila - Practical Assessment, Research & 2009 - pareonline.net

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165–1188.

Church, A., Dutta S. (2013). *The Promise of Big Data for OD.* OD Practitioner. 45(4).

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

Creswell, D., and Plano-Clark, V. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks: CA: Sage Publications

Cummings, T., Greiner, L. (2009). *Dynamic strategy-making*: *A real-time approach for the 21st century leader.* Jossey Bass. ISBN: 978-0-7879-9663-5.

Cummings, T., Worley, C. (2014). *Organization Development and Change, 10th ed.* Cengage Learning Publishing. ISBN: 978-1-13319-045-5.

Davenport, T. H., Barth, P., & Bean R. (2013). How 'big data' is different. MIT Sloan Management Review, 54(1).

Dobbs, Manyika, Woetzel (2015). *The four global forces breaking all the trends.* Retrieved from: http://www.mckinsey.com/insights/strategy/the_four_global_forces_breaking_all_the_trends

Dumbill, E. (2013). Making sense of big data.

Erwin, R. (2015). Data literacy: Real-world learning through problem-solving with data sets. American Secondary Education, 43(2), 18-26

Forrester, 2012. The Big Deal About Big Data For Customer Engagement Business: Leaders Must Lead Big Data Initiatives To Derive Value. Available from 〈http:// www.forrester.com/The Big Deal About Big Data For Customer Engagement/fulltext/-/E-RES72241〉

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D. (2015), "How 'big data' can make big impact: findings from a systematic review and a longitudinal case study", International Journal of Production Economics, Vol. 165, pp. 234-246.

Friedman, T. L. (2006). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus and Giroux.

Fyke, J.P., Buzzanell, P.M. (2013) *The Ethics of Conscious Capitalism: Wicked Problems in Leading Change and Change Leaders.* Human Relations, 66(12), 1619-1643

Galbraith, J., (2014). *Organization Design Challenges resulting from Big Data.* Journal of Organization Design. 3(1). 2-13.

Galbraith, J., Downey, D., Kates, A. (2002). *Designing Dynamic Organizations: A Hands-On Guide for Leaders at All Levels*. New York: AMACOM.

Gallos, J. (ed.) (2006). *Organization Development: A Jossey-Bass Reader.* San Francisco: Jossey-Bass. ISBN: 978-0-78798-426-7

Gamage, P. (2016). New development: Leveraging 'big data'analytics in the public sector. *Public Money & Management*, *36*(5), 385-390.

Gartner. (2012). IT Glossary – Big Data, http://www.gartner.com/it-glossary/big-data

Gehrke, J., 2012. Quo vadis, data privacy? Ann. New York Acad. Sci. 1260 (1), 45–54. http://dx.doi.org/10.1111/j.1749-6632.2012.06630.x.

George, G., Haas M. R., Pentland, A. (2014). *Big Data and Management.* Academy of Management Journal. 57(1).

Gobble, M. M. (2013). Outsourcing innovation. *Research-Technology Management*, *56*(4), 64-67.

Groves, P., Kayyali, B., Knott, D. & Kuiken, S. (2013). The big data revelation in Healthcare: Accelerating value and innovation.

Gudivada, V. N., Baeza-Yates, R. A., & Raghavan, V. V. (2015). Big Data: Promises and Problems. *IEEE Computer*, *48*(3), 20-23.

Hartmann P.M., Zaki, M. Feldmann N., & Neely, A. (2014). Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms, Cambridge Service Alliance.

Head, T., Sorensen, P., Baum, B., Yaeger, T., Cooperrider, D. (2007). *Organization Behavior and Change: Managing Human Resources for Organizational Effectiveness, 14th ed.* Champaign: Stipes Publishing. ISBN: 978-1-58874-749-

Head, Thomas C. (2009). Strategic Organization Development: A Failure of True Organization Development? In T. Yaeger & P. Sorensen (Eds.), *Strategic Organization Development: Managing Change for Success.* (pp 23-42). Champaign: Information Age Publishing.

Hofstede, G. (2010). Geert hofstede. *National cultural dimensions*.

Hopkins, B., & Evelson, B. (2012). Forrester: Big data–start small, but scale quickly. *Computer Weekly. Retrieved August*, *8*, 2012.

Hopkins, M. S., LaVelle, S., Lesser, E., Shockley R., Kruschwitz, N. (2011). *Big Data, Analytics and the Path from insights to value.* MIT Sloan Management Review. 52 (2).

IBM. (2011). *The FOUR V's of big data*. Retrieved from http://www.ibmbigdatahub.com

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, *57*(7), 86-94.

