ABSTRACT

Big data and analytics can lead to new avenues for IT-enabled innovations. The fundamental premise of this study examines the main collaborative and innovative capabilities of big data. Innovations can deliver important outcomes for governments and businesses. Further, this paper develops a theoretical framework for the examination of multiple types of collaborations using big data resources. Challenges and risk for collaborations and innovations are examined, with recommendations for future projects.

KEYWORDS: big data, analytics, collaboration, innovation

INTRODUCTION

Big data refers to “datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze” (McKinsey Global Institute, 2011). The big data phenomenon is the recognition that a vast amount of data is generated and collected daily. Seven pillars of commonality provide a foundation for big data: volume, velocity, variety, viscosity, variability, veracity, and volatility (DeSouza, 2014). All seven characteristics attest to the extensive scope, depth, and multi-dimensional complexity of big data. This research is driven by two main questions: 1. how can big data and analytics drive innovation and 2. How can inter-organizational collaborations serve as a catalyst for innovation in the age of big data?

IT-ENABLED INNOVATION THROUGH COLLABORATIONS

Innovations can be broad and varied. For this paper the specific context focus for innovations are IT-enabled innovations. IT-enabled innovations refer to all innovations that were developed and or contain technology-based components (ranging from hardware to software components). IT-enabled innovations represent a key revenue driver for many companies in both private and public domains. More specifically, IT-enables innovations focus on IT as a core competence of a business, and drives value creation and a competitive advantage (Harmon & Demirkan, 2015). Innovations can occur on a spectrum of development ranging from continuous to discontinuous. A continuous innovation is one that has had incremental changes added to it from its previous version. For example a new version of a software product can be classified as a continuous
innovation. Discontinuous innovations are much more disruptive and result in more pronounced behavioral modifications from end-users.

Consider the example of the iphone. The iphone was an innovative (more discontinuous than continuous) product released by Apple in 2007. Since its early release this innovation has spurred the growth of many new innovative products such as mobile applications (apps) and additional mobile devices. Today apps are pervasive and used by individuals, for-profits, non-profits and government agencies. Mobile devices have changed the way that individuals and businesses operate on a daily basis.

In the United States collaborations between academic institutions and both government and private industry has seen the birth of many historic IT-enabled innovations such as the Internet and the world-wide-web. Collaborative projects utilize the expertise harnessed from the different partners. This can range from technical knowledge, analytical capabilities, equipment and infrastructure, to direct capital for funding the project. Crowdsourcing, which can occur as internal crowdsourcing; community crowdsourcing; open crowdsourcing; or broker crowdsourcing harnesses knowledge and expertise from a broad base of participants to generate new ideas (Simula & Ahola, 2014). More recently, multi-partner collaborative projects developed through crowdsourcing and crowdfunding are continuing the path to the development of IT innovations.

**Big data, analytics and innovation**

The affordability and ubiquity of collecting and storing large repositories of data is fueling the growth of big data. Businesses are also using the troves of data to answer previously unanswered questions, and to pose new questions that provide more insight and understanding about the business. Herein lies the value of analytical tools for illuminating the hidden secrets residing in the big data repositories. A recent study identified the areas of big data analytics – driven by data mining and statistical methods; mobile analytics – driven by mobile devices and applications; network analytics – driven by links, multi-nodes and social networks; text analytics – driven by unstructured text content; and web analytics – driven by web services and cloud applications, as the next wave of emerging analytics research areas (Chen, Chiang, & Storey, 2012). Through these different analytical methods, tools and technologies a business can better understand events that happened within the organization; evaluate the current state of the business; and also makes predictions and delineate best choices for the future.

Through sense-making analytics (complex analysis of multiple streams of big data) we can now derive meaning and connections from previously divergent topics. For example big data analytics can be used to reduce medical payment fraud within Medicare and Medicaid services, where data is collected from disparate entities such as government employees, healthcare providers and independent contractors (TechAmerica Foundation, 2012). The innovation here is derived from the improved processing enabled through the use of IT and big data resources.

