JIM

Journal of Innovation Management

The International Journal on Multidisciplinary Approaches on Innovation

Volume 4 Issue 1 | open-jim.org

ISSN 2183-0606

FEUP Edições

Journal of Innovation Management
JIM 4, 1 (2016) i-iv
HANDLE: http://hdl.handle.net/10216/83250

Table of Contents

Editorial

Letters

Articles

ISSN 2183-0606
http://www.open-jim.org

JIM 4, 1 (2016) 1-3
HANDLE: http://hdl.handle.net/10216/83250

Editorial

Coping with big: Does big data lead to 'bigger' innovation?

Marko Torkkeli[1], Anne-Laure Mention[2], João José Pinto Ferreira[3]

[1]Lappeenranta University of Technology, Finland; [2]Luxembourg Institute of Science and Technology, Visiting Professor & Deputy Director of Centre d'étude de la Performance des Entreprises University of Liège; [3] FEUP - Faculty of Engineering, University of Porto, INESC TEC - INESC Technology and Science, Portugal

marko.torkkeli@lut.fi, anne-laure.mention@list.lu, jjpf@fe.up.pt

This Spring Issue will discuss about big data and multiple aspects of its usability and applicability. Many of us have seen blockbuster movies *Back to the future (premiere in 1985)*, *The Terminator (1984)* or *Minority report (2002)*. The unifying element of the above mentioned movies is that manuscripts are introducing a superior competitive advantage factor. The protagonists create an advantage by having either real-time data (sometimes from the future) or all relevant (big and historical) data with enormous computing capacity over competitors. A bit after first two of those movies premiered, NASA scientists Cox and Ellsworth (1997) published an article where term 'big data' appeared first time (Press, 2014).

Intelligence needs to be topped up in a way to create advantage. Data has been there for a long time, in all forms and sizes. It is applied in almost single every business sector and it is getting faster in sense of usability. The data storage capacity has been exponentially increasing over time, but the usability of this wealth of data remains a critical issue.

This Issue aims to deepen our current understanding of the Big Data phenomenon, from multiple perspectives: definitional, conceptual, analytical, and empirical. Drivers, as well as obstacles, to the adoption and diffusion of big data are unearthed, providing grounds for managerial and policy implications. All papers adopt a comprehensive approach to big data, embracing both technological, processual, organizational and human aspects that are inherent to any type of innovation. The potential offered by big data to generate "bigger" novelties, and to create a wider, more sustainable impact from innovation, remains an essential question, to which this Issue partially answers.

In the first Letter of this Issue, Hanna discusses the drivers and barriers of e-commerce, which is portrayed as a techno-managerial innovation. Distinctive national features and peculiarities influence the speed of diffusion and adoption of e-commerce, at multiple levels: across industries and sectors, across firms within a nation, and within the boundaries of firms with differentiated levels of depth and breadth of extent and use. Hanna further elaborates on the role played by national policies aimed at promoting the adoption of e-commerce and highlights the importance of developing e-skills and increasing the general awareness and digital literacy of stakeholders.

ISSN 2183-0606
http://www.open-jim.org

In their Letter from Academia, Maglio and Lim depict how big data analytics can leverage the value offered by services, rendering them smarter. The Scholars further identify four types of smart service systems enabled by big data, namely smart customization and prevention, smart operations management, smart coaching and smart adaptation and risk management. A common feature of these smart service systems stems from the fact that these are the outcome of "embedding human knowledge and capabilities in technologies to serve human purposes for effective value co-creation", as described by the Scholars.

In "Data, Dialogue and Innovation: Opportunities and Challenges for Open Government in Canada", Roy revisits the Canadian experience, as a precursor of open data strategies. The Scholar details the tensions and the need for reforms addressing the various architectural facets of the public sector, embracing technological, administrative, political and social aspects. The paper also caters for avenues for facilitating systemic openness and collective innovation across sectors and government.

In their contribution, Segarra et al. explore how big data can be used as a lever by companies to boost their revenues and create value. The Authors develop and apply a set of tools to strategically analyze bid data capabilities and their potential for value creation, in a sequential manner. The empirical validation of the framework is performed using a single, in-depth case analysis of all operating segments of Amazon.com and how big data analytics contribute to customer satisfaction and sales in the retail industry.

In their literature review, Ylijoki and Porras reveal 17 definitions of big data, discuss the shortcomings of these current definitions and elaborate enhancements to the terminology, unveiling areas of further research.

In the fourth contribution of this Issue, Prescott tackles the critical question of the competitive advantage that firms can gain from big data. Anchored in the resource based view and dynamic capabilities literature stream, and using an interpretive approach, this exploratory research concentrates on the impact of digital data genesis on firm competitive advantage, explored through the lenses of improvements to product and service offerings in a single case study setting.

The final paper of this Issue focuses on the application of big data in the agricultural industry, illustrating how it can simultaneously foster economic and environmental benefits. The paper also highlights the influence of organizational collaboration as well as intellectual property contexts in the way big data can deliver its full potential in agriculture.

We wish you a stimulating journey in your reading of this issue of the Journal of Innovation Management.

Innovatively Yours,

Marko Torkkeli, Anne-Laure Mention, João José Pinto Ferreira
Editors

References

Cox, M. and Ellsworth, D. (1997) Application-controlled demand paging for out-of-core visualization, VIS '97 Proceedings of the 8th conference on Visualization '97, IEEE Computer Society Press Los Alamitos, CA, USA, ISBN:1-58113-011-2

Press, G. (2014), 12 Big Data Definitions: What's Yours? Accessed 30th April, 2016. http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#7be0d87321a9.

JIM 4, 1 (2016) 4-10
HANDLE: http://hdl.handle.net/10216/83250

E-commerce as a techno-managerial innovation ecosystem: Policy implications

Nagy K. Hanna

Wits University, World Bank
nagyhanna@comcast.net

Policy Letter

Innovation can be viewed as a adoption and dissemination of something new in a given context. E-commerce is thus an innovation when it is introduced to a new environment in an emerging market or when adopted by a new class of user industries. As a techno-managerial innovation, it requires business adaption, organizational learning, and supportive environment that could lead to wide diffusion and transformational impact. Several global forces drive the adoption of e-commerce such as global competition, trade liberalization, and increasingly, ICT advances and Internet diffusion. National factors, such as governance, education, and infrastructure, then shape and differentiate the speed of adoption across enterprises within a country, the breadth and depth of use within an enterprise, and ultimately the impact on the firm and the nation. Understanding the national environment, the policy, technological and infrastructural contexts, and the common drivers and barriers to adoption and effective use within firms should provide a guide to promoting e-commerce as a techno-managerial innovation, and realizing its full potential for the nation.

1 Potential for e-commerce

E-commerce is transforming trade. In the US, total online transactions grew from $3trillion (2006) to $5.4 trillion (in 2012), equivalent to a third of US GDP, with 88 percent of value of these transactions are business-to-business (B2B)[1]. Globally, business-to-customer (B2C) transactions alone are expected to grow from $1.5 trillion (2014) to $2.4 trillion in 2017. E-commerce is growing four times faster than world economy. A sizeable share of e-commerce is cross-border, estimated to average 16 percent among the six main markets—US, UK, Germany, Brazil, China, and Australia. The growth of cross border e-commerce will be specially fast in Asia Pacific, growing 3.7 times between 2011 and 2017, with the largest growth is in

[1] US Census Bureau E-Stats Report, 2012

ISSN 2183-0606
http://www.open-jim.org

China, where cross-border transactions are estimated at $160 billion in 2018. Worldwide, the cumulative Average Gross Rate (CAGR) of the cross-border e-commerce is 25 percent (Souminen, 2014).

Consumers also benefit from e-commerce, sometimes at the expense of firms, as e-commerce enables consumers to search for lower prices online and improves their bargaining position, and reduces the information advantage previously enjoyed by dealers and middlemen.

E-commerce is a huge export and growth opportunity for small and medium enterprises (SMEs), in particular. It increases export participation and broad-based trade from SMEs. When trade transactions are cross-border, e-commerce increases export diversification and expands the gains from trade. It gives consumers a wider variety of goods and services at lower cost. It gives online exporters a more staying power than offline exporters.

Since the commercial use of the Internet is less than two decades old, and SMEs tend to lag in adoption of new technologies, we may be seeing just the beginning of transformations that will be apparent in coming years with the diffusion and mastery of this innovation. While only large enterprises companies had the capital, global networks, and scale economies to enter and compete in world markets, e-commerce enables small businesses to leapfrog and gain visibility at low cost among global buyers and most distant markets.

2 Drivers and barriers to e-commerce

Many factors shape the adoption, pace of diffusion, and ultimate impact of new and versatile techno-managerial tools and practices such as e-commerce. Taking an ecosystem view can help in identifying these factors and dealing with the key stakeholders who influence adoption and diffusion. Stakeholders at the national and local levels would include relevant public agencies, the adopting enterprises, the customers, and various potential intermediary institutions such as business development associations, consulting firms, small business development services, research centers, and universities. This ecosystem can be quite complex, more than is often assumed by the designers of e-commerce diffusion programs. It span public policy, public platforms for e-commerce and electronic payments, enterprise learning and capability development, and consumer education.

The key global drivers of e-commerce are: global competition, cross-border trade liberalization, global production and distribution networks, the practices and strategies of MNCs, telecom deregulation, mobile and Internet diffusion, and the emergence of online marketplaces and e-commerce movement. International policies and institutions (WTO, ITU, World Bank) can also facilitate e-commerce via open rules and effective regulations for trade, investment, intellectual property, and telecommunications. Integration of countries into global production networks involves the adoption of e-commerce as a condition of participation. MNCs try to standardize their internal practices worldwide and push their suppliers and partners to align their processes and practices, including e-commerce, with those of the MNC. They rely on ICT, and particularly B2B e-commerce to improve coordination, cut inventory, shorten time-to-market, reduce errors, and become demand-driven.

The impact of these global forces is mediated by national contexts at the macro level: the economic, political, social, and technological context and the policy environment of the country. Also, it is influenced by firm and industrial sector organization and capabilities at the micro and meso levels. Understanding drivers and barriers at these levels is critical to devising policies and programs to diffuse e-commerce practices and to augment their transformational impact for business adopters.

At the national level, e-commerce drivers are economic factors, while barriers are more institutional, legal, and capability factors. National policy context influencing e-commerce comprises: openness to trade and investment, telecom and Internet regulation, security and ease of online payments, consumer protection, legal environment and enforcement of rule of law, online privacy and data security, intellectual property protection, customs and trade compliance costs (particularly for cross-border traders), and data protectionism, among others. National infrastructures and financial institutions also play a key role in the diffusion of e-commerce: access and quality (reliability, speed, cost) of Internet and communication infrastructure, transport and distribution systems, postal system, and financial services (eg, credit cards) and financial regulation.

Most visible in developed economies, the use of ICT for business is contributing to growing intensity of competition and environmental complexity, and in turn, greater complexity arising from globalization is leading to increased adoption of e-commerce.

Country economic structure and socio-economic institutions also matter for the use of e-commerce. Countries that are heavily dependent on international trade like China (with more than half of its GDP based on trade), Singapore and Taiwan (almost totally depended on foreign trade) are likely to be open to external influences such as e-commerce practices, and might also learn faster from foreign MNCs. Meantime, Brazil and Mexico’s large income inequality is likely to factor in retarding the use of e-commerce (Kraemer, 2006). A substantial proportion of firms in the developed countries have integrated their processes with suppliers and business partners, indicating substantial use of e-commerce fir supply chain coordination. Singapore proactively promoted e-commerce and business process integration to act as a production platform for foreign MNCs. The small scale of local markets in many developing economies may give global factors a more leading role as drivers of adoption than in large, inward-oriented countries, like Brazil.

Industry structure and sectoral differences also matter. Financial institutions were among the first to go global and to drive e-commerce practices via B2B, BPO, call centers, etc. Emerging patterns of e-commerce diffusion suggest that global networks drive B2B e-commerce, as in manufacturing, while local competition drives B2C e-commerce, as in wholesale/retail distribution. Put differently, B2B e-commerce supports upstream activities and tends to be more global; B2C supports downstream activities and to be more localized. Firms that operate more globally realize more benefits than firms that operate locally, as they are able to achieve economies of scale from their e-commerce investments, and their broader experience with ICT enables them to utilize e-commerce more effectively.

The overall diffusion and impact of e-commerce is likely to be a gradual and learning process, as in ICT use in business transformation in general. It must adapt to national institutional conditions. In China, for example, heavy investments in IT infrastructure and the Internet has been counterbalanced by relatively rigid institutional infrastructures and business processes. Hence, the gradual diffusion of e-commerce in China has been focused on the internationally oriented coastal regions and cities, where complementary organizational factors have been faster to emerge.

Government can also provide digital platforms and induce business use of e-commerce via e-government procurement systems, online government to business applications, trade-net, and other e-commerce promotion initiatives that create network effects and a critical mass of users. Government policy may require the use of Internet for government procurements, offering incentives for to help small enterprises go online. A national e-commerce movement can be fostered by business media, industry associations, venture capitalists, in collaboration with governments. It can be primed, transformed, or made more inclusive by developing local online marketplaces, such as the Ethiopia Commodity Exchange for agricultural markets, an ICT-based marketplace that serves the entire value chain: farmers, traders, processors, exporters, and consumers (UNCTAD, 2011). The evolutionary character of e-commerce points to the need for continual monitoring and evaluation of the use and impact of e-commerce over time, and to identify and overcome the barriers to its diffusion and effective use, especially for SMEs.

3 Challenges to adoption, diffusion, and impact for SMEs.

Surveys suggest the strongest drivers of e-commerce use are the desire to expand markets, to improve coordination with customers and suppliers, and to entre new markets (kreamer, et al, 2006). Linkages to MNCs can be both a driver and enabler for SMEs. Other factors include demand by local consumers, and the cost and quality of access to the Internet and online marketplaces. The biggest barriers to adoption are concerns over privacy and security of data (highest in the financial sector) and inadequate legal protection for Internet purchases.

The impact of e-commerce on firms and economies is significantly dependent on their managerial and technical capabilities. It is a function of the spread and intensity of use of ICT within the enterprise, and within its entire value chain, from suppliers to business partners, to customers. E-commerce is not just about sales (as in contrast to Amazon and eBay), and online selling requires other activities as complements, if not prerequisites. Firms with a higher and more strategic technology use tend to realize greater value from e-commerce investments.

Managers play important roles in promoting greater depth and breadth in ICT use across the value chain: marketing, sales, customer services, procurement, production, information sharing, and value chain coordination. Technical and managerial competencies for effective use of ICT innovations are mainly acquired through learning by doing. As ICT applications become a strategic necessity, such competencies become a major differentiator of e-commerce adopters at the firm, sector, and national levels.

4 Mobile for e-Commerce

Mobile phones, being the main communication tools for small entrepreneurs in developing countries, have great potential for e-business applications. Mobile telephony is also likely to be the primary tool for connecting the vast majority of low-income population to business and information society, at least in short to medium term. For example, SMEs that export agricultural products may be alerted to business opportunities and receive timely price information for their products. Mobile commerce, mobile banking and payments and mobile content are spreading in most developing countries. The potential is great, provided there is an enabling regulatory environment. In many countries, prepaid mobile services are used to provide mobile public payphones, and this improves accessibility in rural areas. As mobile handsets grow in sophistication and add new functionalities, such as digital photography and multimedia messaging and other utilities, they will provide a gateway to digital literacy.

The Philippines presents an interesting example of leveraging mobile phones for e-commerce by farmers. With very low penetration of Internet, the government tapped into the culture of SMS to provide two mobile applications, one, for farmers to post prices of their products, and the other, for mobile users to compare price of the top most traded products in a province. The program works with cooperatives to provide a level playing field and by giving players access to a common source of reliable and on-line market prices, are helping farmers maximize their selling prices.

For developed countries, the impact of broadband on business has been fairly documented in terms of reducing costs, increasing revenues, and expanding markets—but the evidence of broadband impact on business is still fairly thin among developing countries, even at the firm level. The experience of developed countries is therefore critical to guide broadband and e-business diffusion and impact in developing countries. A study of broadband deployment in business in several Latin American countries showed that deployment was associated with considerable improvements in business organization, including knowledge diffusion within organizations, and speed of business reengineering and network integration (Momentum Research Group, 2005). Broadband may also help firms in specializing in core activities and outsource the rest. Broadband may also help in building distinctive capabilities by allocating activities more efficiently between workers tackling complex and creative tasks and more transactional workers (Johnson, Manyika and Yee, 2005).

The highest productivity gains appear in firms that commit to integrate broadband, and IT in general, with reengineered business processes. Organizations that align their investments in network infrastructure, network-based applications, business processes and organizational behavior experience greater increases in business outcomes than organizations that disproportionately focus on one or more of these elements.

5 Conclusion

E-commerce holds the potential of becoming a major techno-economic innovation and an entry towards broader export, innovation and business transformation. It can lead to innovation in business models. It can provide a platform for innovations in business processes, relationships, products, and services. Some countries have therefore initiated e-commerce diffusion programs to help early adopters and SMEs, and increase the scale and impact of this innovation in selected sectors or economy-wide.

National policies to promote the adoption of e-commerce as a techno-managerial innovation should include: effective logistics and delivery infrastructure, trade facilitation system, digital infrastructure and platform development, affordable and reliable access to this infrastructure, secure payment solutions for online purchases, and enforcement of e-commerce laws and regulations. These policies should be complemented by developing e-commerce skills among small businesses, promoting digital entrepreneurship and innovation management, developing government e-procurement and e-trade networks to incentivize enterprises to adopt online transactions, and raising digital literacy and general awareness of all stakeholders about e-commerce.

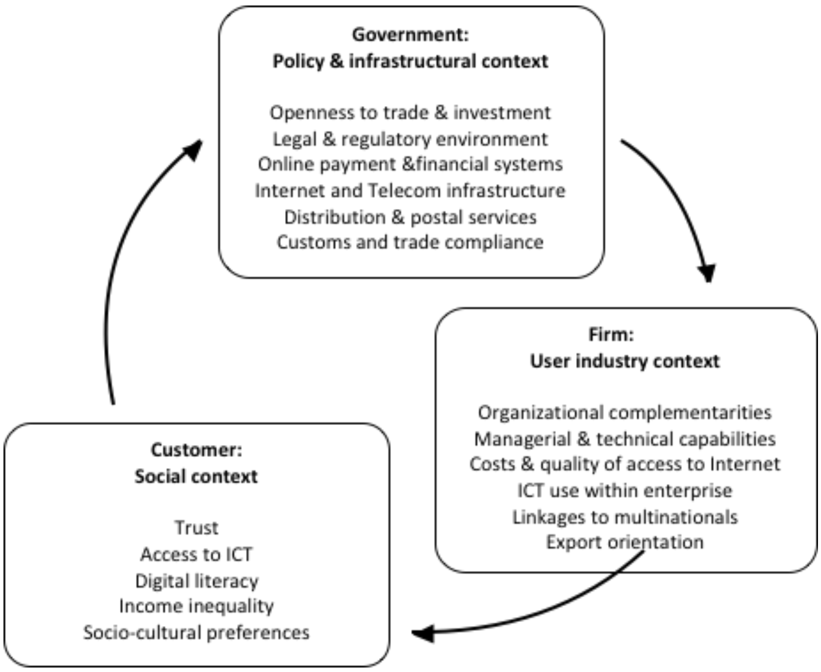


Fig 1. Factors influencing the e-commerce ecosystem

Some promising lessons are emerging. An ecosystem approach can be helpful in designing these diffusion programs and in mobilizing the relevant stakeholders to fund and sustain them. E-commerce diffusion programs should focus on actual usage and payoffs from e-commerce, rather than focusing solely on ICT investment. They should be tailored to country context, to address policy and institutional factors, such as payment systems, privacy and data security, legal protection for online transactions, customs and trade compliance procedures, and Internet governance. They should address issues of access to Internet and broadband, and the diverse forms of digital divide. They may also address those infrastructural issues (eg, postal, transport, logistics, electricity) that impact SMEs' e-commerce most adversely. And perhaps most important to SMEs, diffusion programs should promote capability development and advisory services to SMEs to leverage and integrate e-commerce adoption into their business strategies and practices. Such programs may be sponsored by central government (ministry of industry, trade, small business), cities and local governments, trade and business associations, universities, or combined sources via partnerships.