Jelinek, M., & Litterer, J. (1988). Why OD must become strategic. *Research in organizational change and development*, *2*(1), 135-162.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed.). Belmont: Cengage Learning.

Khilji, E. (2002). Modes of convergence and divergence: An integrative view of multinational practices in Pakistan. *International Journal of Human Resource Management*, *13*(2), 232-253.

Kiron, D., & Shockley, R. (2011). Creating business value with analytics. *MIT Sloan Management Review*, *53*(1), 57.

Koutroumpis, P., & Leiponen, A. 2013. *Understanding the value of (big) data*. In Proceedings of 2013 IEEE international conference on big data. 38–42. Silicon Valley, CA, October 6–9, 2013. Los Alamitos, CA: IEEE Computer Society Press

Kuketz, D. (2013). 13 New Business Models Created By Big Data. Retrieved from: http://www.utopiainc.com/insights/blog/505-13-new-business-models-big-data

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387-394.

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. Proceedings of the VLDB Endowment, 5(12), 2032–2033.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. MIT Sloan Manag. Rev. 52 (2), 21–32.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D.,Christakis, N., Contractor, N., Fowler, J.,Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). *'Computational Social Science'*. *Science* vol. 323, pp. 721-3.

Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers*, *17*(2), 243-259.

Lawrence, P. R., & Lorsch, J. W. (1967). Differentiation and integration in complex organizations. *Administrative science quarterly*, 1-47.

Lorsch, J.W. & Mathias, P.F. (2001). When professionals have to manage. Harvard Business Review, July-August: 78-83

M. Halaweh and A. El Massry. 2015. Conceptual Model for Successful Implementation of Big Data in Organizations, Journal of International Technology and Information Management, Vol. 24, No. 2, 21–29

Marshall, B., Cardon, P., Poddar, A., & Fontenot, R. (2013). Does sample size matter in qualitative research?: A review of qualitative interviews in IS research. *Journal of Computer Information Systems*, *54*(1), 11-22.

Malik, P. (2013). Governing big data: Principles and practices. *IBM Journal of Research and Development, 57*(3), 1-13. doi:10.1147/JRD.2013.2241359

Mantelero, A., & Vaciago, G. (2015). Data protection in a big data society. Ideas for a future regulation. *Digital Investigation*, *15*, 104-109.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute. Retrieved from http://www.mckinsey.com

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. *McKinsey Global Institute*, 21.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McAfee, A., Brynjolfsson, E., (2012). *Big Data: The Management Revolution.* Harvard Business Review. 90.

McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition, and productivity.* June 2011. Lexington, KY: McKinsey & Company.

Miller, H. G., & Mork, P. (2013). From data to decisions: A value chain for big data. *IEEE IT Pro, 15*(1), 57-59. doi:10.1109/MITP.2013.11

Miller, S. (2014), "Collaborative approaches needed to close the big data skills gap", Journal of Organization Design, Vol. 3 No. 1, pp. 26-30.

Munford, M. (2014). *Rule changes and big data revolutionise Caterham F1 chances*. The Telegraph, Technology Section, 23 February 2014. Available from: http://www.telegraph.co.uk/technology/technology-topics/10654658/Rule-changes-and-big-data-revolutionise-Caterham-F1-chances.html.

Nadler, D. A. (1976). The use of feedback for organizational change: Promises and pitfalls. *Group & Organization Management*, *1*(2), 177-186.

Opresnik, D. & Taisch, M. (2015). The value of big data in servitization. *International Journal of Production Economics*, 165, 174-184.

Oracle (2013). Big Data for the Enterprise An Oracle White Paper. http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51-59.

Parry, M. (2010, May 28). The humanities go Google.

The Chronicle of Higher Education. Retrieved from http://chronicle.com/article/The-Humanities-Go-Google/65713/

Perrey, J., Spillecke, D., Umblijs, A., 2013. Smart analytics: how marketing drives short-term and long-term growth. McKinsey Q.

Punathambekar, A., & Kavada, A. (2015). Debating Big Data. *Media, Culture & Society*, 0163443715594870.

Rooney, B. (2012). Health Care is the Next Frontier for Big Data. *Wall Street Journal. Wall Street Journal*.

Rijmenam, M. (2014). Big Data Will Revolutionize Education. Retrieved from: https://datafloq.com/read/big-data-will-revolutionize-learning/206

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of big data. How innovative enterprises extract value from uncertain data. IBM Institute for Business Value. Retrieved from http://www-03.ibm.com/systems/hu/resources/the real word use of big data.pdf

Shull, A. C., Church, A. H., & Warner Burke, W. (2014). Attitudes about the field of organization development 20 years later: The more things change, the more they stay the same. In *Research in organizational change and development* (pp. 1-28). Emerald Group Publishing Limited.