Further, managerial decision making in both the private and public sectors is increasingly more data driven with less emphasis being placed on pure intuition and individual conjecture. With big data analytics business managers can capture and measure a myriad of variables to improve the quality of decision making (McAfee & Brynjolfsson, 2012). As more government agencies move towards using big data, there is also an increased ability to establish new products and services through collaborations with the private sector (TechAmerica Foundation, 2012). Arguably, collaborative innovations represent the next area of growth for the application of big data analytics, with big data serving as a bridge between innovation and collaboration.
Collaborations and innovations
A recent report indicates that the use of big data in public agencies is still in its infancy; it is primarily focused on the aggregation of internal data; and in need of collaborative leadership for the successful completion of projects (DeSouza, 2014). The discussion on collaboration in the context of the above DeSouza report focused on projects across different departments and organizations within government. As an extension, the scope of big data projects lends itself to collaborations into both the public and private sectors. In this paper public (also called public sector) refers to organizations that are under the umbrella of a federal, state or local government agency, such as a state’s department of transportation. The private sector refers to businesses not under the government’s control – these can be privately held or publicly traded firms. Lastly, non-profit organizations can be considered as either public or private depending on their primary sources of funding.

Public-private partnerships (PPPs) represent endeavors between private and public entities that are “long-term contracts for the delivery of assets and services” (Liu & Wilkinson, 2014). However, more recently public-private collaborations (PPCs) have emerged which differ from PPPs because they are less formal and are much more flexible structures (Guzmán & Sierra, 2012). Guzmán & Sierra (2012) use the example of a private company maintaining a public beach, and getting the naming rights for that beach to illustrate a PPC. These types of collaborative projects present innovative solutions to problems that involve partners with common goals of improving neighborhoods and building communities. The web, through social media sites such as Facebook and Nextdoor foster environments for information sharing and collaborative problem solving within communities.

PPCs can be varied in scope, size, purpose, and duration, and analysis of data can reveal more information about where collaborative opportunities exist. With the advent of big data, more opportunities for innovation emerge as a result of insights and understanding gathered from data mining and big data analytics. Arguably, one area for utilizing big data is for the development and implementation of collaborative projects. One area that has great potential for public-private collaboration is healthcare management. Further, the White House has explicitly focused attention on the collaborative potential of big data analytics between public and private entities to address healthcare problems (applieddatalabs.com, n.d.).

The benefits of partnerships and collaborations between government institutions and private businesses include greater efficiencies for task completion; cost reductions to tax payers; and improved quality of products and services (Cellucci, 2010). Even though there are clear and defined benefits of public-private collaborations, a primary challenge is to determine where and when such collaborations are appropriate (Guzmán & Sierra, 2012). The presence of effective big data analytics can provide more insights on this phenomenon, given the multiple views of the data that now exists. Big data analytics can serve as the driver to shed light on what sectors and types of projects can be more appropriate for PPCs. Even given global conditions of fiscal austerity and political/economic upheavals, strategic opportunities can emerge from analyzing big data in the public and also private arenas.

Collaborations between governments and different types of institutions have always existed, but the introduction of big data into this landscape can lead to new avenues for sharing data across different types of businesses and agencies. OpenData500 is a project launched by the Governance Lab at New York University, specifically examining how businesses are leveraging value from open government data to develop new products and services (www.opendata500.com). This project is not specifically looking at big data, but it highlights the
growing intersection of data driven initiatives between the public and private sectors. Big data involves not only collection of data, but also asking the right questions of the data.

Even though private-private collaborations are not the focus of this paper, they can provide benchmarks and templates for public sector entities. Collaborations between private companies represent a hallmark for business ventures and innovations. Outsourcing ventures represent one form of collaboration that many businesses have embraced. Companies engage in outsourcing to reduce internal costs and to focus resources on their core competencies. A second area for collaborations in the private sector involves the pooling of resource to reduce risks, or solving a common problem. One direct outcome of collaboration is innovation.

THEORETICAL FRAMEWORK
The theoretical basis for this paper uses the resource dependency theory which examines the resources that an organization needs that exist outside of its sphere of control (Pfeffer & Salancik, 1978). As firms are better able to utilize and exploit external resources, they are better poised for success. Indeed, big data is an asset to both private and public organizations. As such, big data is a resource to both the entity that owns it and the external constituents that exist outside their immediate sphere of control. Dependence on external resources for an organization’s success shows the development of positive symbiotic relationships in such as supply chain management and consumer retailing (Hofer, Jin, Swanson, Waller, & Williams, 2012). Technologies such as email, texting, social media all support collaborations in a business environment. As individuals and business units build trust and share resources, the foundation elements for creative development are laid. Subsequently, as public organizations harness the power of big data analytics, there will be improved confidence to share ideas and methods to promote innovation with business partners.