Figure 1 sums up the various factors that influence the adoption and use of e-commerce, and overall e-commerce ecosystem of a country; this ecosystem view should guide the design of e-commerce diffusion programs.

JIM 4, 1 (2016) 11-21
HANDLE: http://hdl.handle.net/10216/83250

Innovation and Big Data in Smart Service Systems

Paul P. Maglio[1], Chie-Hyeon Lim[1,2]

[1] University of California, Merced, USA
{pmaglio,clim28}@ucmerced.edu

[2] Pohang University of Science and Technology, Republic of Korea
arachon@postech.ac.kr

Letter from Academia

As traditionally measured, services, which include everything from transportation to retail to healthcare to entertainment to hospitality and more, account for most economic activity. Taking a more modern view, we define *service* as value creation that occurs within systems of interacting economic actors. Service systems have been getting smarter over time, as big data analytics have been used to generate information and automate operations that create ever more value for people in the service systems. In this short letter, we describe some of our perspective on the use of big data analytics in smart service systems, suggesting one framework for thinking about big data in this context and outlining a set of research issues.

Keywords. Smart Service System, Big Data, Innovation.

1 Introduction

Service is everywhere. Taking a traditional view, services include transportation, retail, healthcare, consulting, outsourcing, entertainment, hospitality, and much more, accounting now for more than 80% of economic activity in the US and other industrialized countries (Spohrer and Maglio, 2008). Taking a more modern view, we think service includes all economic activity in which individuals, organizations, and technologies work together, applying specialized competences and capabilities to make all actors better off together than they are separately (Spohrer and Maglio, 2010; Vargo et al., 2008). On this view, service—also known as *value co-creation*—underlies all economic exchange (Vargo and Lusch, 2004). The key to effective service lies in arranging the capabilities among multiple actors or stakeholders so they can create the most value together (Maglio et al., 2009).

Specifically, we define s*ervice systems* as configurations of people, information, organizations, and technologies that operate together for mutual benefit (Maglio et al., 2009). Service systems differ from other types of sociotechnical systems in that they depend on entities sharing capabilities to increase mutual value. In this way, change in service systems results from rearranging where and how system capabilities are located (Breidbach and Maglio, 2015), often transforming the way systems work by embedding sophisticated capabilities into technologies, such as self-service

ISSN 2183-0606
http://www.open-jim.org

technologies to generate more overall value (e.g., Campbell et al., 2011). In service systems, value creation is difficult to measure and anticipate: service systems depend not only on people, information, organizations, and technologies, but also on interactions among these, which has emergent consequences. We think a key problem in understanding service systems lies in understanding the critical role of people and their relationships with other components, such as information and technology (Maglio et al., 2015).

Service innovation results from transformations of existing service systems or establishment new service systems. It can often be achieved through offering a new core benefit or developing a new way to deliver a core benefit (Berry et al., 2006). Service innovation takes multiple forms in multiple industries (Miles, 2008), and is evident in traditional service industries and also manufacturing industries (Baines et al., 2009). It sometimes results from use of specific methods (e.g., Bitnet et al., 2008; Bettencourt, 2010) and sometimes requires organizational and cultural changes (e.g., Rothenberg, 2007; Reinartz and Ulaga, 2008). Sustainable service innovation requires continuous iterations of new service development, service operations, and service improvement (Kim et al., 2009). Stimulating service innovation is a timely research topic that has a large gap between importance and current knowledge (Ostrom et al., 2015) and *service science* emerged with the aim to achieve service innovation scientifically and systematically (Maglio and Spohrer, 2008).

A *smart service system* is "a service system capable of learning, dynamic adaptation, and decision making based upon data received, transmitted, and/or processed to improve its response to a future situation" (Medina-Borja, 2015). *Big data* typically describes large and complex sets of data representing digital traces of human activities (Manyika et al., 2011), and may be defined in terms of scale or volume (Zikopoulos, et al., 2012), analysis methods (Chen, et al. 2012), or impact on organizations (McAfee and Brynjolfsson, 2012). *Big data analytics* involves cognitive and computational processes that uncover patterns in big data (George et al., 2015). Smart service system innovation can take a great advantage of big data analytics (Medina-Borja, 2015) and recent studies show examples of the contribution of big data analytics to smart service innovation (e.g., Lim et al., 2015; Opresnik and Taisch, 2015).

Smart service systems are a kind of *human-centered service system,* meaning that knowledge, capabilities, and value are all determined by the people in the system (Maglio et al., 2015). Big data analytics can create human value in smart service systems in many ways; for instance, for customers, customer data may get converted into information that is useful in customer value creation processes in smart service systems (Saarijärvi, 2011), and for firms, customer data can be analyzed to understand patterns of customer behavior (Boyd and Crawford, 2011) to learn why customers make certain decisions or behave in certain ways (Huang and Rust, 2013) and to design new services or improve existing services (Lim et al., 2015). Use of big data can foster a mutually beneficial relationship between a firm, its customers, and possibly society in smart service systems (Kumar et al., 2013). Thus, a key problem in innovating smart service systems lies in taking advantage of big data analytics to create *human value.*

In this brief letter, we describe how big data analytics can help foster new service innovations, creating smart service systems by embedding human knowledge and capabilities in technologies to serve human purposes for effective value co-creation.

2 Using Big Data in Smart Service Systems

Cities around the world collect massive amounts of data related to urban living, and these data contribute to the production of useful information for citizens, visitors, city officials, and local employees (Caragliu et al., 2011). Automobile manufacturers analyze vehicle condition and driving data collected from onboard devices via telematics, and they provide various types of information to drivers, for instance, about fuel efficiency, safety, consumption, and navigation (Lim et al., 2015). Insurance companies collect patient data and provide healthcare-related information to patients to improve healthcare safety, reduce cost, and develop sustainable relationships (OECD, 2013). These are just a few examples of smart service systems enabled by big data analytics.

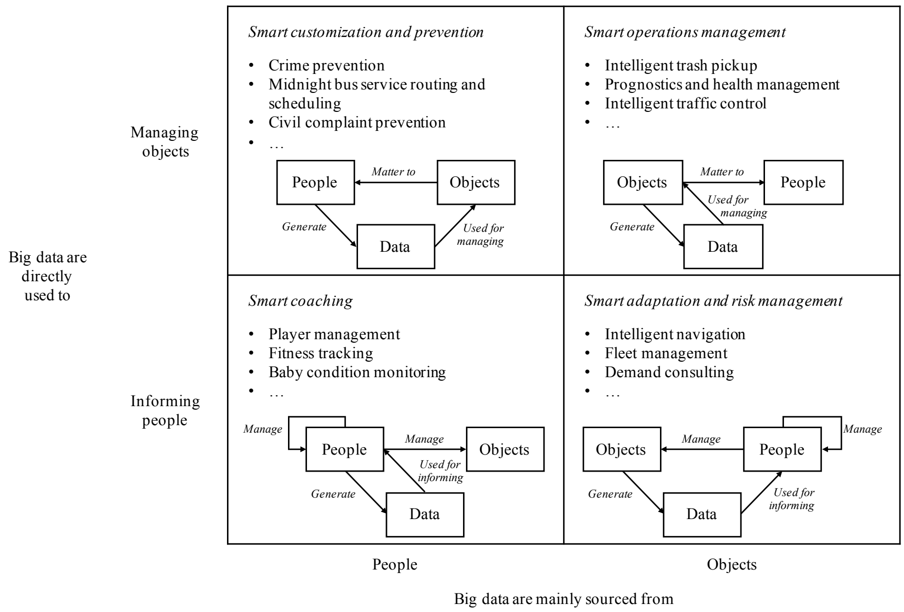


Fig. 1. Four ways to use big data in smart service systems.

We have studied big data analytics in smart service systems from many cases available in journal articles, books, technical reports, Internet news, and blogs. We found it useful to organize the cases by the source of data, either collected mainly from people or from objects, and by the use of data, either informing people to help people and manage objects or managing objects directly (see Figure 1). Our two-by-two matrix shows four categories of innovation in human-centered smart service systems. We describe each category in turn.

First, smart operations management (shown in the upper right of Figure 1) relies on data from objects (e.g. product condition, environment, and event log data) to manage objects (e.g. vehicles, infrastructure, and city administration). Cases in this category aim to improve operational processes of certain service systems by controlling objects within the system efficiently and effectively based on enhanced understanding of them through analysis of data. Representative examples include intelligent trash pickup, which collects data from trash bins using Radio Frequency Identification tags and schedules trash collection location and time (Purohit and Bothale, 2012); prognostics and health management, which uses the heavy equipment condition data to cope with potential product breakdowns and maximize product availability for stakeholders (Lee et al., 2014); and intelligent traffic control in Singapore, which collects data from roads and taxis to anticipate future traffic and control traffic lights for the citizens and visitors (Lee, 2013).

Second, smart customization and prevention (shown in the upper left of Figure 1) relies on data from people (e.g. human health, behavioral, and purchase history data) to manage objects. Cases in this category aim to understand problems and needs of people in certain service systems through analysis of data and then to customize system operation to specific needs or prevent problems. Numerous cases of smart cities (Caragliu et al., 2011; Kitchin, 2014) correspond to this category: Representative examples include crime prevention in San Francisco, which used crime records to predict future crime locations, patrol the locations, and prevent potential crimes (Lee, 2013); midnight bus service routing and scheduling of Seoul, which analyzed mobile call records and taxi-use data from the citizens and visitors to identify where they were and how they moved in the city late at night, enabling optimization of bus routes and schedules based on late-night demand (KLID, 2014); and civil complaint prevention of Busan, which analyzed 10 years of civil complaint records in a district and identified strategies to manage illegal parking, dust scattering from construction, streetlights, and other sources of citizen complaints (KLID, 2014).

Third, smart coaching (shown in the lower left of Figure 1) relies on data from people to help people (e.g., players, moms, teachers, and company employees) manage themselves and others (e.g. babies, students, and company customers). Cases in this category aim to provide evidence-based coaching or management based on enhanced understanding of human behaviors and context: Representative examples include player management, which involves data-driven evaluation and improvement of athletes, such as baseball players, golfers, and swimmers, using detailed data of their actions over time (Jung et al., 2010); fitness tracking using smart bands or other wearable devices, which collect data of daily life, such as behavior, health, and food menu data, to help people achieve specific fitness-related outcomes, such as walking 10,000 steps (Takacs et al., 2014); and baby condition monitoring, which collects data from babies and environment to give parents their status and predicted behaviors (http://www.sproutling.com/).

Fourth, smart adaptation and risk management (shown in the lower right of Figure 1) relies on data collected from objects to help people. Cases in this category aim to analyze the data about objects that affect specific human goals to help them adapt the objects and manage risks: Examples include intelligent navigation in Milan, which analyzes data on factors affecting traffic flow, such as real-time traffic situations, accidents, weather, construction, and event data from sensors placed all over the city, to provide navigation information to the citizens and visitors (Lee, 2013); fleet management, which collects data from transportation processes of trucks and uses the data to improve efficiency and productivity of the processes for drivers (Volvo, 2009); and demand consulting, which analyzes card usage records, types and revenue of retailers, and government data to assess which new types of businesses might be needed in any specific area for decision making by retail entrepreneurs (MISP, 2014).

People pay for goods and services to get jobs done, whether farming, driving, dating, or other business (Ulwick, 2005). Innovation is a perennial mission of firms, enabling jobs to get done better than before (Bettencourt, 2010). Our case analysis shows that big data analytics contributes mainly to the creation of useful knowledge to manage objects or inform people so that people can do their jobs (about either people or objects) better. In particular, as shown in Figure 1, big data analytics can create value for people *regardless of data source*. Our two-by-two matrix is a framework for understanding similarities and differences among the cases, helping us view big data analytics with human-centered and service-centered thinking. Exploration and exploitation of big data should be human-driven rather than technology-driven, and a prerequisite to big data use in a smart service system context is identification of the right information to generate human value. Our four categories can be applied to help innovate in any human-centered service system, such as in cities, health care, information and communication technology, education, and manufacturing.

3 Research Issues for Smart Service System Innovation

Smart service systems are everywhere. Yet relatively little is known about them and about innovation for them. Ostrom et al. (2015) identified and evaluated twelve service research priorities through roundtable discussions and surveys with service researchers around the world. Although all the priorities and related topics were deemed important, the study concluded that “using big data to advance service” had the largest gap between importance and current knowledge of the field. We also see a number of research issues for smart service system innovation among the fields of big data analytics, service science, and innovation management. These issues relate to design, evaluation, description, and automation of smart service systems.

First, consider service design. Service innovation depends on new or improved service ideas, concepts, processes, business models, and more. We see service design as the bridge that connects an opportunity or idea with full business development. Developing new methods for service design is a research area aimed at creating actual service value (Zomerdijk and Voss, 2010), and various service design methods exist, such as the TRIZ-based service design (Chai et al., 2005), multilevel service design (Patrício et al., 2011), and casebook-based service design (Kim et al., 2012). Although such methods may aid in service design in multiple contexts, none specifically address the design of smart service systems that rely on big data analytics. Another limitation of the current service design literature is that no method seems to exist for designing services starting from big data about customers. Such data-driven methods for service design could take the guesswork out of the customer understanding and enable the efficient design of information content for customers. The four categories proposed in this letter can be used as archetypes (i.e., design models) for the design of smart service systems. A data-driven service design method should consider the human-data-object relationship shown in Figure 1.

Second, consider service evaluation. Smart service system innovations are expected to develop in multiple industries with the rapid advancement of technologies for collecting data from people and objects (Medina-Borja, 2015). Developing scales for evaluating the quality of smart service systems from the perspective of human-centered big data use is another important research issue to improve emerging smart service systems. The perceptions of smart service system quality are not well-known, compared to those for traditional service (Parasuraman et al., 1988), electronic service (Ladhari, 2010), and mobile service (Akter et al., 2013). Novel quality scales may be required if customers perceive smart service systems differently from other service types. Such quality scales should consider the human-data-object relationship within smart service systems. Scale development itself may be facilitated by the use of big data that indicate customers' perceptions on specific smart services.

Third, consider service description. Using big data analytics effectively in service design and evaluation requires having a model that describes the service and the data together. Existing service description methods describe services from the perspective of service delivery processes, employee visibility, use of technology, and interactions among players involved (e.g., Bitner et al., 2008; Lim et al., 2012; Sampson, 2012; Teixeira et al., 2012; Lim and Kim, 2014), but cannot describe services based on a language of data. A generic model to describe a service with a set of variables that can be measured with a set of data, such as customer and context variables, could facilitate the design of smart service systems, data-driven service design, and evaluation of smart service systems; such a model would be useful in integrating different data analytics results (e.g., a statistical relationship between specific variables) to design and evaluate services as well as planning data analytics from a service-oriented perspective. Research on data-driven service description may involve analysis of existing service cases, such as the analysis shown in Figure 1.

Finally, consider service automation. Human actions in service systems can be categorized as informational, physical, and interpersonal actions (Apte and Mason, 1995), and the automation of these actions has evolved from automated teller machines of banking services to warehouse robots of shipping services and has made these services smarter. Big data analytics can contribute to automation of information actions: Decision-making and information exchange tasks in service systems can be substituted by big data analytics to create cognitive systems that can act on their own or provide support for people (Kelly and Hamm, 2013). The examples in Figure 1 mainly show the contribution of big data analytics to the automation of information actions in service systems. Physical actions, such as physical tasks and operations in service systems can be automated, for instance, through the use of robots and other control systems that substitute mechanical work for human work, such as rehabilitation and factory robots (e.g., Agarwal et al., 2015). Big data analytics can contribute to algorithms for controlling physical actions (e.g., Yun et al., 2014). Recent robot examples that interact with humans directly and emotionally (e.g., https://www.jibo.com) show that automation of interpersonal actions also can take a great advantage from big data analytics for recognition and prediction of emotion. Though the inseparable relationship between big data analytics and automation-based smart service system innovation has been discussed recently (e.g., Porter and

Heppelmann, 2014), little is known about such innovation. The human-data-object relationship in Figure 1 may prove useful in such research.

4 Concluding Remarks

Innovation in human-centered smart service systems will be enhanced by a shared vocabulary among disciplines, which is one of the main goals for the development of a unified service science (Spohrer et al., 2007; Maglio et al., 2015). Researchers have used different perspectives on human-centered smart service systems, such as data collection, analytics, and information delivery, but relatively little is known about how different perspectives can work together to create value with data. Such a framework could help build a theoretical background of human-centered smart service systems, stimulate applications of such services both in academia and by businesses, and foster human-centered service value creation with big data. For example, we see the potential for a new framework for *service-oriented data analytics* (SODA), a standard approach to collecting, transforming, and analyzing data to discover useful information for a service system. In the context of human-centered smart service systems, we can even see *cognition as a service* as a composable piece in larger systems (Spohrer and Banavar, 2015). Innovation in human-centered smart service systems depends critically on big data collection and analytics that serve specific human purposes and enable creation of specific human value.

5 References

Agarwal, P., Fox, J., Yun, Y., O'Malley, M. K. & Deshpande, A. D. (2015). An index finger exoskeleton with series elastic actuation for rehabilitation: Design, control and performance characterization. *International Journal of Robotics Research*, Accepted for Publication.

Akter, S., D'Ambra, J. & Ray, P. (2013). Development and validation of an instrument to measure user perceived service quality of mHealth. *Information & Management*, 50(4), 181-195.

Apte, U. M. & Mason, R. O. (1995). Global disaggregation of information-intensive services. *Management science*, 41(7), 1250-1262.

Baines, T. S., Lightfoot, H. W., Benedettini, O. & Kay, J. M. (2009). The servitization of manufacturing: A review of literature and reflection on future challenges. *Journal of Manufacturing Technology Management*, 20(5), 547-567.

Berry, L. L., Shankar, V., Parish, J. T., Cadwallader, S. & Dotzel, T. (2006). Creating new markets through service innovation. *MIT Sloan Management Review*, 47(2), 56-63.

Bettencourt, L. (2010). *Service innovation: How to go from customer needs to breakthrough services*. London: McGraw-Hill.

Bitner, M. J., Ostrom, A. & Morgan, F. (2008). Service blueprinting: A practical technique for service innovation. *California Management Review*, 50, 66–94.

Boyd, D. & Crawford, K. (2011). Six provocations for big data. Accessed 5th Nov 2015. http://ssrn.com/paper=1926431.

Breidbach, C. F. & Maglio, P. P. (2015). A service science perspective on the role of ICT in service innovation. *European Conference on Information Systems (ECIS),* Muenster, Germany.

Campbell, C. S., Maglio, P. P. & Davis, M. M. (2011). From self-service to super-service: How to shift the boundary between customer and provider. *Information Systems and eBusiness Management,* 9(2) 173-191.