Strawn, G.O. (2012), "Scientific research: how many paradigms?", EDUCAUSE Review, Vol. 47 No. 3, pp. 26-34.

Stringer, R. (2002). *Leadership and organizational climate*. Upper Saddle River, NJ: Prentice-Hall.

Styrk B. (2015). *How do Organizations prepare and clean Big Data to achieve better Data Governance? A Delphi Study.* (Doctoral dissertation). ProQuest (UMI Number: 3682586)

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). Using multivariate statistics.

TechAmerica Foundation (2012). Big Data – A Practical Guide To Transforming The Business of Government. Available at: https://www304.ibm.com/industries/publicsector/fileserve?contentid=239170

Tormay, P. (2015). Big data in pharmaceutical R&D: Creating a sustainable R&D engine. *Pharmaceutical medicine*, *29*(2), 87-92.

Vaidhyanathan S. (2010) More on Google, China, censorship, and motivations. The

Googlization of Everything, 14 January. Available at: http://www.googlizationofeverything.com/2010/01/more_on_google_china_censorshi.php

Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D., 2015. How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. Int. J. Prod. Econ. http://dx.doi.org/10.1016/j.ijpe.2014.12.031.

Weisbord, M. (2012) *Productive Workplaces: Dignity, Meaning, and Community in the 21st Century.* Pfeiffer Publishing, 3rd edition. ISBN: 978-0-47090-017-8.

White, M., 2012. Digital workplaces: vision and reality. Bus. Inf. Rev. 29 (4), 205–214.

Worley, C., Hitchin, D., Ross, W. (1996). *Integrated Strategic Change: How OD builds competitive advantage*. Addison – Wesley Publishing Company. ISBN: 0-201-85777-4

Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., ... & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems*, *34*(1), 1.

Yaeger, T., Sorensen, P. (ed) (2009). *Strategic Organization Development: managing change for success.* IAP-Information Age Publishing, Inc. ISBN: 978-1-60752-210-2

Young, J. M., (2014). *An Epidemology of Big Data.* (Doctoral Dissertation) ProQuest (UMI Number: 3620515)

Appendix A: List of Databases

1. Academic Search Complete
2. AHFS Consumer Medication Information
3. ATLA Catholic Periodical and Literature Index
4. ATLA Religion Database with ATLASerials
5. Audiobook Collection (EBSCOhost)
6. Business Source Complete
7. Business Source Elite
8. Child Development & Adolescent Studies
9. CINAHL Plus with Full Text
10. Communication & Mass Media Complete
11. Consumer Health Complete - EBSCOhost
12. eBook Collection (EBSCOhost)
13. Educational Administration Abstracts
14. ERIC
15. Funk & Wagnalls New World Encyclopedia
16. GreenFILE
17. Health Source - Consumer Edition
18. Health Source: Nursing/Academic Edition
19. Humanities International Complete
20. Library, Information Science & Technology Abstracts
21. MAS Ultra - School Edition

22. MasterFILE Premier

23. Mental Measurements Yearbook with Tests in Print

24. Military & Government Collection

25. MLA Directory of Periodicals

26. MLA International Bibliography

27. Newspaper Source

28. OmniFile Full Text Select (H.W. Wilson)

29. Primary Search

30. Professional Development Collection

31. PsycARTICLES

32. PsycEXTRA

33. PsycINFO

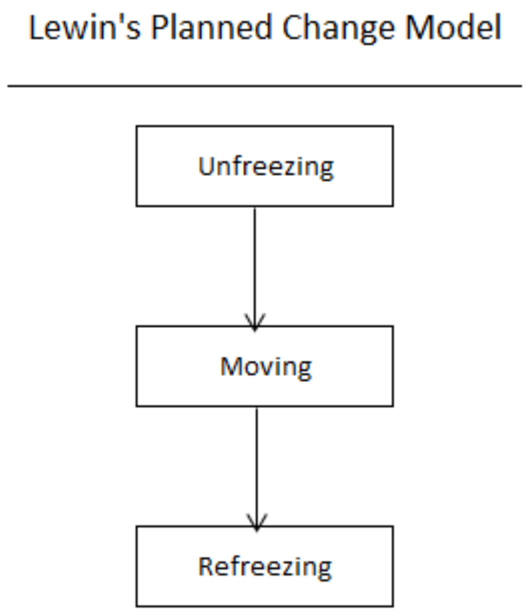
34. Regional Business News

35. RILM Abstracts of Music Literature (1967 to Present only)

36. SPORTDiscus with Full Text

37. MathSciNet via EBSCOhost

Appendix B: Kurt Lewin Change Model



Appendix C: Interview Cover Letter

Background and Purpose

I am a student at Benedictine University in Lisle, Illinois engaging in research related to my dissertation project. The purpose of this research project is to examine which organizational and technical characteristics affect Big Data Implementations and external factors that moderate that relationship. You are invited to participate in a research project because of your Big Data experience.