INNOVATIONS IN CONTEXT
Innovations occur within different contexts. In this section of the paper three context areas for IT-enabled innovations are highlighted. These three areas: healthcare, cybersecurity and weather are selected because of their strong focus on big data initiatives.

Healthcare
Big data projects have the potential to reduce healthcare costs, improve the management of chronic diseases, reduce health insurance fraud, and increase the early detection of diseases (Lamont, 2014; TechAmerica Foundation, 2012). Watson, IBM’s famous Jeopardy winning computer system, is working collaboratively with leaders in the healthcare domain to use big data to solve complex medical problems. Doctors at the world renowned Memorial Sloan-Kettering Cancer Center in New York City are submitting entire patient medical records along with supplemental voice inputs to Watson, to generate the best treatment options for patients suffering from several different types of cancers (Klie, 2013). In addition to Sloan, and IBM other partners, such as Nuance Communications, were involved in bringing the above mentioned innovative project to life. Through its very powerful processing capabilities Watson can process thousands of data points to produce an optimal recommendation for the patient. This is an IT-enabled technological innovation involving both private and public sector entities that can save lives and improve the quality of treatment to patients.

The ballooning cost of healthcare, now nearly one fifth of GDP, combined with changes mandated by the Affordable Care Act and a more concentrated focus on patient outcomes is driving the need for innovative solutions informed by the results of big data analytics (Lamont, 2014). There is an inherent opportunity associated with the revelations brought about through big data analytics that cannot occur if hospitals, medical professional, and other important
stakeholders are working in isolation. There are many opportunities for collaborations leading to innovations across organizations and also across countries. The recent Ebola outbreak in West Africa saw participation from global public and private partners to reduce the spread of the disease. As more data is collected by domestic and international partners in the healthcare field, we increase the likelihood of addressing critical and acute issues related to patient care through collaborative innovations.

Cybersecurity
The field of cybersecurity focuses on protecting data, hardware, and software from unauthorized access gained through computer networks, simply put it is the protection of cyberspace. The Internet, the world’s most popular public network, has become infamous as the source of data breaches and systematic coordinated attacks on organizations’ data resources. Big data analytics in the area of cybersecurity processes data gathered from web sources, computer networks, social networks, deep forensics, and human portrait information such as psychological and political data (TechAmerica Foundation, 2012). Complex algorithms and techniques are used to mine these varied data streams. The ability to fight off cyber-attacks and cyber terrorism requires a concerted collaborative approach with global private and public sector entities involved.

Cyber-attacks threaten to expose sensitive internal data, weaken a company’s image, and reduce public trust in the ability of businesses to protect data. Outside of nuisance attacks designed to exhibit power and paralyze the functioning of organizations, data is the prized jewel that cyber criminals are in search of. Ironically, data is also the same component that is needed to thwart the efforts of cyber-criminals. In a recent report based on data collected from over 18,000 IT and data security practitioners, 61% of them expect to use big data analytics as part of their future strategy to thwart the growth of cyber-attacks (Ponemon Institute, 2013). This is good news since it highlights the potential value and diversity of opportunities harnessed though big data analytics. Effective cyber-security mechanisms require a proactive rather than a reactive approach along with participation from other external entities such as product vendors and business partners.

The opportunities to innovate through collaborations in the area of cybersecurity are extensive because of the many different types of issues that occur under the cybersecurity umbrella. A taxonomy of cybersecurity identifies four areas of operational cyber security risks: 1. actions of people eg. accidental or deliberate; 2. systems and technology failures eg. hardware malfunctions; 3. failed internal processes eg. status monitoring; and 4. external events eg. regulatory compliance (Cebula, Popeck, & Young, 2014). For each of these four categories, effective analysis of data can be used to reduce cybercrime and improve the IT security arsenal of organizations through innovations and collaborations.