Caragliu, A., Del Bo, C. & Nijkamp, P. (2011). Smart cities in Europe. *Journal of urban technology*, 18(2), 65-82.

Chai, K. H., Zhang, J. & Tan, K. C. (2005). A TRIZ-based method for new service design. *Journal of Service Research*, 8(1), 48-66.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012) Business intelligence and analytics: From big data to big insights", *MIS Quarterly,* 46(4), 1-24.

George, G., Haas M. & Pentland, A. (2015). Big data and management, *Academy of Management Journal*, 57(2), 231-326.

Huang, M. & Rust, R. (2013). IT-related service: A multidisciplinary perspective. *Journal of Service Research*, 16(3), 251-258.

Kelly III, J. E. & Hamm, S. (2013). *Smart machines: IBM's Watson and the era of cognitive computing*. New York: Columbia University Press.

Jung, J., Park, H., Kang, S., Lee, S. & Hahn, M. (2010). Measurement of initial motion of a flying golf ball with multi-exposure images for screen-golf. *Consumer Electronics, IEEE Transactions on*, 56(2), 516-523.

Kim, K. J., Hong, Y. S., Shin, D. M., Cho, N. W., Jung, J. Y., Lee, Y. H., Park, H. Y., Hong, J. W., Kang, W. M. & Shin, H. Y. (2009). Services innovation: Research framework and research issues. *Journal of Korean Institute of Industrial Engineers*, 35(4), 226-247.

Kim, K. J., Lim, C. H., Lee, D. H., Lee, J., Hong, Y. S. & Park, K. (2012). A concept generation support system for product-service system development. *Service Science*, 4(4), 349-364.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14.

KLID (2014). *A study on developing strategies for big data application in private and public sector through case-study analysis*. Korea Local Information Research & Development Institute.

Kumar, V., Chattaraman, V., Neghina, C., Skiera, B., Aksoy, L., Buoye, A. & Henseler, J. (2013). Data-driven services marketing in a connected world. *Journal of Service Management*, 24(3), 330-352.

Ladhari, R. (2010). Developing e-service quality scales: A literature review. *Journal of Retailing and Consumer Services*, 17(6), 464-477.

Lee, J.,Kao,H.A. & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP*, 16, 3-8.

Lee, W. (2013). Smart city management using big data. BDI Focus, BDI.

Lim, C. H. & Kim, K. J. (2014). Information service slueprint: A service blueprinting framework for information-intensive services. *Service Science*, 6(4), 296-312.

Lim, C. H., Kim, K. J., Hong, Y. S. & Park, K. (2012). PSS Board: a structured tool for product-service system process visualization. *Journal of Cleaner Production*, 37, 42-53.

Lim, C. H., Kim, M. J., Heo, J. Y. & Kim, K. J. (2015). Design of informatics-based services in manufacturing industries: case studies using large vehicle-related databases. *Journal of Intelligent Manufacturing*, Online First. Accessed 5th Nov 2015. http://link.springer.com/article/10.1007/s10845-015-1123-8

Maglio, P. P. & Spohrer, J. (2008). Fundamentals of service science. *Journal of the Academy of Marketing Science,* 36(1), 18-20.

Maglio, P. P., Vargo, S. L., Caswell, N. & Spohrer, J. (2009). The service system is the basic abstraction of service science. *Information Systems and e-business Management, 7,* 395-406.

Maglio, P. P., Kwan, S. J. & Spohrer, J. (2015). Toward a research agenda for human-centered service system innovation. (Commentary) *Service Science*, 7(1), 1-10.

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, San Francisco, CA.

McAfee, A. & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 61-68.

Medina-Borja, A. (2015). Smart things as service providers: A call for convergence of disciplines to build a research agenda for the service systems of the future. (Editorial) *Service Science,* 7(1), ii-v.

Miles, I. (2008). Patterns of innovation in service industries. *IBM Systems Journal*, 47(1), 115-128.

MSIP (2014). *Case book of big data applications*. Ministry of Science. ICT and Future Planning, Republic of Korea.

OECD (2013). *ICTs and the health sector: Towards smarter health andwWellness models.* OECD Publishing. Accessed 5th Nov 2015. http://dx.doi.org/10.1787/9789264202863-en.

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

Ostrom, A. L., Parasuraman, A., Bowen, D. E., Patrício, L., Voss, C. A. & Lemon, K. (2015). Service research priorities in a rapidly changing context. *Journal of Service Research*, 18(2), 127-159.

Parasuraman, A., Zeithaml, V. A. & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.

Patrício, L., Fisk, R. P., Falcão e Cunha, J. & Constantine, L. (2011). Multilevel service design: From customer value constellation to service experience blueprinting. *Journal of Service Research*, 14, 180-200.

Porter, M. E. & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, 92(11), 11-64.

Purohit, S. S. & Bothale, V. M. (2011). RFID based solid waste collection process. *Proceedings of Recent Advances in Intelligent Computational Systems (RAICS)*, 457-460.

Reinartz, W. & Ulaga, W. (2008). How to sell services more profitably. *Harvard Business Review*, 86(5), 90.

Rothenberg, S. (2007). Sustainability through servicizing. *MIT Sloan Management Review*, 48(2).

Saarijarvi, H. (2011). *Customer value co-creation through reverse use of customer data*. Tampere University Press.

Sampson, S. E. (2012). Visualizing service operations. *Journal of Service Research,* 15, 182-198

Spohrer, J. & Banavar, G. (2015). Cognition as a service. *AI Magazine,* 36(4).

Spohrer, J., Maglio, P. P., Bailey, J. & Gruhl, D. (2007). Steps toward a science of service systems. *Computer*, *40*, 71-77.

Spohrer, J. & Maglio, P. P. (2008). The emergence of service science: Toward systematic service innovations to accelerate co-creation of value. *Production and Operations Management*, *17*(3), 1-9.

Spohrer, J. & Maglio, P. P. (2010). Toward a science of service systems: Value and symbols. In P. P. Maglio, C. A. Kieliszewski & J. C. Spohrer (Eds.), *Handbook of service science*. New York: Springer.

Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C. & Hunt, M. A. (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17(5), 496-500.

Teixeira, J., Patricio, L., Nunes, N. J., Nobrega, L., Fisk, R. P. & Constantine, L. (2012). Customer experience modeling: from customer experience to service design. *Journal of Service Management*, 23(3), 362-376.

Ulwick, A. W. (2005). *What customers want: using outcome-driven innovation to create breakthrough products and services*. New York: McGraw-Hill.

Vargo, S. L. & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68,1-17.

Vargo, S. L., Maglio, P. P., and Akaka, M. A. (2008). On value and value co-creation: A service systems and service logic perspective. *European Management Journal*, *26*(3), 145-152.

Volvo (2009). ITS4mobility technical description. Volvo. Accessed 5th Nov 2015. http://www.volvobuses.com/SiteCollectionDocuments/VBC/Global%20-%20ILF/Downloads/ITS4mobility_Technical_description.pdf.

Yun, Y., Kim, H. C., Shin, S. Y., Lee, J., Deshpande, A. D. & Kim, C. (2014). Statistical method for prediction of gait kinematics with Gaussian process regression. *Journal of Biomechanics*, 47(1), 186-192.

Zomerdijk, L. G. & Voss, C. A. (2010). Service design for experience-centric services. *Journal of Service Research*, 13(1), 67-82.

JIM 4, 1 (2016) 22-38
HANDLE: http://hdl.handle.net/10216/83250

Data, Dialogue, and Innovation: Opportunities and Challenges for "Open Government" in Canada

Jeffrey Roy

School of Public Administration, Dalhousie University
roy@dal.ca

Abstract. In a rapidly evolving online environment where the inter-relationship between information and innovation is evolving from primarily closed and inward structures to much more open and networked governance arrangements, the public sector faces growing pressures and new opportunities to reform and adapt. Open data and big data are now widely embraced initiatives to spur innovation both inside of and outside of the public sector. Their capacity to foster innovation is nonetheless shaped by critical tensions between traditional government structures and culture on the one hand and more open and participative notions of governance on the other hand. Within such a context, this article examines the current Government of Canada Open Government Action Plan and its three main dimensions: information, data, and dialogue. The analysis reveals that despite some progress in the realm of open data, information and dialogue are constrained by the aforementioned tensions and the need for wider reforms to various architectural facets of the public sector – administratively, technologically, politically, and socially. Across each of these layers, we consider the sorts of wider reforms required in order to facilitate systemic innovation within the government and across sectors.

Keywords. Open Government, Innovation, Governance, Digital Data, Information, Dialogue, Engagement, Transparency.

1 Introduction

In a rapidly evolving online environment where the inter-relationship between information and innovation is evolving from primarily closed and inward structures to much more open and networked governance arrangements, the public sector faces growing pressures and new opportunities to reform and adapt. Accordingly, 'open government' has become a new mantra across much of the democratic world, arguably stemming from President Obama's 2009 inaugural Presidential Directive on Openness and the various US federal government initiatives that ensued. The Organisation for Economic Co-operation and Development (OECD) describes open government as 'The transparency of government actions, the accessibility of government services and information, and the responsiveness of government to new ideas, demands and needs' (Ubaldi, 2013). More than sixty countries have now subscribed to the Open Government Partnership, a global declaration of principles meant to facilitate political buy-in and a commitment to specific actions plans.

While one important aim of open government is heightened transparency in improving democratic awareness, oversight an involvement, an equally important objective is to enable innovation to occur both within the public sector and across the

ISSN 2183-0606
http://www.open-jim.org

private sector and civil society at large. Central to this direction is the notion of open data – releasing raw data sets previously stored internally and viewed as proprietary, in order to spur their shared utilization in ways that drive collective learning and create new forms of both public and private value. There are expanding calls for governments to leverage and embrace data-driven innovation:

> *Open up public data, particularly publicly funded data. Clearly, the economic potential of these and other important public datasets can only be fully exploited if the most innovative and creative entrepreneurs have full access to data;*
>
> *Find creative ways of tackling privacy, security, and intellectual property concerns while allowing the exploitation of the full economic potential of big data (p.8, Sousa, 2013).*

This accompanying lens of big data - defined by the OECD as the tremendous expansion of volume, variety, and velocity of data flows across all sectors, both reinforce open government tendencies while also contradicting them. This is so since big data encompasses and builds upon open data in one sense: nevertheless, the management and usage of such vast and largely automated data systems are often invisible and/or explicitly shielded from public purview due to a variety of reasons (chief among them being national security). Traditional government, moreover, for reasons explained more fully through this article, features contradictory stances toward transparency and secrecy even prior to the overlay of data-driven reforms (Roy, 2013a/b). The resulting tensions surrounding information management and data openness lie at the heart of efforts to pursue innovation via open government.

Such tensions are compounded by a third and equally complex dimension of open government - namely public engagement and dialogue (Lee and Kwak, 2011; Roy, 2013a). As an illustration, the participatory spirit of the Government of Canada Open Government Action Plan (a primary focus of this article) has been recognized by the Information Commissioner of the Australian State of New South Wales:

> *Since its introduction in 2011, 'Canada's Action Plan on Open Government' has been refined in response to greater recognition of social, economic and technological developments. Modifications in 2014 provided opportunities for citizens to better understand and participate in government and its processes; and drive innovation and maximise economic opportunities to create a more cost-effective, efficient and responsive government (p.1, Tydd, 2015).*

The objective of this article is to dissect the governance of this action plan and its capacities for spurring innovation both internally and across society at large. This critical case study approach is based upon the following evidentiary layers: first, a selective literature review on open government and how this concept encompasses the inter-related elements of open data, big data, and innovation (building upon prior contributions of this author to the scholarly literature); secondly, direct observation of the Government of Canada's consultative exercise undertaken to help inform this action plan as well as likeminded forums with public sector managers from all levels of government; thirdly, classroom interactions and discussions with dozens of mid-career public servants pertaining to the information culture and open government initiatives of the federal government; and fourthly, a series of semi-structured interviews with fifteen government and industry managers engaged directly or indirectly with various elements of the Open Government Action Plan. Supplementing this Canadian case study are scholarly and applied examples from likeminded democratic countries, notably the United States and the United Kingdom.

The article is organized as follows. Following this brief introduction, section two examines the evolution of open data and open government in Canada. Section three then focuses on big data – and countervailing tendencies toward more closed government (in line with the inertia of traditional democratic structures and public sector governance). Section four examines tensions between data-driven and more deliberative notions of innovation and learning within democratic contexts, and if and how governments are attempting to surmount such tensions. Building upon this analysis, section five proposes a set of directions for public sector governance reforms, both administratively and democratically, that will better allow for the pursuit and realization of systemic openness and collective innovation. The article then concludes with a summary of the main lessons learned and some proposed future research directions.

2 Open Data and Open Government

New participatory mechanisms, systemic openness, and virtualization are underpinning an emerging governance ethos that, for the public sector, is often termed as the emergence of Gov 2.0. At the heart of Gov 2.0 are drivers of collective intelligence and more collaborative forms of governance that are typically associated with a widening online universe and less hierarchical and control-minded forms of governance (Shirky, 2008; Wyld, 2010; Lips, 2012). From both external vantage points on new societal formations (such as Wikipedia and a myriad of social media-driven movements) as well as internal to the public sector (what Lips characterizes as 'public administration 2.0'), governments are increasingly challenged to move beyond a typology of hierarchies and markets and embrace usage of networks typically more open and collaborative in formation and execution (Stoker, 2005; World Economic Forum, 2011; Kostakis, 2011; Reddick and Aikins, 2012; Gil-Garcia, 2012; Roy, 2013a/b).

Governments are embracing such changes which present significant structural and cultural shifts. Aligned with the spirit of such principles, the spreading of web 2.0 experimentation within government is specifically meant to foster collaboration and democratize the creation and exchange of ideas:

> *The role of citizens in an open government environment – enriched by open government data – can be one of democratic innovators. In an ongoing open innovation process, citizens can draw on open data, and propose both policy-areas to tackle and technical approaches to take (p.186, Maier-Rabler and Hubler, 2011).*

The potential recasting of governance in terms of expectations and roles is profound. Rather than gathering information and ideas via highly regimented and contained mechanisms (shaped by a proprietary mindset), this alternative presentation of openness and ideas begins from the premise that the ownership of information and ideas is fundamentally diffused and shared. At the same time, however, such an ethos of openness invariably faces strong pushback from both the traditions of proprietary protection and its organizational cousin that is particularly prevalent in the public sector - namely hierarchical and informational control. For example, one early study of the usage and acceptance of new social media within the public sector found such tensions deeply engrained within Canadian government where information is viewed predominantly as a proprietary asset. The authors conclude that the most significant impediment to Gov 2.0-inspired reform is the 'clay layer' embedded by a hierarchical public service culture (p.3, Fyfe and Crookall 2010). Along with much fanfare in that country with respect to open data initiatives, an independent review of information management processes within the British government found 'concern about publishing data externally' (p.3, Read, 2012).

By contrast, within the rubric of open government, the notion of open data is based upon the 'notion that public sector information is a resource, the release of which will maximize its social and economic value to citizens' (ibid.). In the Netherlands, for example, an impetus for non-proprietary public data came from the Dutch courts in April of 2009 when a City of Amsterdam's appeal to impose restrictions and fees over several its data holdings was rejected (Ubaldi 2013). Such clashes between proprietary and openness, and control and empowerment shape the pursuit and effectiveness of open data and its wider ramifications (Bermonte, 2011; Roy, 2013a). Outside of government too, similar tensions between proprietary and open systems are prevalent across many segments of industry and society (Wyld, 2010; Public Administration Committee, 2011; World Economic Forum, 2011).

Yet a widening ethos of openness draws sustenance from: i) the Internet as a platform for democratization in the broadest sense; ii) the search engine and a widening array of self-expressive and interactive web 2.0 tools and platforms; and iii) most recently the advent of mobility. As Young puts it, the cloud as a symbolic basis of a wider virtual universe driven by a myriad of smaller and more powerful and mobile computing devices, a penchant to share more and more personal information online – especially via social media, and a new form of enhanced and shared networked intelligence (Young, 2012). At the same time, however, accompanying optimistic portrayals of the potential benefits of such intelligence come offsetting concerns pertaining to individual privacy, while open data has similarly sparked fresh concerns about the digital divide and accentuating new forms of 'data divides' (Halonen, 2012).

Indeed, as important to government efforts to release data is society's interest and ability in accessing and making use of it. Open data's origins are interwoven with a growing community of activists and apps developers working initially within the confines of privately-developed operating system platforms such as Apple and Android (the latter built from open sourced coding and thus more portable across a range of companies and devices). The participative flavour of such movements can and has also extended beyond commercial pursuits, as exemplified in February 2013 by the inaugural open data day (that has since grown into a global network of more than 200 community-based events around the world).

In Canada, as an important precursor to explicit open data strategies, one early example of the potential for government to embrace an ethos of openness came from the City of Nanaimo, British Columbia, on the west coast of Canada which effectively abandoned its prior model of internalized and proprietary and infrastructure and information holdings within the realm of geographic information systems and spatial data mapping. Citing the benefits of open innovation through greater usage and access and heightened redundancy and security, the municipal government opted for open source tools (including freely available Google Earth online) and shifted its data imaging that it previously regarded as a proprietary asset to Google's cloud-enabled platform (Birch, 2008). Some five years later, this same municipality would become the first in Canada to adopt a 'Pan-Canadian Open Government License' for its data holdings.

More than thirty Canadian municipalities, of all sizes, have now undertaken open data strategies. One such example, adopted in 2012 by the City of Halifax puts forth the following drivers of doing so (summarized here): restrictive data policies limiting the public good; costly and inefficient public data sharing processes; local community movements seeking greater data access and usage; open data as a driver of economic growth; and open data as a platform for increased transparency and citizen engagement (Halifax Regional Council, 2012). Whereas the first and second drivers apply mainly to the internal apparatus of information management by the municipality, the latter themes underscore the wider societal and participatory dimensions of open data as a key source of collective innovation across both civic and economic pursuits as well as the interdependencies across both realms.

Spurred by this local emergence of initiatives on the one hand, and by the emergence of an international network of countries committed to open government principles on the other hand, the Government of Canada released its own Open Government Action Plan in 2012, having since updated it to include a series of initiatives and objectives for the timeframe of 2014-2016. The action plan is based upon three inter-related dimensions: information, data, and dialogue. Whereas 'information' centres mainly upon transparency about government operations and research, 'data' is most closely associated with online technologies and open data platforms in line with the aforementioned strategies of many Canadian municipalities. Finally, 'dialogue' is based upon the engagement of citizens in democratic governance, thereby illuminating at least the potential for inter-linkages between open data, open government and the wider participatory currents of Gov 2.0.

Any reasoned assessment of open government in Canada would nonetheless conclude that of these three dimensions, only 'data' has been acted upon with any degree of seriousness as reflected by tangible initiatives. Along with releasing a large number of data sets via its own open data portal, the Government of Canada has sought to collaboratively forge a pan-Canadian license for such data sharing that would be portable across all government levels in Canada. It has also created a national 'hackathon' (the Canadian Open Data Experience) to spur usage of the public data sets through coding and applications development in three categories: youth, commerce, and quality of life. Additionally, Canada has been recognized internationally for having made some strides in its open data endeavours (World Wide Web Foundation, 2015).

Therefore, while open data arguably remains at its inception – with little research as yet seeking to quantify its impacts and return on investment, we can at least point to specific undertakings and progress. Why, then, is it so difficult to make similar claims with respect to information and dialogue, and how does the absence of seriousness in these two dimensions impact the governments capacities to be an innovation catalyst? The next two sections address each of these important questions in turn.