Procedures

Thank you for your participation in the 45 minute interview. Be assured that your responses will be kept confidential and anonymous. Your participation is voluntary and individual responses are kept confidential. The purpose of this phone interview is to lean on experts and gain knowledge from those who have been involved in Big Data implementations and projects. This will allow the researcher to compile an anonymous survey, to be distributed through LinkedIn Groups, outlining organizational characteristics, technical competence and external factors.

Confidentiality and Risk

This study is being undertaken for my research only and will not be shared in an identified context. It is my intention to publish my dissertation research upon completion. Once the study is complete, the data will be kept on file in a safe and secure location at Benedictine University in Lisle, IL USA.

We conform to all aspects of the Ethics Code of the National Institutes of Health (NIH) to protect identifiable research information from forced disclosure. Names and other identifiable information will not be used in published research. All responses will be transferred to a secure, password-restricted server. Access to raw data will be tightly restricted to only those individuals directly involved in data analysis. Although the Steering Committee members of this dissertation and Benedictine University may have access to the redacted data, this author will retain the sole ownership of all raw data. Once data analysis is completed, this committee will review the proposed data analysis to ensure it complies with the confidentiality policies listed above. The results of this study will be used for scholarly purposes only and may be shared with Benedictine University representatives in a summarized format. Once the study is complete, the data will be kept on file in a safe and secure location at Benedictine University.

If you have any questions about the research study, please contact Talha Ashraf at talha_ashraf@yahoo.com - or- Dr. Alandra Weller Clarke, Chair IRB, Benedictine

University at AClarke@ben.edu or 630-829-6295. This research has been reviewed according to Benedictine University IRB procedures for research involving human subjects.

Appendix D: Interview Protocol

Interview #_______________
Date_______/_____/_______

Interview Protocol

Script

Welcome and thank you for your participation today. My name is Talha Ashraf and I am a PhD student at Benedictine University in Lisle, IL. This interview will take about 45 minutes and will include series of questions regarding your experiences surrounding Big Data initiatives in organizations. I would like your permission to record notes during this interview, so I may accurately document the information you convey. If at any time during the interview you wish to discontinue the use of the interview itself, please feel free to let me know. All of your responses are confidential. Your responses will remain confidential and will be used to develop a better understanding of a survey to better understand effects of organizational characteristics and technical competence for Big Data implementation. The purpose of this study is to increase our understanding and add to management literature on this subject.

At this time I would like to remind you of your written consent to participate in this study. I am the responsible investigator, specifying your participation in the research project: Effects of Organizational Characteristics and Technical Competence on Big Data implementation and what external factors affect the relationship. You and I have both signed and dated each copy, certifying that we agree to continue this interview. You will receive one copy and I will keep the other locked and separate from your reported responses. Thank you.

Your participation in this interview is completely voluntary. If at any time you need to stop, take a break, or return a page, please let me know. You may also withdraw your participation at any time without consequence. Do you have any questions or concerns before we begin? Then with your permission we will begin the interview.

Interview Questions:

1. What are some of the initiatives that your organization is currently undertaking? Please elaborate on the variety of initiatives.

2. Of the initiatives you mentioned above, what is the organization trying to achieve with these initiatives? Can you categorize these initiatives and if so, can you give examples (ex: technology, business, marketing, etc)?

3. What are some success criteria that the organization uses to measure these initiatives? Furthermore, in your previous experience are these characteristics similar, different or a combination?

4. Do you have any initiatives that are categorized as Big Data projects? If so, what do you think differentiates these projects from the other projects that are being undertaken in your organization?

5. What characteristics would you identify with Big Data initiatives in your organizations that you believe will help with the success of these initiatives? Please elaborate on what success criteria would be for the organization and for you individually.

6. What would you do differently knowing what you know today about Big Data initiatives?

7. Can you describe examples, from your experiences or from what you have heard/know, of any unsuccessful Big Data or technology initiatives that were undertaken? What were some of the characteristics that stood out from an organization perspective? From a technical perspective?

8. Do you have any closing thoughts or comments that you would like to share?

***** If participant wishes to discontinue study, ask if they would be willing to share why:**

Thank the participant for his/her participation.