Weather
The daily vagaries of atmospheric changes, ranging from snow to rain to sun is called the weather. The long term examination of these conditions is called climate. Weather and climate both affect different aspects of business events including retail decisions, energy needs, product acquisition, and distribution channels. Accurate prediction of weather conditions saves operational costs for both private and public sector firms. Both individuals and businesses lose millions of dollars annual due to weather related events. Publicly traded companies saw an 81% increase in stating that weather conditions had an effect on their business during the first quarter of 2014 as compared to the same period of the previous year (McGrath, 2014). The National Oceanic and Atmospheric Administration (NOAA), which serves as an umbrella to several other organizations such as The National Weather Service (NWS), uses big data
analytics to improve predictions on weather, climate and water events with a goal of enhancing outcomes to benefit the U. S. economy (TechAmerica Foundation, 2012).

Innovation in the area of big data and weather is not just about predicting the next hurricane; it is also about how these predictions affect government institutions and businesses. As weather and climate change predictions increase in accuracy and reliability, the businesses that are more agile will have better responses to customer and supplier demands. Companies such as DHL, Lowe’s, Home Depot, and other players across different sectors of the economy are all leveraging weather intelligence to improve the quality of services they provide (Nash, 2013). Vendors of seasonal products will fall behind the competition if they fail to recognize the value of weather data, particularly when niche companies are filling the void and making extensive data analytics available.

PROPOSED METHODOLOGY
This paper represents the first phase of a large scale project: big data for innovation and collaboration in the public sector. To further develop this project a three phased methodology is proposed. The first phase is identifying existing public entities that are currently harnessing the power of big data and data analytics. This will be a detailed content analysis of public sector data that is available.

Phase two will be a case study of one of these entities. This will involve in-depth interviews with the different partners involved. This case study will serve as the template for phase three. Phase three will be a longitudinal case study looking at the impact of the innovation(s) on the constituents, citizens, communities, etc. that is impacted by both the collaboration and the innovation. The long term impact of big data analytics will be a foundational element of the third phase of this study. A case has been made for recognizing big data itself as an IT-enabled innovation (Chae, 2015), so it will be important to demonstrate how this innovation is providing value in the public sector.

RISKS AND LIMITATIONS
Since this paper focuses on the convergence of three areas: IT-enabled innovations, collaborations, and big data, specific risks will be addressed for each of these topics. Firstly, IT-innovations, like any other type of innovation run the risk of not delivering on its promises. If the innovation whether it is a product or service fails, there can be significant loss to the partners involved in the collaborative venture. To reduce the likelihood of these occurrences careful upfront planning and pre-testing would be a useful practice. Secondly, a primary risk associated with collaborative projects is the development of adversarial relationships across the different partners (Hofer et al., 2012). Establishing relevant business contracts and effective team management would be needed to reduce conflicts across collaborators.

Lastly, there are risks associated with the use of big data and big data analytics. Companies using big data analytics are challenged to use their findings in an ethical and responsible way as not to do harm (Waterman & Bruening, 2014). There are also security and privacy risks associated with the collection of high volume data. Organizations may also see financial risks associated with the acquisition of technologies and expertise to develop the big data resource and resulting innovations.

FUTURE DIRECTIONS
Big data analytics is a growing phenomenon. To examine its value in the creation of IT-enabled innovation would require empirical testing to determine a causal relationship. More in-depth
studies of both successes and failures will provide a richer set of insights about this emerging area. Test the assertions and theoretical relevance of the material presented in this paper through case studies of collaborations can also be a future project. There are many opportunities in the specific context areas described above: healthcare, cybercrime, and weather for more in-depth studies.

CONCLUSION
This paper highlights the value of collaborations in the development of IT-enabled innovations. The main contribution of the paper identifies the use of big data and big data analytics as a bridge for new opportunities for collaborative innovations. The collaborative framework is supported by the resource dependency theory, which advocates for the added value of resources external to a firm. This paper also looked specifically at three content areas where big data for collaborative innovation is acutely relevant. A proposed methodology is presented to further advance this work. Lastly, risks and future directions are presented. The content of this paper provides value to both practitioners that are interested in IT-enabled innovations and the use of big data analytics, and researchers in the field that are exploring answers to emerging questions.

REFERENCES