3 Big Data and Closed Government

While openness and ownership of data resources lie at the heart of the Government of Canada Open Government agenda, the traditional public sector culture of information resources and management runs directly counter to such currents: on this point there is widespread agreement from scholars, media observers, and public sector oversight bodies alike (Bermonte, 2011; Read, 2012; Ubaldi, 2013; Roy, 2013a). There is also evidence that the Canadian Westminster model is a particularly egregious example of centralized and control-minded information management, even relative to its Parliamentary peers in Australia and the United Kingdom (Aucoin et al., 2011; Roy, 2013a).

There are both political and operational dimensions to such a charge. Politically, governments struggle with messaging and communications and a traditional media apparatus that often promotes information protectionism and spin (Martin, 2010). Operationally, governments have traditionally managed their digital infrastructure via proprietary software and hardware systems that reinforce confidentiality and a control-minded mindset in terms of procurement and contracting, leading one British Parliamentary review to characterize such conditions as a ‘recipe for rip-off’ (Public Administration Committee, 2011; Roy, 2013a). Such a political and organizational apparatus as it has evolved over past decades is thus poorly suited to a genuine cultural commitment of open government (beyond narrow and precise actions such as opening up specific data sets).

Accordingly, the Government of Canada’s organizational architecture for managing its Open Government agenda personified such tensions. The unit responsible for open government is housed within the central agency (Treasury Board) responsible for expenditure management controls for the government as a whole. The same Minister is thus dually responsible for overseeing a traditionally inward and control-minded organization and facilitating an alternative governance mindset predicated upon openness. While in fairness it should be noted that the Chief Information Officer’s Branch is also located within the same central agency – thus providing a platform for government-wide management of digital infrastructure, the previously noted lament from the British Parliamentary Committee also applies in equal measure here: a predominantly proprietary approach to managing such systems is much more in line with the control-minded aspects of Treasury Board than its novel embracement of systemic information and data openness.

The resulting context for dialogue – the third dimension of the Open Government agenda, is predictably minimalist and constraining, featuring a highly general set of promises to improve opportunities for public input and engagement in policy and services processes. Beyond the aforementioned example of the 2015 Hackathon, there are few if any concrete initiatives designed to spur wider public conversations about data openness and innovation. Similarly, while the government plan invokes new social media platforms for expanding public engagement, research has instead shown that Government's usage of social media is largely focused on communications and information provisioning, rather than listening and interaction (Fyfe and Crookall, 2010; Roy, 2013a). Here once again the organizational architecture further constrains innovative approaches for outward governance in so far as the two central agencies tasked with leading this dialogue dimension (Treasury Board and Privy Council Office) and poorly suited to such a role given their centralizing functions (ibid.).

In contrast to such positioning, evidence from comparative research suggests that innovative public engagement requires a newly created organizational unit with the competencies and culture conducive to such an alternative role (World Economic Forum, 2011; Mergel, 2012; Dalakiouridou et al., 2012). Beyond a sub-unit of Treasury Board and the CIO Branch with responsibilities primarily focused on open data initiatives, the absence of new organizational actors in order to facilitate a government-wide focus on information management and public engagement (or dialogue) reinforces the notion of closed government and can only constrain the pursuit of innovation within and outside of government.

Importantly, this point has been recognized by other governments even within Canada, notably the Province of British Columbia which created an inter-Ministerial task force that, in turn, has recommended the creation of a new and autonomous agency to shift beyond its existing open data effort (that in many respects mirrors the model of the Government of Canada) and to focus on collaborative opportunities within and outside of government to leverage big data for new and wider forms of innovation:

> *The intent for the centre is to create a hub of interaction, exploration, analysis and innovation. It is targeted at a wide range of users to enable enhanced engagement and collaboration between these groups on key questions of interest to each (p.6 BC Centre for Data Innovation, 2014).*

As discussed previously, an important lesson of public administration is that traditional top-down and control-laden structures of Westminster-stylized governance – personified by central agencies such as the Government of Canada's Treasury Board, are poorly suited to devising more outward and collaborative forms of governance. As such, this BC example responds to this truism at least conceptually in proposing an alternative governance mechanism more suitable to charting new public sector capacities more appropriate for a networked and data-rich era.

An accompanying challenge surrounds the extent to which such capacities are genuinely participatory and encompassing of meaningful public dialogue, a stated aim of the Government of Canada Open Government Action Plan. Underpinning this challenge is the question of whether data-driven innovation requires such participation, or instead whether increasingly computational and algorithmic processes are the primary drivers of value creation and innovation. Understanding this challenge, and government responses to it, require a deeper consideration of innovation and its relationship to both data and dialogue.

4 Data versus Dialogue in Innovation

The inclusion of dialogue as one of three pillars of the Government of Canada's Open Government Agenda (along with data and information) is testament to the importance of human interaction and participation as drivers of innovation, either within or between organizations or across societies at large. Indeed, building upon the Obama impetus, Lee and Kwak articulate a set of four escalating levels of such participation culminating in the realization of 'ubiquitous engagement' as the ultimate set of conditions for driving collective innovation in such a manner (Lee and Kwak, 2011).

At the same time, there are tensions between data and dialogue – reflecting wider tensions between automated and analytical processes on the one hand (that are at the heart of big data systems), and more human, interactive and deliberative processes on the other hand. A number of prominent voices have expressed concern that the former comes at the expense of the latter. A case in point is technology critic Nicholas Carr who surprised many with his characterization of Google as a bureaucratic leviathan, less in terms of how the organization treats its workers internally and more in championing an algorithmic society that reduces individual freedom and cognition as more and more decisions are instead automated (and thus standardized):

> *What Taylor did for the work of the hand, Google is doing for the work of the mind. In Google's view, information is a kind of commodity, a utilitarian resource that can be mined and processed with industrial efficiency. The more pieces of information we can "access" and the faster we can extract their gist, the more productive we become as thinkers (Carr 2008).*

A similarly critical tone underpins a Guardian article entitled, 'The rise of data and the death of politics' which essentially argues that in an increasingly data-driven environment, traditional forms of discursive politics loses out as consumerism and immediacy further reinforce individualization and analytical capacities over more collective forms of engagement (Morozov, 2014). Nabatchi strikes a similar chord in her portrayal of a 'citizenship and democratic deficit' increasingly prominent in today's online world (Nabatchi, 2010).

The salient point to recognize here is that such viewpoints tie together democracy and innovation in ways that are often under-appreciated or ignored by big data proponents within government and perhaps especially in private sector companies cultivating data capacities of one sort or another. In three essential ways, government's ability to pursue and foster innovation both internally and across society is contingent upon dialogue: first, collective intelligence as discursive and participatory processes; secondly, diversity and inclusion as innovation stimulants; and thirdly, political literacy underpinning governmental investments and actions and how such actions are gauged and adapted over time.

On the first point, there are of many prominent voices countering the assertions of Carr and others as to the evolution of governance in an increasingly online and inter-connected world (many referenced in earlier sections of this article depicting the rise of Gov 2.0). Underpinning many such voices is a philosophy of governance predicated upon grassroots engagement, spontaneous forms of activity, and learning and adaptation. Linking such attributes to big data and social innovation, Desouza and Smith suggest that the promotion of 'virtual experimentation platforms' are essential in order to 'increase our understanding of how to use big data': such platforms are predicated about social interaction and human collaboration (Desouza and Smith, 2014). Sifry makes a similar critique of the technocratic rise of big data systems that largely favour private value creation over civic engagement and public value, suggesting that devising ways to cultivate smarter citizens is as important if not more important than designing smarter governance systems (Sifry, 2014).

This latter point speaks to the second theme above, namely the importance of diversity and inclusion to cultivating an environment conducive to innovation. The embracement of hackathons and apps competitions by governments at all levels in Canada is a case in point, in extending the public sector's reach to a wider set of competencies and a wider realm of creativity than otherwise available internally or even via more traditional partnering arrangements with outside experts such as consultancies. Yet by the same token, socio-economic polarization driven largely by educational attainment is viewed as significant challenge in this increasingly digital and data-centric world (Jaeger 2012; Janssen et al. 2012). One study of such initiatives from a British think tank pointed to widening 'data divides' that basically reflect the application of historically-rooted and potentially now reinforced forms of digital divides to this new data intense landscape (Halonen, 2012). If the public sector is to encourage smarter citizens as well as smarter governance systems – and foster civic-based forms of innovation to spur public value creation alongside private value creation, societal inclusiveness and collective innovation must be viewed as intertwined objectives.

The third point above - namely political literacy, follows from the preceding themes in so far as an engaged and inclusive society well versed in the opportunities and risks of big data and seeking collective innovation for both private and public purposes will likely yield – and more to the point at hand, requires a digitally and data literate political class. By contrast, a highly technocratic data regime – as we often see exemplified by national security efforts invariably generating controversial outcomes in secret before eventual exposure (typically by stakeholders and activists outside of formal political institutions), breeds suspicion and distrust (Roy, 2015).

While global rankings lauding the US and UK as open government leaders may be contested by some due to the source of these rankings, it does bear noting that both jurisdictions share well documented ambitions for various facets of open government (including open data and bid data) that are underpinned by strong political commitments of elected leaders. Though beyond the scope of this paper to provide any sort of objective and detailed assessment of the British Government, it does bear noting that in what many regard as the most privacy-sensitive public service domain of health care, in comparison to Canada, the UK has taken some notable steps in rebalancing individual privacy and the pursuit of collective societal innovation in this space (Callaway, 2013). We can postulate that the relatively higher level of political literacy on display in that country in recent years is an important variable in enabling action of this sort and the required public understanding and support.

While health care falls predominantly within provincial jurisdiction in Canada, the wider relevance of this observation remains. The single initiative undertaken by the legislative branch of the Canadian Parliament in the realm of open data, for instance, led to political gridlock and traditional stances around privacy issues based upon partisan perspectives, and Canadian Governmental action in terms of data surveillance for security purposes remains shielded from political oversight in a much more substantive manner than in most other democratic countries (Roy, 2015). The sharply critical stances of independent oversight bodies, notably the federal Information and Privacy Commissioners, further reinforce the adversarial (rather than more open and collaborative conditions conducive to more open and inclusive governance) nature of Westminster politics. In such an environment, information is more likely to be viewed and processed as proprietary by political and administrative actors – and big data capacities are more likely to evolve in a more technocratic manner as a result.

In sum, despite the inclusion of dialogue as one of three dimensions in the Government of Canada's Open Government Action Plan, not only is this dialogue stunted by weak capacities within the executive branch (as discussed), but fractious political institutions and a relatively disengaged citizenry further reduce the scope for meaningful and discursive public engagement that, in turn, drive innovation both within and across government on the one hand, and within and across the economy and society on the other hand.

5 Toward Open and Innovative Democratic Governance

The over-riding lesson from the preceding analysis is that positive linkages between data, dialogue and innovation are not likely to emerge organically within a traditional democratic and governmental apparatus such as Canada's Westminster's Parliamentary regime. In fact, the risks are considerable that the tensions and frictions between traditional government and open government could not only constrain innovation within the public sector but also heighten cynicism and distrust amongst the citizenry at large. Underscoring this latter point are heightened signs of voter apathy and distrust in Canada and elsewhere, a point further reinforced by a recent public opinion survey of Americans probing them as the impacts of open government initiatives on governmental performance and trust (PEW, 2015).

What, then, must change, if governments are to devise meaningful governance capacities to leverage the benefits of open data and big data for greater innovation both internally and across society? Essentially, four inter-related governance architectures of the public sector must be transformed, including: the organizational, the technological, the political, and the societal. We examine each in turn, drawing upon preceding examples and discussion in order to highlight the Government of Canada's shortcomings in each realm and the sorts of reforms required going forward.

First, with respect to organizational architecture, as exemplified by the Treasury Board of the Government of Canada, traditional structures of executive branch government (predicated more on principles and policies of closed government than open government) are poorly suited to the pursuit of systemic openness and innovation. With respect to open data and big data specifically, the Province of British Columbia example provides recognition of this point, while a likeminded

British observer calls for the creation of a new 'Advanced Analytics Team' within Cabinet Office to champion novel cross-governmental approaches and new governance mechanisms to enable collaborative innovation (Yiu, 2012).

It bears noting here that the private sector firms driving data accumulation and processing (social media companies, cloud computing and data-mining experts etc.) are doing so through entirely new organizational structures – and many data advocates are unrealistically calling for governments to embrace similar mindsets and techniques (or instead to readily partner within them in doing so). The challenge for government is much more complex and lies in creating new spaces for innovation and experimentation within deeply embedded structures, most of which have been predicated limited openness and strong degrees of hierarchical control. The example noted above of virtual experimentation platforms is a case in point, necessitating the creation of new outward governance capacities for data sharing and public engagement. Discursive and participative forms of innovation will otherwise be stymied.

Secondly, in terms of technological architecture, an excessive reliance on proprietary hardware and software systems reinforces the traditionalism of government and further reinforces resistance to systemic openness and what some have term to be the fostering of 'open source democracy' (Kostakis, 2011; Maier-Rubler and Huber, 2011; Harrison et al., 2012; Roy, 2013a/b). Echoing more recent calls for more open source government by former White House Chief Technology Officer Beth Noveck, the 2011 British Parliamentary report invokes openness of technology solutions as a basis for wider forms of participative value creation and innovation:

> *We see a clear opportunity for Government to adopt this model. IT enabled public services should be provided on an open platform with open interfaces. Government should provide the necessary open infrastructure that empowers people inside and outside of Government to innovate (p.47, Public Administration Committee 2011).*

There is evidence of the UK government having embraced such a model and mindset in its most recent reforms (Fishenden and Johnson, 2014). Conversely, the Government of Canada has pursued its own digital refurbishment in largely the same sort of predominantly proprietary mindset discredited by the 2011 British review (Roy, 2013b). The result is a reinforcement of traditionalism across the technological and organizational apparatuses that invariably constrain the Government of Canada's Open Government Action Plan ability to foster more systemic governance openness.

Thirdly, a new political architecture is required in order to meaningfully embrace the centrality of dialogue and new forms of public engagement as the linchpin between data and innovation, rather than mere recognition of its importance as is the case within the Government of Canada model at present. In their quest for ubiquitous engagement within the US federal government, for example, Lee and Kwak underline the importance of 'creating and nurturing a self-sustaining ecosystem for public engagement is an important touchstone of open government efforts (p.25, Lee and Kwak, 2011). In an effort to create such an eco-system, the Obama Administration thus created a new office of public engagement in 2009 commiserate with this novel and outward facing functionality (Mergel, 2012). The UK example is once again illuminating, at least in so far as the legislative branch of the British Parliament has recognized their own incapacities for digital dialogue with the public at large and in calling for deep-rooted reforms to address this deficiency (Speaker's Commission on Digital Democracy, 2015). The resulting recommendation for a new space within Parliamentary to formally house public deliberations and integrate them within the workings of Parliament (among other proposals) provides an example of how democratic and political innovations are intertwined.

Perhaps nowhere is the need for public engagement and dialogue more pronounced than on matters of privacy and redefining the balance between autonomy and openness, a debate viewed as either precarious or polarizing across large segments of populations. New forms of citizen involvement and oversight are thus essential for governments to pursue and realize the benefits of big data on the one hand, and to ensure democratic accountability and facilitate collective learning on the other hand. To quote from the New South Wales Information Commissioner in Australia, 'the increasingly digital environment requires greater coordination and oversight to ensure maximized civic engagement and public trust in the management of government information (p.54, Tydd, 2014).

Fourthly, and finally, the intertwined objectives of digital and data inclusiveness must be embraced as a social, political and economic objective for a jurisdiction such as Canada – or instead, a widening of existing digital divides is certain to follow. As we have seen with the Government of Canada's apps competition (following many similar examples of other jurisdictions in Canada and elsewhere), mobile devices are viewed as a critical enabler of citizen involvement with data resources and how such resources can be leveraged for public interest pursuits. In the UK more broadly, mobility has been viewed as a key platform to stimulate online usage of government resources for those otherwise disenfranchised groups that have thus far shunned or been unable to partake in online processes (Roy, 2014). On the other hand, Benton provides a more sceptical tone of mobility in this regard and the ability of smart phone usage to lessen what are already significant digital divides:

> *A final facet of the digital divide is that of smart-phones – which minorities and disadvantaged groups are more likely to rely on as their main method of accessing the Internet – often display an inferior version of full websites, and thus may provide a second-class form of Internet access. In seeking to capitalize on the opportunities that smart phones offer, policy-makers have to walk a fine line between improving access among those who would not otherwise have a line to city services and perpetuating a two-tiered system (p.11, Benton 2014).*

Geographic cleavages between urban and rural dwellings also present an important challenge to more inclusive governance. This quote's focus is that of so-called smart cities, where technological innovation and social diversity tend to be most intense. Indeed within the Canadian context, open data initiatives and big data companies tend to be most highly concentrated in large city centres, risking a further alienation of rural and remote communities – many continuing to struggle with affordable and reliable high speed Internet access in broadband or mobile form (Roy, 2014). Such cleavages risk greater divides and exclusion as Morozov warns: 'algorithmic regulation, whatever its immediate benefits, will give us a political regime where technology corporations and government bureaucrats call all the shots' (p.11, Morozov, 2014). Despite a fair bit of evidence to suggests that most democratic governments are well-intended in promoting open government as a means to greater participation and inclusion – and to civic and social innovation along with economic innovation, there are legitimate questions about whether the capacities for such an agenda are developing in concert with an otherwise predominantly commercializing and individualizing online culture.

6 Conclusion

On the one hand, incentivized in part by the general evolution of the Internet era and the internationalizing agenda of many countries (especially those of the OECD), the Government of Canada has sought to develop an Open Government Action Plan predicated upon three main dimensions: information, data, and dialogue. A number of specific initiatives have been devised, most especially in the realm of data, notably the creation of an open data portal and an annual hackathon event to encourage innovation through the wider sharing and usage of such data holdings for economic and social purposes. On the other hand, an information management regime steeped in historical tendencies toward selective and reactive transparency, as well as a stated focus on public dialogue that runs fundamentally counter to the control-laden, bureaucratic structures and culture of a Westminster-stylized machinery of government render such an action plan problematic in many respects.

The Government of Canada is arguably emblematic of the wider struggles of the public sector generally, in much of the world, to reconcile the tensions between traditionalism and reform embedded within the evolution of an open government (Ubaldi, 2013). This article has argued that in resolving such tensions, governments must develop new governance architectures organizationally, technologically, politically, and socially. As is the case in Canada (especially if one adds consideration of provincial and municipal efforts excluded from the analysis of this article), as well as in other jurisdictions, notably the US and the UK, governments are beginning to experiment to varying degrees within and across each of these realms.

One over-riding conclusion from this analysis is the growing need for inter-disciplinary endeavours within as well as outside of government. Within the public sector, to draw from the Government of Canada example, the three inter-related dimensions of information, data and dialogue all stem from highly differentiated traditions and skill sets, even as they must be increasingly integrated going forward. Similarly, if government is to orchestrate a societal focus on big data and shared innovation, creating multi-stakeholder venues welcoming of varying disciplines and perspectives is of paramount importance. Creating a collectivized and openly discursive forum for risk management, for example, in order to identify and mitigate the unintended consequences of big data in proactive and reactive manners, can both bring new competencies into government and widen public learning and trust (Quigley and Roy, 2011).

With respect to promising future research directions, more investigation is required into the specific determinants of innovation through data-driven efforts and how such efforts are shaped by the various types of architectures identified in the preceding discussion (i.e., organizational, technological, political, and social). As more and more governments experiment with alternative governance arrangements to pursue data-driven innovation and open government, comparative case studies can provide further illumination in this regard. It is also of paramount importance for governments and researchers alike (separately and via new partnerships) further study how open data sources are being accessed and if and how they are shaping big data systems (and, in turn, how these systems are being utilized and to what end). Additionally, the attitudes and mindsets of elected officials and senior public servants are invariably key determinants in shaping or constraining any public sector reform agenda, and more study is required here in order to expose and understand social, managerial, demographic and political cleavages at play in developing and overseeing open government going forward.

In sum, open government has been embraced with remarkable speed by jurisdictions around the world, viewed by many as a pathway to not only greater public understanding and accountability but also systemic innovation both within the public sector and across increasingly digitized and networked societies. Yet the realization of public value remains in its infancy, constrained in many respects by inward democratic governance regimes. In the case of the Government of Canada, the rhetorical foundations for breaking free from this inertia are at least partially established, whereas the constructing of new realities remains very much a work in progress.

7 References

Accenture. (2009). Web 2.0 and the next generation of public service: Driving high performance through Cloud. Accenture.

Aucoin, P., Jarvis, M., & Turnbull, L. (2011). Democratizing the Constitution: Reforming Responsible Government. Emond Montgomery Publications.

Barkin, R. (2011). The Cloud Comes Down to Earth. American City and County. Accessed 9th June 2015. http://americancityandcounty.com.

Bent, T., Cumming, D., Osmon, R. K., LeFebvre, J., & Reynolds, S. (2008). Change your world or the world will change you: The future of collaborative government and Web 2.0. Deloitte & Touche LLP.

Benton, M. (2014). Smart inclusive cities: How new apps, big data, and collaborative technologies are transforming immigrant integration. Migration Policy Institute.

Bermonte, A. (2011). Senior leaders' use of web 2.0 and social media in the Ontario Public Service. Theses and dissertations. Paper 680. Toronto: Ryerson University.

Birch, J. (2008). Nanaimo's GOOGLE Experience. Municipal World, 118(6), 5-9.

Halifax Regional Council. (2012). Council report: Open data initiative. Item No. 11.1.4. Halifax Regional Municipality.

Callaway, E. (2013, October 17). UK push to open up patients' data. Nature, 502, 283.

Carr, N. (2008). Is Google making us stupid? What the internet is doing to our brains. The Atlantic. Accessed 9th June 2015. http://www.theatlantic.com.

Dalakiouridou, E., Tambouris, E., & Tarabanis, K. (2012). eParticipation and online social networks: The case of the European Institutions. European Journal of ePractice, 16, 94-108.

Desouza, K., & Smith, K. (2014). Big data for social innovation. Stanford Social Innovation Review.

Eaves, D. (2009). MuniForge: Creating municipalities that work like the web (Web log comment). Retrieved from: Eaves.ca.

Fishenden, J, & Johnson, M. (2014). A tale of two countries: the digital disruption of government. Paper presented at CAPAM 2014 Biennial Conference, Putrajaya, Malaysia.

Fyfe, T., & Crookall, P. (2010). Social media and public sector policy dilemma. Toronto: IPAC.

BC Centre for Data Innovation. (2014). Final report of the working group. Government of British Columbia.

Giggey, R. (2011, May 7). The G4: Setting city data free. Canadian Government Executive, 17(8).

Gil-Garcia, J. R. (2012). Enacting Electronic Government Success: An Integrative Study of Government-wide Websites, Organizational Capabilities, and Institutions. New York: Springer.

Grimmelikhuijsen, S. (2012). A good man but a bad wizard. About the limits and future of transparency of democratic governments. Information Polity, 17, 293-302.

Halonen, A. (2012). Being open about data analysis of the UK open data policies and applicability of open data. London: Finnish Institute.

Harrison, T.M., Guerrero, S., Burke, B.G., Cook, M., Cresswell, A., Helbig, N., Hrdinova, J., & Pardo, T. (2012). Open government and e-government: Democratic challenges from a public value perspective. Information Polity, 17, 83-97.

Helbig, N., Cresswell, A.M., Burke, B.G., & Luna-Reyes, L. (2012). The Dynamics of Opening Government Data (A White Paper). Albany: Center for Technology in Government.

Hubbard, R., Paquet, G., Wilson, C. (2012). Stewardship: Collaborative Metagovernance and Inquiring Systems. Ottawa: Invenire Books.

Jaeger, B. (2012). New Frontiers in the Digital Divide: Revisiting Policy for Digital Inclusion. Paper presented at 2012 EGPA Annual Conference, Bergen, Norway.

Janssen, M., Charalabidis, Y., and Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management (ISM), 29(4), 258-268.

Kostakis, V. (2011). The Advent of Open Source Democracy and Wikipolitics: Challenges, Threats and Opportunities for Democratic Discourse. An

Interdisciplinary Journal on Human in ICT Environment, 7(1), 9-29.

Laja, S. (2012, January 9). Stockton and Darlington councils launch shared cloud services for schools. The Guardian. Accessed 9th June 2015. http://www.theguardian.com/uk.

Langford, J., & Roy, J. (2008). Moving Towards Cross-Boundary Citizen-Centred Service Delivery: Challenges and Lessons from Canada and Around the World. Washington, D.C: IBM Center for the Business of Government.

Lee, G., & Kwak, Y. (2011). An Open Government Implementation Model: Moving to Increased Public Engagement. IBM Center for The Business of Government.

Lips, M. (2012). E-government is dead: Long Live Public Adminstration 2.0. Information Polity, 17, 239-250.

Maier-Rabler, U., & Huber, S. (2011). Open: the changing relation between citizens, public administration, and political authority. Journal of Democracy, 3(2), 182-191.

Martin, L. (2010). Harperland: The Politics of Control. Penguin Group Canada.

Mergel, I. (2012). The social media innovation challenge in the public sector. Information Polity. 17, 281-292.

Morozov, E. (2014, July 20). The rise of data and the death of politics. The Guardian.

Mossberger, K., & Wu, Y. (2012). Civic engagement and local e-government: Social networking comes of age. UIC Institute for Policy and Civic Engagement.

Munson, S., Avrahami, D., Consolvo, S., Fogarty, J., Friedman, B., & Smith, I. (2012). Sunlight or sunburn: A survey of attitutdes toward online availability of US public records. Information Polity, 17, 99-114.

Nabatchi, T. (2010). Addressing the Citizenship and Democratic Deficits: The Potential of deliberative democracy for public administration. The American Review of Public Administration, 40(4), 376-399.

Horrigan, J. & Rainie, L. (2015). Americans' views on open government data. Pew Research Center.

Public Administration Committee. (2011). Twelfth Report - Government and IT: A Recipe For Rip-Offs: Time For A New Approach. London: Parliament of Great Britain.

Quigley, K., & Roy, J. (2011). Cyber-security and risk management in an interoperable world: An examination of governmental action in North America. Social Sciences Computer Review, 30(1), 83-94.

Read, M. (2012). Practical steps to improve management information in government. An Independent Report Commissioned by the Minister for the Cabinet Office and Chief Secretary to the Treasury. London: Government of the United Kingdom.

Reddick, C. G., & Aikins, S. K. Eds. (2012). Web 2.0 technologies and democratic governance: Political, policy and management implications. Springer: New York.

Roy, J. (2013a). From Machinery to Mobility: Government and Democracy in a

Participative Age. Springer: New York.

Roy, J. (2013b). Cloud computing and Gov 2.0: Traditionalism or transformation across the Canadian Public Sector? International Journal of Public Administration in the Digital Age (forthcoming).

Roy, J. (2014). Mobility and Government: Opportunities and Challenges for Service Delivery and Information Management. Toronto: Institute for Citizen-Centred Service.

Roy, J. (2015). Secrecy, security, and digital literacy in an era of meta-data: Why the Canadian westminster model falls short. Intelligence and National Security 30(1) 1-23.

Sifry, M. (2014). Transparency, big data and internet activism. UTNE Reader. Accessed 9th June 2015. http://www.utne.com.

Sousa, S. (2013). How should the government approach the big data challenge? Realising economic opportunities and building an enterprising state. Big Innovation Center.

Shirky, C. (2008). Here Comes Everybody: The Power of Organizing Without Organizations. New York: Penguin Group.

Speaker's Commission on Digital Democracy. (2015). Open up! Digital Democracy Commission. Parliament of Great Britain.

Stoker, G. (2005). Public Value Management – A New Narrative for Networked Governance? American Review of Public Administration, 36(1), 41-57.

Tydd, E. (2015, March). Around the world with open government. Public Administration Today, 42, 52-54.

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. OECD Working Papers on Public Governance. Paris: OECD Publishing.

Wanna, J., & O'Flynn, J. (2008). Collaborative Governance: Collaborative Governance. Australia: The Australian National University Press.

World Economic Forum. (2011). The Future of Government: Lessons Learned from around the World. World Economic Forum: Global Agenda Council on the Future of Government.

World Wide Web Foundation. (2015). Open data barometer global report (2nd ed.). World Wide Web Foundation.

Wyld, D. C. (2010). Moving to the Cloud: An introduction to Cloud computing in government. IBM Center for The Business of Government.

Young, N. (2012) The Virtual Self. Toronto: McClelland & Stewart.

Yiu, C. (2012). The big data opportunity: Making government faster, smarter and more personal. Policy Exchange. London, UK.

JIM 4, 1 (2016) 39-68
HANDLE: http://hdl.handle.net/10216/83250

A Framework for Boosting Revenue Incorporating Big Data

Laura Segarra[1], Hamed Almalki[1], John Elabd[1], Jonathan Gonzalez[1], Michal Marczewski[1], Moaad Alrasheed[1], Luis Rabelo[2]

[1]Students of University of Central Florida, Department of Industrial Engineering and Management Systems, Orlando, Florida

{laura.l.segarra, halmalki, john.elabd, gonzo, marcz, moaad}@knights.ucf.edu

[2]Faculty of University of Central Florida, Department of Industrial Engineering and Management Systems, Orlando, Florida

luis.rabelo@ucf.edu

Abstract. Complex industry partnerships, innovative strategies, and cross-cutting industry competition, challenge business leaders in making strategic and operational decisions that support growth and competitiveness. Companies seeking to inform their business decisions by leveraging "big data" face challenges in processing and analyzing such large and rapid datasets. However leveraging big data can create value for businesses. Although various frameworks exist for implementing analytics, few accommodate the implementation of big data analytics. Our goal is to develop a framework by studying big data on a micro and macro level and examining how companies can use big data to boost revenue through creating value. This research is augmented by an in-depth examination of industry giant Amazon.com. Our results provide a framework that enhances traditional analytical frameworks through the integration of big data analytics. Our findings indicate that an integrated framework provides enhanced insights to decision makers seeking to create value for their businesses.

Keywords. Big Data, Information Analysis, Innovation, Business Management.

1 Introduction

Business leaders face many challenges in establishing and maintaining a competitive advantage in today's fierce and cross-cutting industry. In order to develop ways of differentiating from competitors, while creating business value, business leaders traditionally develop strategies by assessing the business' operational environment along with the capabilities and resources of the company (Harvard Business School, 2006). Advancements in technology and management approaches, such as Business Intelligence Systems and Six Sigma programs, have allowed business leaders to make more informed decisions through the use of data analytics and tools that integrate performance metrics, scorecards, and management reporting (Davenport, 2006). However, as technology continues to advance, the of sheer volume of information generated, variety of sources data is generated from, and velocity in which data is generated, pose challenges for businesses that seek to capture, store, manage, and analyze data that is both large in scope and scale. Giving rise to the term "Big Data", technological advancements have paved the way for data to grow exponentially on a

ISSN 2183-0606
http://www.open-jim.org

global scale through the introduction of new capabilities and mobile devices such as high definition video, smartphones, tablets, GPS, social media, and the "Internet of things" (Gobble, 2013; Fosso Wamba et al., 2014).

1.1 Big Data

According to Gobble (2013), the term "Big Data" refers to the case of having extremely large data sets that require innovative methods in the collection, storage, organization, analysis and sharing of such data. More broadly, big data refers computer network data that cannot be adequately managed and processed through commonly available software and databases due to the enormous rate and size of its production along with its unstructured nature (Manyika et al., 2011; Gobble, 2013; Dewey, 2014). The collection and interpretation of big data is accomplished through strong computing ability that actively engages many digital data streams and uses algorithms to analyze the data in search of meaningful and useful correlations (Davenport, 2014). Key sources of big data include public data, private data, data exhaust, community data, and self-quantification data (George et al., 2014).

The subject of big data has captured the attention of academic researchers and business practitioners alike due to its reported potential for creating business value. Research suggests that the ability to capture and analyze big data efficiently and effectively can lead to the extraction of market and business insights that create business value through the creation of new products and services and also can create value across the global economy through improved competitiveness and productivity. Applying analytics to big data enables companies to create entirely new business models, develop new products and services, improve products while they are in use, and tailor offerings to meet the needs of specific market segments (Manyika et al., 2011; Fosso Wamba et al., 2014;). While traditional analytic approaches often assume stability by focusing on making decisions around exceptions, big data analytic approaches accept a continuously changing environment and focus on the ability to recognize change and react quickly. Smaller data sets associated with traditional analytics sets are known for their use in generating reports that support internal strategic decisions with respect to inventory, price structure, customer base, and offerings to customers. On the other hand, big data analytics uses continuous data sampling to provide additional insights that further enhance strategic decisions and may assist business leaders in identifying new business opportunities, which may also include customer-facing interfaces (Davenport, 2014).

Creating and delivering customer value is at the core of any business strategy and requires research to provide value propositions consistent with customer expectations and needs. Therefore, businesses need gain insights into customer behavior, preferences, and the products or services that customers purchase and use. Understanding the customer's perceived value, the ability to forecast future value perceptions, and the capability to address unique customer requirements are central elements in developing and sustaining a competitive advantage (Nicola et al., 2014). The customer value assessment model proposed by Nicola et al. (2014) provides a quantitative approach for comparing value proposition to customer needs and internal and/or external tangible or intangible assets. The incorporation of big data capabilities can further enhance such approaches through rich data sources and advanced computational capabilities that provide additional insights across a value network along with real time identification and tracking of key factors in determining customer value perceptions. Harasim & Klimontowicz (2013) recognize the fundamental relationship between the diffusion of innovation and customer habits in the retail payment market. User behavior was identified as a critical driver for innovation, and future business strategies are expected to focus increasingly on customer-driven innovation. As customer needs and expectations shift toward real-time payments, ease of use, predictability, and e-payments (Harasim & Klimontowicz, 2013), technologies associated with big data capabilities can assist in meeting such changing customer requirements by providing businesses with key insights derived from customer behavior and trends. For example, Amazon.com uses customer data to provide their customers with suggestions on merchandise they may be interested in on their website by indicating "recommendations for you" or "customers who bought this item also bought". Also, customers can rate products and post reviews in terms of their satisfaction level which can assist Amazon in making internal decisions on its product offerings (Amazon.com, Inc., 2015). Pricing optimization can further be enhanced through the incorporation of external big data associated with influences on consumer demand and competitor prices. Automated algorithms can even adjust prices automatically in response to particular events or trends. Big data analytics may also be extended to other traditional analytics for assessing supply chain risks. Supply chain decisions may be enhanced by leveraging external big data on a company's suppliers and even their suppliers' suppliers with respect to their capabilities, financial standing, quality, reliability, reputation, and practices. Big data can also further enhance traditional market and competition analyses by uncovering new competitive factors and using much more encompassing data sets for trend analysis, benchmarking, and segmentation for deriving strategic alternatives. Management practices for the use of big data in internal decision making have not fully been resolved due to the constant influx of data and lack of establishing decision criteria and timeframes for fluctuating analysis outputs (Davenport, 2014).

Estimates reveal that 1 in 3 business leaders do not trust the quality of the information used in the decision making process (Fosso Wamba et al., 2014). However, additional operational insights, efficiency gains, and enhancements to decision making processes are possible through the use of real-time performance data and automated algorithms (Manyika et al., 2011). Academic and industry research indicate that retailors that apply big data analytics stand to realize a return on investment of up to 20% (Fosso Wamba et al., 2014) and improve operating margins by at least 60% (Manyika et al., 2011). However, businesses struggle to incorporate big data analytics into their practices due to lack of infrastructure, analytic skills, trust, and understanding. In 2011, only 25% of the manufacturing industry leaders believed that digital technologies would significant affect their businesses and was observed to possess insurmountable quantities of data that were never utilized for creating value. And in 2013, 56% percent reported that their companies had not made significant progress in implementing big data projects (Dutta & Bose, 2014).

Research indicates that companies rarely make use of their innovative data and those attempting to put big data analytics into practice can become overwhelmed and are unable to extract any insights of use. Although big data technologies currently exist, a consensus on tools and techniques for managing and using big data to extracting valuable insights is not well established (Gobble, 2013). However, as customer needs shift toward more personalized and custom services that are compatible with the technology platforms that these customers use, insights achievable through big data analytics become increasingly important in identifying specific customer needs along with the innovative capabilities of a businesses to meet these needs (Harasim & Klimontowicz, 2013). Companies are currently trying to gain a better understanding of big data analytics and the associated benefits through pilot projects or the development of a strategy for incorporating big data into their practices (Dutta & Bose, 2014). In addition, there is reportedly a significant shortage in people with skills to perform in depth analytics and managers to make use to such analytics (Manyika et al., 2011; Gobble, 2013).

Scholarly frameworks that integrate the use of big data have not yet been resolved and continue to be of interest to researchers (Dutta & Bose, 2014). Few publications address big data opportunities for introducing new scholarly management tools, practices, and theories. Rather than relying on limited data such as quarterly and annual reports, a shift toward a micro-perspective that incorporates big data can assist scholars in the assessment of business cases, along with the evolution of strategies, practices and behaviors, virtually real time (George et al., 2014).

The following sections provide a description of four different tools, frameworks, and analysis methods commonly used in case study, business analysis, and decision making and include the SWOT analysis, business model, matrix of change, and the strategy map coupled with the balanced score card. These tools, frameworks, and analysis methods form the foundation for the subsequent sections of this paper which contrast the application of traditional analytic techniques with the application of big data analytics in the case of Amazon.com. In addition, we propose a framework that integrates these well-known management tools and frameworks with big data analytics to create a cohesive methodology for the purpose of helping businesses boost their revenue.

1.2 SWOT Analysis

SWOT (Strengths, Weaknesses, Opportunities and Threats) is a tool, developed by Harvard Business School during the 1960's, for finding, collecting, understanding and evaluating internal and external data. As shown if Figure 1, the data is based on four categories: Strengths, Weaknesses, Opportunities and Threats.

Strengths 1. Consider internal information.	**Weaknesses** 1. Consider internal information.
Opportunities 1. Consider external information.	**Threats** 1. Consider external information.

Fig. 1. SWOT Analysis.

The SWOT analysis is one of the most frequently used tools to analyze strategies. Its simplicity and flexibility makes this tool widely used (Al-Araki, 2013). When building a SWOT analysis internal information should be used when considering strengths and weaknesses. On the other hand, external information should be used when considering opportunities and threats.

1.3 Business Model

The business model is a great tool that summarizes the business for the purpose of obtaining the right strategies. "Business models are clearly related to strategy" (Bertels et al., 2015, p.2). It directs the implementation of strategy at a specific point in time and by using it we can analyze the innovation through the business model lens. The business model is a holistic concept where all the three essential innovation lens are presented: technology, value network, and economics (Bertels et al., 2015). Figure 2 depicts a business model framework along with several questions that should be answered when designing a business model (Osterwalder and Pigneur, 2010).

<table>
<tr>
<td colspan="2" rowspan="2"><u>Key Partners</u>
Who are the key partner and suppliers?
Which key resources are acquired from partners and Which key activities are performed by partners?</td>
<td colspan="2"><u>Key Activities</u>
What key activities do the value propositions, distribution channels, customer relationships and revenue streams require?</td>
<td colspan="2" rowspan="2"><u>Value Proposition</u>
What core value is delivered to the customer?
Which customer problem is addressed?
Which customer needs are satisfied?
What products and services are offered to each customer segment?</td>
<td colspan="2"><u>Customer Relationships</u>
What relationship does the customer expect?
Which have been established, how costly are they, and how are they integrated with the model?</td>
<td colspan="2" rowspan="2"><u>Customer Segments</u>
What group(s) of people or organizations are served?
Who is value created for?
Who are the most important customers?</td>
</tr>
<tr>
<td colspan="2"><u>Key Resources</u>
What key resources do the value propositions, distribution channels, customer relationships and revenue streams require?</td>
<td colspan="2"><u>Channels</u>
How are customer segments reached?
How are channels integrated, which ones work best, and which are most cost effective?</td>
</tr>
<tr>
<td colspan="5"><u>Cost Structure</u>
What costs inherent in the business model are the most important?
Which key resources and key activities are the most expensive?</td>
<td colspan="5"><u>Revenue Streams</u>
For what value are customers willing to pay, what and how do they pay and prefer to pay?
How much does each individual revenue stream contribute total revenues?</td>
</tr>
</table>

Fig. 2. Business Model.

Notably, the business model is a generic platform to tie the strategy with practice, describing the design or architecture the value creation, delivery, and capture mechanisms of a firm (Ritala et al., 2014).

1.4 Matrix of Change (MOC)

The MOC can help businesses to assess and understand the difficult interrelationships in strategy change, feasibility of a new system of practices, and the sequence of practices to be changed (Brynjolfsson & Renshaw, 1997).

Elattar (2014) identified the significance of the MOC in that "It is important to remember that the Matrix of Change does not actually provide a solution to problems in transitional management; rather it paints a picture of the transition process and allows stakeholders to better understand the sort of undertaking that will be required for a successful transition" (p. 96).

The MOC is "a visualization tool for capturing the existing and desired states of the proposed change, the complementary and opposing practices and how best to proceed in the implementation of the change" (Massachusetts Institute of Technology and the Center for Coordination Science, 2015).

As shown in Figure 3, three matrices construct the MOC as follows:

- The horizontal matrix (the current existing practices),
- The vertical matrix (the target practices) and
- The transition matrix, which are interactions among processes.

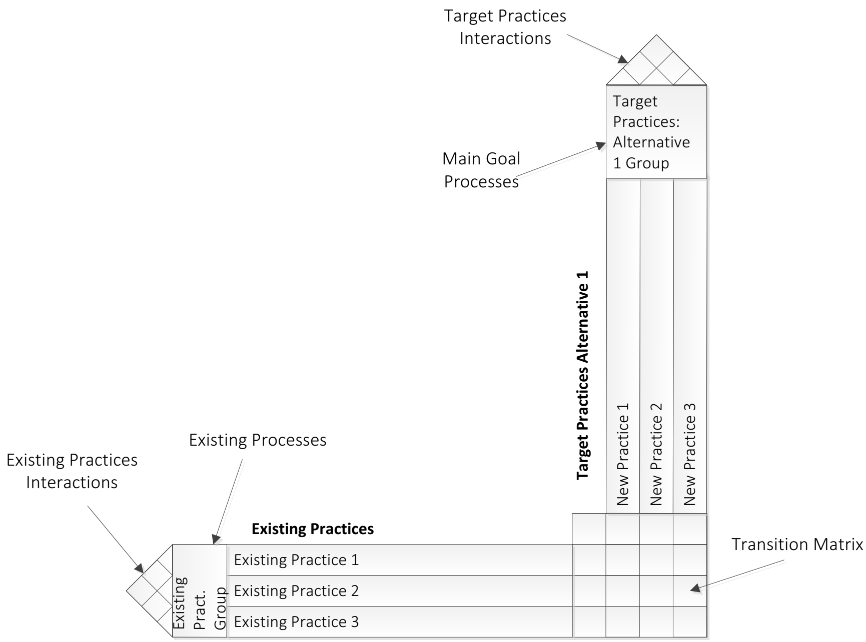


Fig. 3. Matrix of Change (MOC) (Elattar, 2014).

By mapping current practices to the desired future practices through their process interactions, the MOC provides an understanding of how difficult change may be and helps in the formulation of strategies for dealing with such change.

1.5 Strategy Map & Balanced Scorecard (BSC)

A strategy map is at the core of formulating a strategy and identifying primary strategic objectives. As depicted in Figure 4, a strategy map accounts for competitive factors through multiple perspectives including financial, customer, internal process, and learning and growth. Further, the strategy map allows any linkages between objectives to be visually depicted.

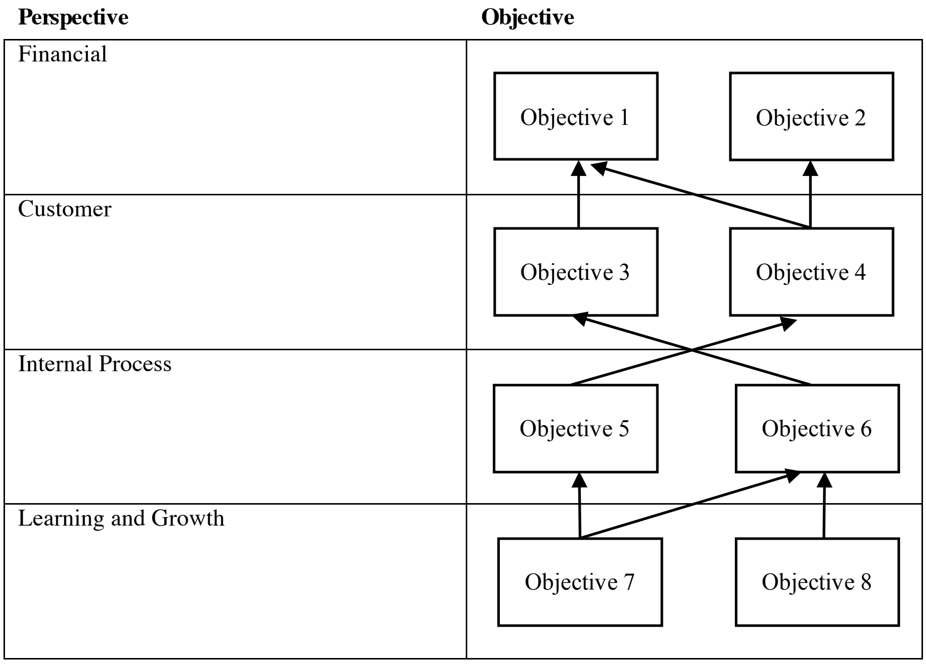


Fig. 4. Strategy Map.

A framework to measure the progress for any organization is highly desirable in assessing success. The balanced scorecard (BSC), shown in Table 1, is a measurement tool with the goal of encouraging businesses to measurement their strategies. The significance of the BSC is that many firms develop strategies without a basis for evaluating, measuring and monitoring their success. More than 50% of the Fortune 1000 use a BSC (Nair, 2004).

Table 1. Balanced Scorecard

Perspective	Objective	Measures	Targets
Financial	Objective 1	How to measure.	Quantify goal.
	Objective 2	How to measure.	Quantify goal.
Customer	Objective 3	How to measure.	Quantify goal.
	Objective 4	How to measure.	Quantify goal.
Internal Process	Objective 5	How to measure.	Quantify goal.
	Objective 6	How to measure.	Quantify goal.
Learning and Growth	Objective 7	How to measure.	Quantify goal.
	Objective 8	How to measure.	Quantify goal.

Consistent with the strategy map, the BSC has four perspectives: learning and growth, the internal processes, customer value, and financial perspective in which the measurement metrics are constructed (Callado & Jack, 2015). In a BSC, a SMART methodology (specific, measurable, assignable, realistic, and time-related) should be applied (Yuanhong et al., 2015).

2 Methodology

Our research of case studies and scholarly literature indicates there is a gap between existing frameworks and the integration of big data analytics in case analysis and strategic decision making. There are many scholarly works that provide concepts and frameworks for achieving various management functions and objectives, however more integrated frameworks that provide steps that companies should implement to realize their full potential through the incorporation of big data analytical techniques are scant.

While authors such as Kaplan & Norton (2001) and Hamel (2002), introduce strategic concepts and tools that address innovation, such as innovative business models, strategy maps, balanced score cards, and profit boosting techniques, the incorporation of big data for enhancing these practices has not yet been addressed. Academic compositions such as those for decision analysis (Clemen & Reilly, 2014) and quality management (Evans & Lindsay, 2014) have not yet addressed the use of big data for growing such capabilities. Works that address technical endeavors associated with engineering activities have not yet addressed ethical issues as a consequence of implementing big data techniques (Martin & Schinzinger, 2005) or the management of engineering and technology as it applies to the use of big data (Morese & Babcock, 2014).

Many researchers such as Brynjolfsson & Saunders (2010), Deighton & Kornfeld (2013), Sahoo, et al. (2014) acknowledge blurred lines between industries, technology convergence, along with platforms and architectures needed for today's digital ecosystems. Other researchers focus on a particular aspect of big data such as data mining (Matsudaira, 2014; Kusiak, 2015) or data-driven marketing (Pousttchi & Hufenbach, 2014). Research conducted by Gobble (2013), Davenport (2014), and Dutta & Bose (2014) have established linkages between the practical use of big data to support strategic decision-making, project management, and innovation. However linkages to integrated frameworks and steps for practical implementation of big data analytics are only in the beginning stages of development, are in short supply in the literature, and deserve further development.

We propose a business enterprise framework for boosting revenue that incorporates the use of big data analytics. The development of our framework began with a review of the literature to identify concepts, frameworks, and methodologies relevant to business case analysis, business solutions, and management practices. In addition, the literature was reviewed to identify case studies and research specific to the collection, storage, analytics, and use of big data. A synergistic approach was used to derive a framework that outlines practical steps for implementing big data capabilities that create value.

To validate the framework, we performed a detailed case study of Amazon.com, across all of their operating segments, in order to provide an understanding of how the company uses big data analytics in providing top customer satisfaction and top sales in the online retail industry. Our case study was informed using publically available information regarding Amazon, their competitors, and global market trends, much of which were in the form of annual, quarterly, stock, and market reports. Our case analysis approach was performed in two steps. First, traditional management tools were used consistent with our framework to assess the company as a whole and each of their operating segments. The second step synergized our review of literature and Amazon's big data capabilities to overlay the use of big data analytics onto the management tools from our framework for the purpose of enhancing traditional methods of analysis. In addition, a panel of fifteen experts from academia (engineering management, computer science, industrial engineering, and business), professional and consulting management, and global online retail, spanning the United States, Latin America, Europe, Asia, and the Middle East, evaluated and validated our framework. This expert panel consisted of individuals experienced in the subject matters (i.e., big data and strategy). The practice included an evaluation questionnaire and individual interview sessions. The evaluation questionnaire consisted of using the case study and the framework to evaluate its consistency and uniqueness. The experts were also interviewed (individually) after the evaluation session. The results were analyzed and incorporated into the framework. Our framework compliments the framework provided by Dutta & Bose (2014) for implementing a big data project. We focus on augmenting traditional analytical tools and methods with big data analytics for better informed decisions that lead to value creation.

2.1 Proposed Framework

The framework shown in Figure 5 uses well-known management tools along with big data analytics to create an integrated methodology for the purpose of helping businesses boost their revenue. The uniqueness of this contribution is performing the framework in sequential processes to guarantee a great result.

The first phase of the framework studies the situation or case by assessing the operational environment and company capabilities. This phase is accomplished through the SWOT analysis (Strengths, Weaknesses, Opportunities and Threats) and the business model, and requires data and information from external and internal sources. The external sources support the development of the opportunities and threats (e.g., technological trends). The internal sources (e.g., financial performance) support the development of the strengths and weaknesses. Big data is a mix of structured, semi-structured, unstructured, and streaming data. Therefore, this phase needs tools that scan the data looking for emerging issues while finding factors that could affect current performance, competitor information, and market information. Tools that emphasize clustering, data mining, and predictive analytics for structured data are able to improve search procedures and identify relationships. For unstructured data, mechanisms are needed for enhancing documentation descriptions and labeling along with enabling search operations such as network analyses. The results from the SWOT analysis helps to inform the company's business model and strategic objectives for how a company generates value and competes differently.

The second phase consists of using the MOC in order to plan the diffusion of change. The utilization of big data and analytics are very important in order to validate the transition matrix and the desired future state. Therefore, the capability of the tools must support risk modeling. Predictive analytics utilizing powerful machine learning paradigms, along with data mining integrated with simulation analytics (Rabelo et al., 2007) offer good foundations for risk modeling. Risk modeling provides justification for the transition matrix along with the corresponding strategies and operational executions with respect to the degree of change, levels of feasibility, order or sequence, and level or consistency of the pace.

The third phase employs the Strategy Map and BSC to provide metrics for measuring performance and determining the best projects through the alignment with strategies that generate an improved and sustainable financial income. Social Analytics and Business Intelligence with transactional data, are able to support the development of schemes to measure performance. In addition, real-time analytics is an important mechanism for enabling customer engagement to be captured from different sources such as mobile apps, digital ads, sensor networks, and web sessions. This flow of information can inform dashboards that help executives visualize performance measures associated with various organizational initiatives (i.e., projects).

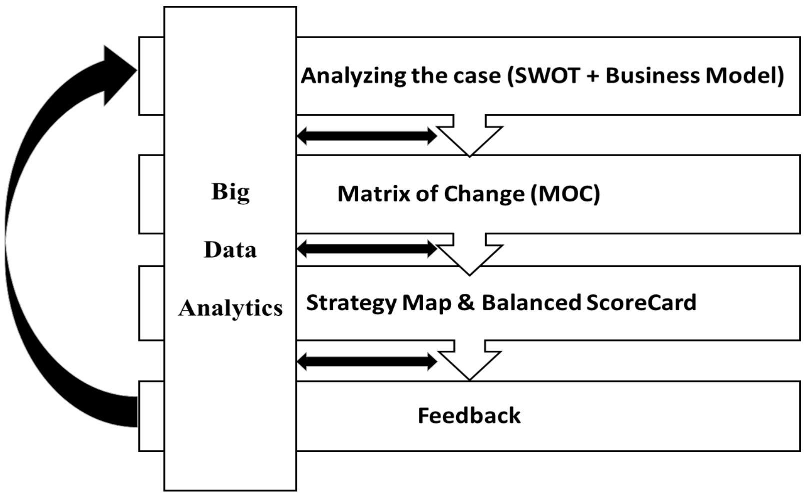


Fig. 5. Revenue Boosting Framework.

In all the phases, the big data analytics methodology will be used to make sure that all analyzing tools are based on informed, authentic and reliable data. The big data analytics methodology will be accomplished for many purposes such as enhancing sales predictions and better matching of products to customers. Finally the feedback of all the framework processes will be closely studied and evaluated for the purpose of making the right decisions in the future decisions. The detailed figures in the subsequent sections show the specific analytical techniques applied to each step within the proposed framework using Amazon.com. The exact techniques depend on the company's goals, objectives and practices which is why we demonstrate the use of our framework for a specific company, Amazon.com, in the subsequent sections.

2.2 Framework Applied to Amazon.com

Amazon.com, Inc. serves consumers, sellers, content creators, and enterprises through their websites and web services. As an online retailer in the Catalog & Mail Order Houses industry and Services sector, two operating segments exist; North America and International. The company's primary source of revenue consists of sales from a variety of products and services to customers. Gross revenue consists of product sales from inventory, while the net share of revenue consists of service sales of items sold by other sellers. Sales are affected by seasonality, and are historically higher during the fiscal year fourth quarter, which ends December 31 (Amazon.com, Inc., 2014; Amazon.com, Inc., 2015).

Several platforms are provided for third-party retailers, marketing and promotional services, and web services for developers, publishing, digital content subscriptions, and advertising services. Amazon Web Services (AWS) serves business customers from data center locations in the U.S., Brazil, Europe, Japan, Singapore, and Australia. Manufacturing and sales of electronic devices include Kindle e-readers, fire tablets, fire TV, echo, and fire phones (Amazon.com, Inc., 2014). International websites include Australia, Brazil, Canada, China, France, Germany, India, Italy, Japan, Mexico, Netherlands, Spain, and the United Kingdom. Other operated websites include www.a9.com and www.alexa.com for search and navigation, and the movie database www.imdb.com (Amazon.com, Inc., 2015).

Amazon has developed and expanded their infrastructure and big data analytic capabilities over many years. As the market reach and customer base grows, so does the amount of data available for analysis. Insights are regularly extracted from customer data due to highly developed data management and analytic capabilities, along with an infrastructure that is both flexible and scalable. A variety of technologies, networks, and tools are implemented in support of real time analytics and productivity management that promotes customer satisfaction and revenue gains. These capabilities are currently being extended to customers through services for digital storage, cloud computing, business enterprise solutions, and big data solutions. Table 2 provides brief descriptions of the company's AWS big data analytics options (Amazon.com, Inc., 2015).

Table 2. Amazon Web Services (AWS) for Big Data Analytics

Analytics Option	Description
Amazon Redshift	Used for analyzing global sales across product mixes, add clicks and impressions, social trends, along with the storage of historical stock market data and the aggregation of gaming data. Measures are provided for quality, operational efficiency and financial performance. Amazon Redshift is also compatible with many business intelligence systems and is designed for data warehouse workloads with structured data. Unstructured data may be prepared and structured for Amazon Redshift through the use of Amazon Elastic MapReduce.
Amazon Kinesis	Used in processing real-time stream data for analysis. Stream data may be rapidly moved from data sources and continuously processed. Data may be transformed and redistributed, analyzed real-time, or decomposed and aggregated across data streams. Amazon Kinesis enables real-time analytics such as customer engagement and website clickstream. Since data is not batched application logs can be pushed directly to an Amazon Kinesis stream for processing. Data processed by Amazon Kinesis may be used in the extraction of metrics and generation of key performance indicators that feed real-time reports and dashboards. Data may be moved and stored through Amazon S3, Amazon Glacier, Amazon Redshift, or Amazon DynamoDB.
Amazon Elastic MapReduce	Uses Apache Hadoop for providing a framework for running big data processing and analytics through the distribution of data sets across compute notes in a Hadoop cluster. The capability is typically used in risk modeling and analytics for threats, ad targeting and click stream, genomics, prediction, and ad-hoc data mining.
Amazon DynamoDB	Stores and retrieves large amounts of data with millisecond latency and is integrated with other services. This capability is commonly used for mobile apps, gaming, digital ads, sensor networks, online shopping carts and managing web sessions.
Amazon Machine Learning	Uses algorithms to find patterns in data for creating machine learning models used in predictive analytics. Predictions may be real-time or scaled. This technology can be used to build predictive models for detecting fraud, recommendations to customers based on prior actions, targeted market campaigns, automatically structuring information, identifying customer attrition risks and mitigations, and a variety of automated solution recommendations.

We focus on applying big data techniques to the business case of Amazon.com (Chen & Zhang, 2014) to demonstrate that implementing big data analytics yields more accurate analysis results compared to more subjective and traditional business analysis methods. For optimization techniques, algorithms may be applied for optimization criterion that reflect the goodness. Statistical techniques are used in identifying correlations and causal relationships between different objectives. Data mining is a technique consisting of information extracted through pattern recognition and involves machine learning and statistics. Machine learning refers to the use of artificial intelligence to design algorithms that allow computers to evolve behaviors based on data. This technique discovers knowledge and makes intelligent decisions automatically. Visualization techniques assist with depicting data in an understandable way such as through the use of tables, images, and diagrams. And lastly, social network analysis examines social relationships for social network, media mining and analysis, along with human behavior modeling.

2.2.1 Amazon SWOT Analysis

A SWOT analysis is performed to evaluate Amazon's strengths, weaknesses, opportunities, and threats in order to better inform subsequent strategy formulation and recommendation activities. Much of the SWOT analysis was derived from publically available financial statements and annual reports for Amazon and their main competitors. The results of the SWOT analysis are shown in Figure 6.

Strengths 1. Top customer satisfaction in online retail. 2. Top sales in online retail. 3. Large regional span with incremental market development. 4. Innovative and wide range of products and services that span multiple markets. 5. Integrated products and services such as Prime subscription that provides the customers many services as a bundle.	**Weaknesses** 1. Low net income 2008-2011, and net income loss in 2012 indicates that Amazon is not profitable. Even though net sales have increased year over year, reaching $61 billion in 2012, Amazon's operating expenses account for 99% of net sales. 2. Slowing sales and growth rates in mature and saturated markets such as the US and the UK. 3. High debt – 67.8%.
Opportunities 1. International e-commerce and mobile device markets are emerging given the appropriate price points and access. 2. Online purchases using mobile devices are growing. 3. Industry technology and capabilities are becoming more and more interconnected. 4. E-books sales exceeded paper book sales. 5. Increasing demand for smartphones and tablets; opportunity to increase tablet market share and enter the smartphone market. 6. International expanding into developing/emerging markets to boost the sales growth in comparison with mature markets like the US with slower growth.	**Threats** 1. The competition from industries that have more resources and better performing products such as Apple, Google, and Microsoft. 2. The competition from Apple's iBook store in the e-book market. 3. Apple's iTunes store is gaining more market share in the multimedia market. 4. The market demand for owning media is significantly declining due to available streaming.

Fig. 6. Amazon SWOT Analysis.

Information that is considered internal to Amazon is used in the derivation of their strengths and weaknesses. On the other hand, information external to Amazon is used to derive relevant opportunities and threats. In addition, often weaknesses and threats can reveal opportunities that can be acted upon.

2.2.2 Amazon Business Model

Using the business model methodology provided by Osterwalder and Pigneur (2010), a business model framework was derived for Amazon and is shown in Figure 7. Publically available financial statements and annual reports were used to derive much of Amazon's business model.

Key Partners	Key Activities	Value Proposition	Customer Relationships	Key Partners
UPS Publishers Suppliers Manufacturers America Online Yahoo! Excite Netscape GeoCities Alta Vista @Home Prodigy	Online Trading Design Productions Investing Financing **Key Resources** Human Resources Inventories Suppliers Partnerships	Low Prices Delivery Speed Convenient Quality Easy Shopping Varieties	Online Self-service Easy payment options **Channels** Amazon.com Apps	Great Prices Great way of shopping Mass Products

Cost Structure		Revenue Streams
Cost of Revenue (sale)	Labor cost	Sale
Packing cost	Assets cost	Prime Subscriptions
Shipping cost	Exchange Rate cost	Fees
Inventory cost	Technology cost	Income from Interest
Administrative cost		Commissions

Fig. 7. Amazon Business Model.

By applying big data analytical techniques, Amazon has the ability to segment populations and use broader datasets for informing their SWOT and business model. The analysis of information on a global scale provides Amazon the means to identify and leverage their most competitive factors for differentiation from their competitors. Following this technique, current and potential competitors are identified through six segments:

1. Physical retailers, publishers, vendors, distributors, manufacturers, and producers of Amazon products.
2. Online e-commerce and mobile e-commerce websites, including those that sell or distribute digital content.
3. Media companies, comparison shopping websites, web portals, and web search engines, either directly or in alliance with other retailers.
4. Companies that provide e-commerce services, including customer service, payment processing, fulfillment, and website development.
5. Companies that provide data storage or computing products or services, including infrastructure and other web services.

6. Companies that design, manufacture, market, or sell consumer electronics, telecommunication, and electronic devices.

An assessment of how competitors use their resources in creating sales, awareness and traffic is significant to quickly identify opportunities and threats that can affect market position. By using real-time analytics, data may be captured as it is generated through data streams, providing the most up to date and accurate information about the current and potential competitors and customers. Competitor pricing may be analyzed in relation to Amazon's and other competitors. Big data processing and analytics that provide risk modeling and threat analytics are useful in assessing large markets. Data mining, global market predictions across products, analytics for social behavior and customer trends are all viable methods in assessing the operational environment and capabilities of a business.

Several potential threats are recognized with respect to current and potential competitors. This includes those that have more resources, longer histories, a larger customer base, and stronger brand recognition than Amazon. These competitors may impact Amazon's business through securing better terms from their suppliers, implementing more aggressive pricing structures, and applying more resources towards technology, infrastructure, marketing, and fulfillment. There is also the potential for other companies to form partnerships in order to strengthen their competitiveness. On the other hand, a competitor's profit margin is an area of significant opportunity due to ability to undercut competitors in the online retail industry.

Amazon is well known for its acute customer insight and has been leveraging customer data for years in order to strengthen their market leadership position. The use of customer facing technologies provides a wealth of data used to improve performance, increase sales, and create value for customers. Identifying patterns in customer data and purchasing habits allows Amazon to personalize the products and services they offer and machine learning allows Amazon to make product and service recommendations based on the customer's habits and preferences. Further, customer behavior may be analyzed to identify needs that are not being fulfilled. The company aggressively invests in infrastructure and technologies to support rapid expansion and leveraging of their customer base.

Any changes that may affect Amazon's business may be assessed through existing data and data streams on an enormous scope and scale. The following two primary competitive factors are identified:

1. Online retail: selection, price, and convenience, including fast and reliable fulfillment.
2. Seller and enterprise services: the quality, speed, and reliability of Amazon's tools and services.

Amazon focuses on creating customer value in order to strengthen their market position and economic model and continues to leverage customer data to achieve this goal. The company has developed and implemented technologies and services that not only benefit customers but also supply them with a wealth of customer data to inform their strategies and operations. Improving the shopping experience of the customer is an ongoing practice. The introduction of 1-Click shopping created convenient purchasing, while customer reviews and pricing comparisons empowered customers with transparency and control over their investments. The company's low prices, vast offerings, recommendation features, and fulfillment performance has paved the way for building customer relationships and brand strength. Further, Amazon has created customer and product data transparency internally to create customer value through superior customer service. For example, the customer that submits an online customer service request receives a phone call from a customer service representative within 30 seconds of clicking the submit button. In addition, the representative is provided with all of the relevant customer and product information needed to address the customer's issue prior to engaging the customer in conversation. Customers are often frustrated by having to provide customer service representatives with information that they know the company already has such as name address, and phone number and generally expect a customer service call to be a negative experience. However, Amazon's customer service representatives are typically able to resolve customer service issues within a matter of minutes. Amazon's management and use of customer big data has gained the trust and loyalty of many consumers and opportunities exist to better tailor the customer experience.

Figure 8 depicts the application of big data analytics to Amazon's SWOT analysis and business model with the result of well-informed strategic goals.

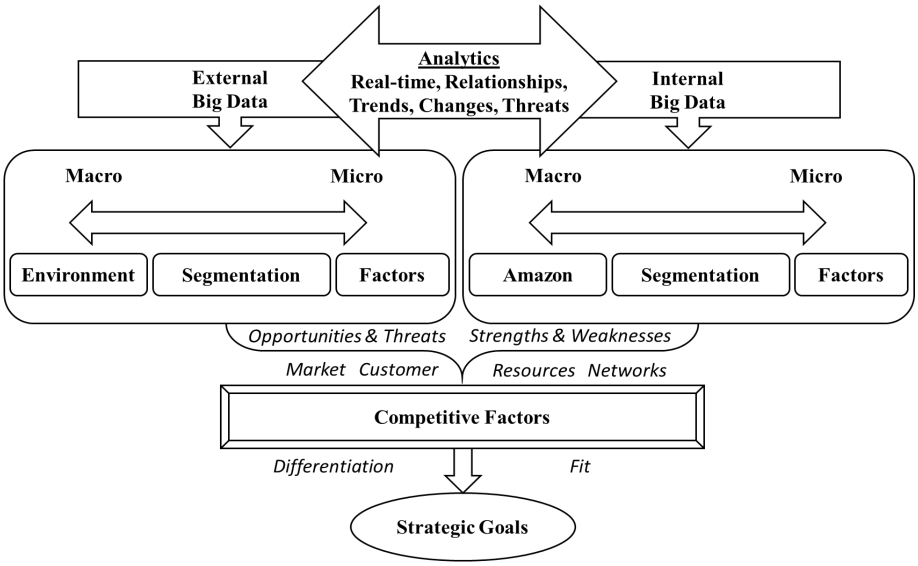


Fig. 8. SWOT and Business Model Applying Big Data Analytics.

How Amazon chooses to compete, the products and services they choose to offer, their basis for differentiation, along with their competitive advantages can be derived from well managed data and the ability to extract insights from that data.

2.2.3 Amazon Matrix of Change (MOC)

The SWOT analysis of Amazon will be used to build Amazon's MOC. Using the MOC will allow us to visualize the case at hand. The MOC analysis is mainly focused on detecting the current and the future practices. Then both practices will be compared to each other to inspect the interactions among them in order to start stakeholder dialogues. Generally there will be two situations: If the transition matrix has more reinforcing than interfering interactions, that would mean the transition will not be so easy and smooth and vice versa. The insights derived from the MOC can be useful in determining the type of leader needed to implement changes along with strategies for achieving successful change.

Figure 9 depepicts Amazon's current practices for the development of the MOC, which are organized by marketing and sales and operations categores. Each practice is evaluated and scored (-2 to +2) with respect to its level of importance. A score of +1 or +2 reflects the degree to which the practices should be preserved during and after the transition. A 0 score indicates no preference, while scores of -1 or -2 reflect problematic practicies. Practices are compared to one another to evaluate whether or not they complement each other. Complementary practices are shown by a "+" sign, whole interfering practice are shown by a "-" sign.

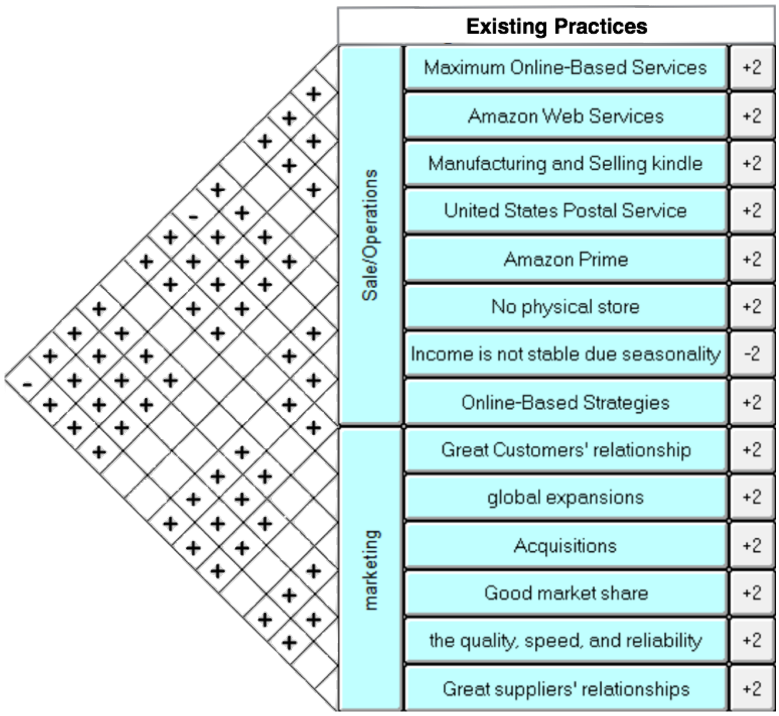


Fig. 9. Amazon Matrix of Change Current Practices.

Amazon's current practices are identified in order to provide a snapshot of the current marketing, sales, and operations capabilities. It is important to capture the situation at hand in order to assess how current strengths and weaknesses impact potential change initiatives.

Figure 10 depicts Amazon's future target practices for the development of the MOC, which are organized by operations, human resources, and marketing strategy categories. Consistent with the current practices assessment, the target practices are evaluated and scored with respect to importance and each target practice is compared with one another to identify complements and interferences.

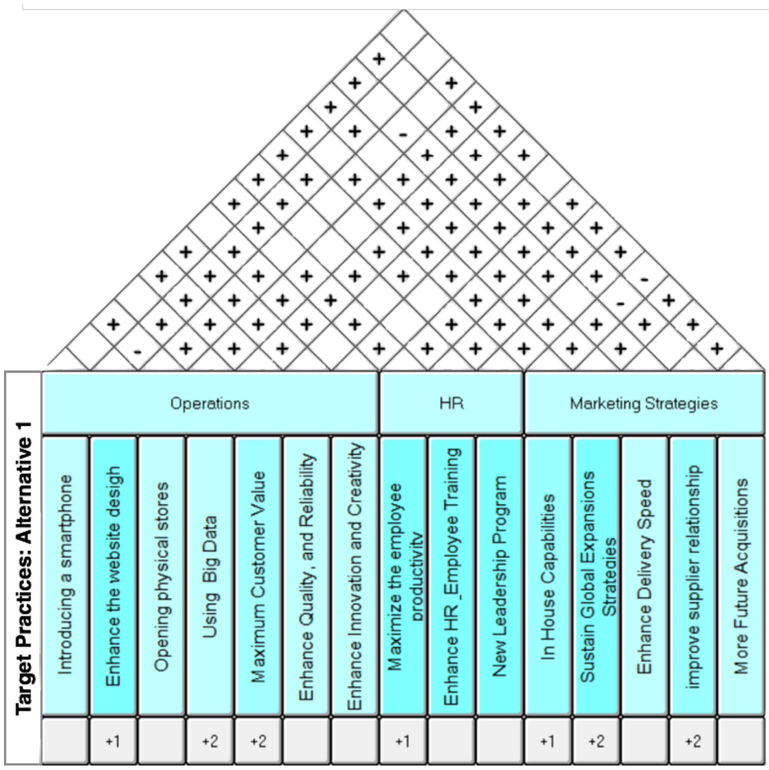


Fig. 10. Amazon Matrix of Change Target Practices.

The future target practices for operations and human resources, along with future marketing strategies represent how Amazon may take advantage of opportunities and mitigate potential threats. From here, we may assess how Amazon's current state affects the desired future state.

The completed MOC for Amazon is shown in Figure 11. Each current practice is compared to every target practice to determine whether or not their interactions will strengthen or interfere with achieving the desired future state, and is indicated by a "+" or a "-" sign within the transition matrix. The completed MOC shows that the positive interactions between the target practices and the current practices are more positive than negative interactions in the transition matrix. This means that the transition will not be so easy and smooth in all leadership styles.

The insights derived from Amazon's MOC give stakeholders and management an idea about the relationship between the current business practices and the proposed future strategies. It is a great insight for better understanding the transition process and they type of undertaking needed to achieve a successful transition.

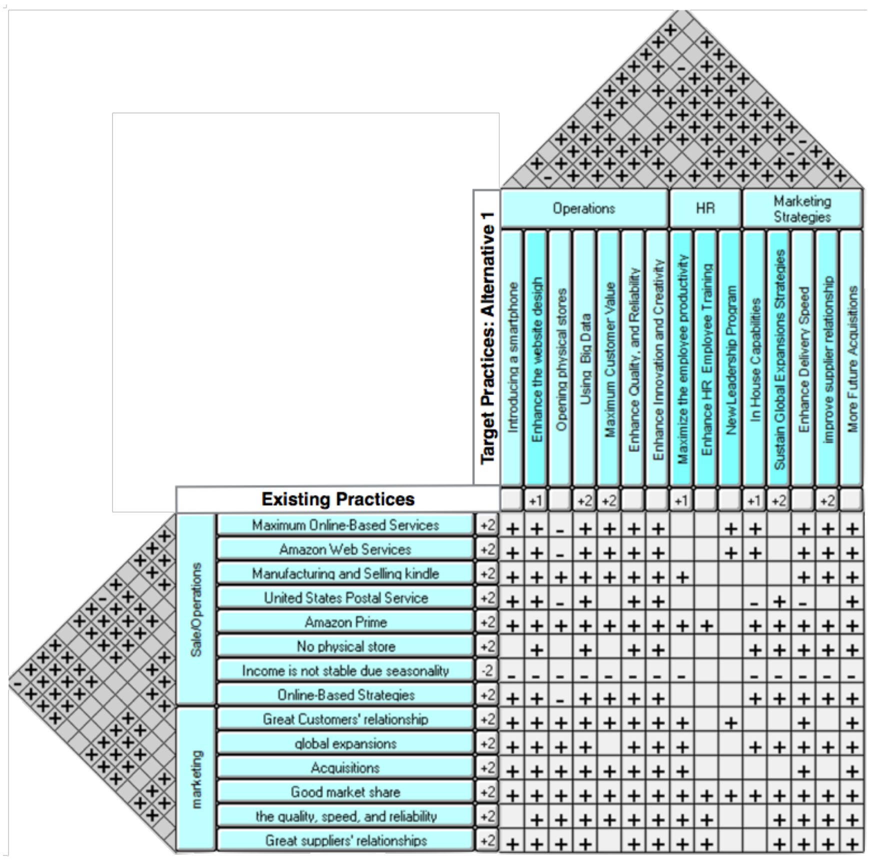


Fig. 11. Amazon Matrix of Change.

Figure 12 depicts the application of big data analytics to Amazon's MOC. Amazon's approach to change deviates from traditional methods. Through Amazon's data management and analytics capabilities they may assess and virtually simulate their current practices and perform benchmarking to identify opportunities for improvement. In fact, the AWS Trust Advisor is now a service that monitors a customer's configurations and then correlates them to known best practices to inform them on existing opportunities for enhancing their performance, security, and cost reduction efforts. However, Amazon engages heavily in physical and virtual experimentation practices associated with big data techniques. In this case, products or services are simulated and deployed similar to a pilot program in order to assess and monitor the usage for either improvement or scrapping of the project. In many cases, this approach is used to strategically lay the groundwork and infrastructure in markets with growth potential. Amazon attempts to address a consumer demand before there is a demand and satisfy a need before there is a need through rapid advancement and deployment of technologies, products, and services. If Amazon's analyses indicate there is potential for gaining market share, they are willing to take the investment risk in deploying new programs knowing full well that some of their investments will pay off and others will not. The philosophy is that valuable knowledge will be acquired from either result. Further Amazon believes that scale is critical in realizing the potential of their business model and thus focuses on growth and expansion.

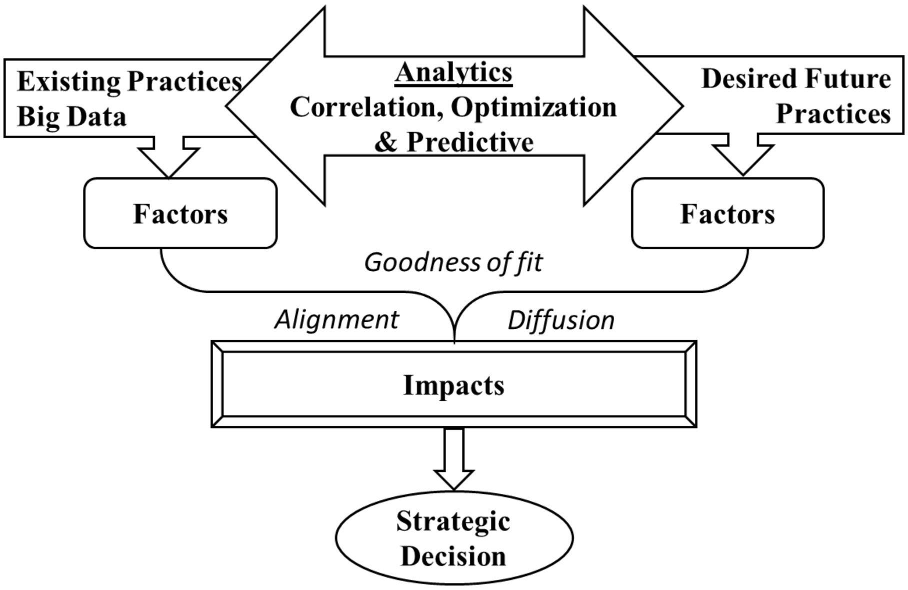


Fig. 12. Matrix of Change Applying Big Data Analytics.

By taking advantage of the wealth of data that Amazon currently houses, analytics may be applied to identify correlations and to generate optimization and prediction information for assessing transition into potential future states. Decision makers stand to gain from this approach as it better informs their strategic decisions through use of accurate and quality data. Further, the breadth of information used in the analysis may provide insights that would not have necessarily been apparent had a limited set of information been employed.

2.2.4 Amazon Strategy Map & Balanced Scorecard

Strategy is about focus and choice, and a strategy map is the cornerstone of strategy formulation. Its function is to outline Amazon's primary strategic objective. To account for the entire set of competitive factors, multiple perspectives are reflected: financial, customer, internal processes and learning and growth. Putting all the perspectives together result in the Strategy Map shown in Figure 13.

The financial perspective concentrates on creating long-term value for shareholders. In Amazon's situation, this should be achieved with Amazon-branded smartphone sales, up-selling retail customers, reducing customer acquisition costs with improved targeting of advertisements, promoting services with higher profit margins, and increasing overall conversion rates; each of these would directly impact the bottom line. In ecommerce, the conversion rate is a crucial KPI (Key Performance Indicator) that represents the ratio of customers browsing products to completed transactions. Enhanced customer experience and customer loyalty has a direct causation correlation with financial objectives, since satisfied customers are more likely to purchase new products, be susceptible to cross-selling recommendations, visit Amazon directly instead of through paid channels, and share their shopping experience with potential customers. These are all key drivers of ecommerce conversion rates.

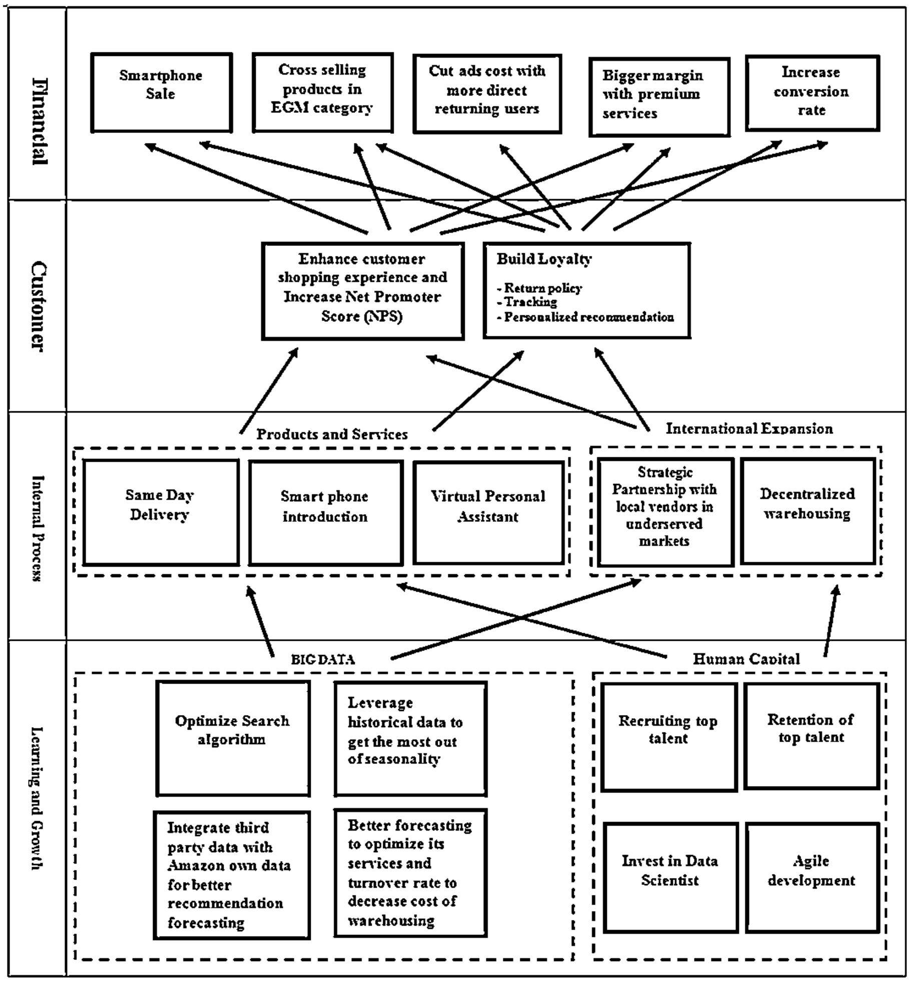


Fig. 13. Amazon Strategy Map.

The customer perspective is linked to the aforementioned financial objectives. It articulates the objectives most relevant to the customer. Material enhancements in the shopping experience are an essential element of the sales process, since they impact customer perception of the brand and, consequently, the future propensity to promote or disparage the brand.

Building loyalty and enhancing the customer experience is the result of Amazon's product offerings, such as a smartphone that would create a captive audience for Amazon's logo and bundled apps/services, superior delivery service like same-day delivery that would create a credible alternative to purchasing products typically sourced in local stores, and a virtual assistant that could answer questions, perform purchases, and arrange delivery through voice commands on mobile devices. An international presence with decentralized warehousing has the potential to grow market share through accessible supply and quick delivery.

The internal processes, grouped as product and services objectives and international expansion objectives, are supported by investment in human capital where Amazon should focus on:

1. Recruiting top talent to create an elite team that supports the internal process objectives such as the launch of new products and services which will require R&D and experienced talent in the field of development of the new products.
2. The retention of the talent will be an essential part for the human capital objective as it would increase the focus on the objective and avoid distraction and knowledge loss from employee turnover
3. Data scientists will be the back bone of the agile teams able to data driven decisions with a higher likelihood of success in supporting the business objectives.
4. While agile development will be the environment/framework where those teams should operate to be able to quickly deploy, learn and reiterate on the products to achieve customer satisfaction and ultimately have the desired financial impact on the business.

Coupled with a strategy map is a balanced score card which provides the measures and targets that support the objectives identified in the strategy map. The balanced score card for Amazon is shown in Table 3.

Table 3. Amazon Balanced Scorecard

Perspective	Objectives	Measures	Targets (2015)
Financial	Smartphone Sales	Number of Smartphones Sold	1.5% of Global Smartphone shipment
	Cross-Selling Products in the EGM category	Number of cross-sold items after each acquisition in EGM	Depends on Amazon historical data of how many people browse "related" products to their initial purchase
	Cut ad costs with more direct returning users	Percent of users returning through direct/free channel after first acquisition	This number will be proportioned with the growth of premium service users
	Bigger Margin on Premium Services	Increase the profit margin and maintain premium services growth	At least 3.2% margin
	Increase conversion rate	Conversion rate on Amazon Market Place	Reach 8% conversion rate to match its peak season
Customer	Enhance customer shopping experience and Increase net Promoter Score (NPS)	Average sessions to buy NPS score	Maintain the lead in the online shopping segment and increase the gap between rivals
	Build Loyalty	Number of returning users with a purchase	No public data
Internal Process	Same Day Delivery	Number of purchases in new product segments where Amazon under-index compared to physical stores	Service the biggest city in every state by the end of 2015 with Amazon fresh
	Develop a Smartphone line	Time to market with first product	Launch by 2015
	Virtual Assistant	Customer engagement with	Launch by 2015

Perspective	Objectives	Measures	Targets (2015)
		the new tool	
	Strategic partnership with vendors in underserved markets	% of revenue increase in currently underserved ma	As needed
	Decentralized warehousing	Number of users served per warehouse in the new markets	Based on pilot result
Learning & Growth	Optimize Search Algorithm	Number of product purchased from 1st page search result	No public data
	Better forecasting to optimize turnover rate	Higher turnover rate	This will depend on product segment
	Leverage historical data for better seasonality prediction	More peak periods per UFI and UFI to UFI (unique feature identifier)	Replicate multiple peak seasons across 2015
	Integrate third party data for better recommendation	Sold items related to events predicted by using the new data set integration	Integrate data of at least one partner in every under-index area of products sale
	Human Capital	Increase of productivity Number of successful new features introduced in A/B testing environment	Target one code rollout a day Increase the number of running experiments per page At least 2 years employee turnover Hire one data scientist for each section of the conversion funnel

The customer shopping experience objective, associated with the customer perspective, is measured by the company NPS (Net Promoter Score) and is described in Table 4.

Table 4. Net Promotor Score Categorization

Category	Description	Score
Promoters	Loyal customers who are brand enthusiasts and will continue to buy and refer others, which positively affects organic growth	9-10
Passives	Users considered as vulnerable to competitive pressures of other companies, especially in an industry with low switching costs.	7-8
Detractors	Customers who may be categorized as unsatisfied customers who may actively damage the brand and slow growth though negative word-of-mouth and social media.	0-6

The NPS is the percentage of customers who wouldn't recommend the company subtracted from the percentage of customers who would promote the company. This methodology is used by leaders as a substitute for the customer satisfaction survey. The interpretation of the net promoter score also depends on the sales stage at which it is placed. For instance, an NPS question asked at the end of the transaction on an ecommerce site better indicates satisfaction with platform rather than customer satisfaction with the product purchase subsequently captured by the product reviews.

Figure 14 depicts the application of big data analytics to Amazon’s strategy map and balanced score card. Amazon’s measures are consistent with their strategic objectives that support market leadership.

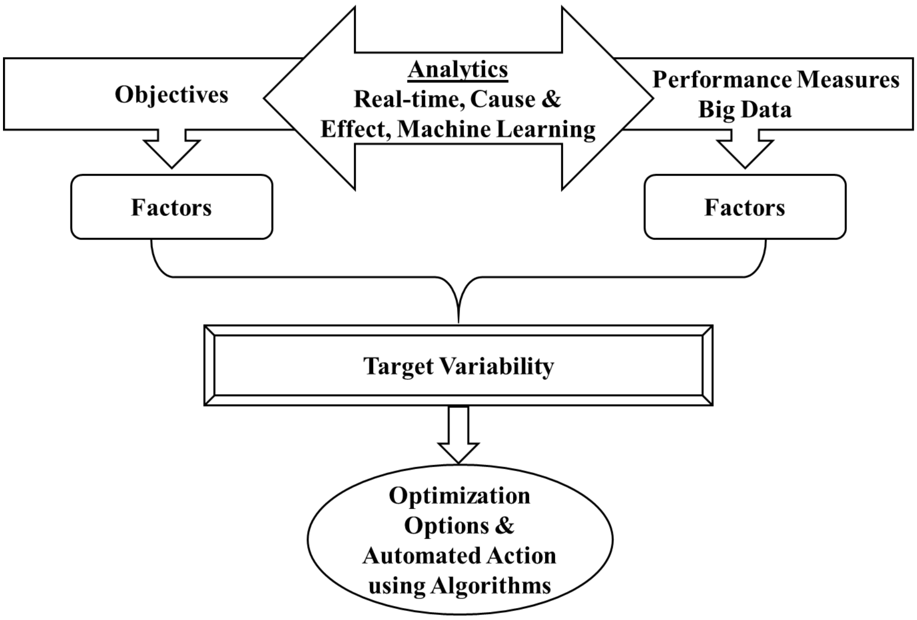


Fig. 14. Strategy Map and Balanced Scorecard Applying Big Data Analytics.

Customer growth, revenue gains, brand strength and the degree to which products and services purchases continue and are repeated by customers are key measures for Amazon's market leadership. Strategically, Amazon seeks to generate revenue through the use of their devices rather than the purchase of devices for better alignment with their customers. Typical industry programs that rely on customers to upgrade products are not building relationships, learning from their customers, and improving the customer's experience in the way that Amazon seeks to accomplish. In fact, real-time analytics, regression modeling, pattern recognition, and machine learning capabilities allow Amazon to institute automated systems that actively seek instances where a customer's experience hasn't met Amazon's standard. In such instances, decision criteria is built into algorithms that will automatically provide customer refunds. This practice is extended into their pre-purchase programs that automatically refund customers the difference between Amazon's guaranteed lowest price and the release price of a product. Again, this is another way that Amazon builds relationships, trust, and loyalty among their customers which is an important factor in increasing their customer base.

Machine learning is at the core of many of Amazon's capabilities to serve their customers efficiently and effectively. Statistical properties of datasets are used to train models in finding patterns within the data through machine learning algorithms. These algorithms are able to quickly optimize models and can easily generate real-time predictions. In 2014 Amazon released the Amazon Echo, which incorporates machine learning based on voice recognition. The device is connected to Amazon's cloud and supported by their web services to provide users with voice controlled access to information, music and more, while continuously learning about its owner and adding more functionality over time based on speech patterns and user preferences.

Big data will be central in guiding new deployments and objectives while acting as a foundation for business decisions in achieving those objectives. By optimizing the search algorithm, Amazon will have the foundation for a good visual assistant as it would mean it was able to answer the customer's "typed" query in an efficient manner which will be the step before converting from the typed query to the "talked" query. Furthermore by leveraging historical data, optimization of the turnover rate would assist with increasing warehousing efficiencies for products with higher turnover rate and ROI. Amazon can also use this information to add efficiency to partners, especially during international expansion which would result in bigger margins for Amazon. By using third party data more accurate forecasting can be achieved in determining which products will be needed in a specific period in a specific area. Such data may include information related to electricity outages, natural disasters, or even medical data from pharmaceutical companies that inform which products sell more, to who and which period.

3 Results

Our research indicates that companies stand to gain from making use of the wealth of data they themselves collect internally and from their customers. Streamlining and optimizing internal and customer interface processes have strong potential for enhancing customer satisfaction, increasing sales, and gaining cost efficiencies. Market insights through customer segmentation, consumer trends, and economic growth can inform a company's strategy development, product and service planning, and resource allocations. Competitors with innovative strategies in cross-cutting industries may be assessed more accurately to identify opportunities that may exist. While machine learning offers enhanced consumer optimization and prediction, a more transparent approach can largely benefit companies such that historical and real-time information may be accessed in a meaningful way by employees to improve the efficiency and effectiveness of their work. Internal transparency can enable companies to gain operational and supply chain efficiencies, allow knowledge base capabilities to significantly grow, and can also better inform decision makers by making more accurate and encompassing information available. Advanced forecasting can make the difference between proactive and reactive operations. Further, well developed forecasting capabilities can enable companies to identify trigger points, align infrastructure and capabilities, and appropriately deploy expansion of products and services. Some companies may even be capable of triggering events that allow growth in certain market sectors.

From the customer perspective, customer and market data can allow companies uniquely meet customer needs and enhance customer satisfaction. Customers are aware that companies have their information, however often times companies are not transparent regarding what they use customer data for and do not show they use the customer's information in a useful way. Customer interfaces, especially customer service, should be a key focus for companies yet are still lacking in many areas of industry. It is frustrating and inconvenient for customers to provide information that they know companies already have. Customer service representatives are all too often not provided with the information they need to effectively deliver customer service.

However, data warehousing and acquisition techniques, along with internal transparency, can provide customer service representatives with the customer information and prescribed solutions needed to effectively and efficiently resolve customer issues.

Given the current technology advances, more and more companies may take advantage of data storage and analytical capabilities offered through third party providers. Also, companies that implement big data analytics may have data that other companies find valuable. If such data is not essential in maintaining a competitive advantage, these companies stand to gain additional profits by partnering with outside companies and allowing access to their data. External data sources that implement big data analytics can allow companies to identify more opportunities and threats than would not otherwise be realized. Additionally, trends in social networks or purchasing behavior can provide can provide much more advanced predictive capabilities for companies that tap into data rich environments.

4 Conclusions

In this paper, an innovative approach has been presented to help managers, stakeholders and advisors to use the big data along with the well-known management tools for the purpose of both enhancing revenue and reducing costs. Our research has several implications and opportunities for future research.

Practical Implications. This study proposed a framework to show and guide organizations how to boost their revenue by helping them understand their businesses through logical steps. The steps of the framework must be accomplished in an order and the big data methodology must be achieved with each steps. The application of big data is used to enhance the result of each phase in this framework. The usage of big data is virtually unlimited; innovation will differentiate between great usage of big data and poor usage of it. As our research shows, a great application of our framework is the use of machine learning to quickly build adaptive models for predictive applications, which may be executed using real-time data streams and scaled to a global level. This is only one example of how companies could leverage big data for not only enhancing decision making tools, but also boosting revenue in a competitive world. As technology and experience with big data applications advance, barriers to implementing big data analytics into business practices is decreasing. Providers of big data services are working toward infrastructures, platforms, and applications that make the capture, management, processing, and analysis of big data possible for businesses without requiring extensive expert knowledge behind such capabilities.

Research Implications. The research shows the emergence of big data analytics in companies worldwide. Integrated frameworks and steps that incorporate big data capabilities such as data acquisition and organization techniques, machine learning, advanced algorithms, and others in support of dynamic and competitive business environments are scant in the academic literature. Incorporating big data into strategic and operational management principles is essential in preparing our future workforce for the work environments they will likely encounter. While traditional management tools provide a great foundation for learning concepts and applying them to business cases, future research is needed to explore ways to incorporate big data analytical methods that support these tools.

We recommend that future research take on an integrated approach from both academic and practical viewpoints. Because today's markets and technologies are rapidly changing and are converging, a solid foundation is needed in this area of research that can accommodate future change while providing practical applications. Without a solid foundation, this line of research runs the risk of generating models for any and every situation and capability with little synergy or practical meaning. We have provided some first steps to such a synergistic approach by providing a framework that couples traditional management tools with big data analytics. Additional research is needed to investigate the types of big data capabilities that can be implemented in achieving business objectives and activities while also considering the level of effort required, along with benefits compared to costs associated with implementation. Future research opportunities exist to explore big data analytical capabilities that can be implemented by companies that do not possess advanced analytical expertise compared to capabilities that require such expertise. Identifying the existing or emerging capabilities that can more easily be implemented and that also create value, can help any company to realize additional benefits. In addition, more research is needed to develop methodologies for aligning company specific objectives with specific big data capabilities for enhancing the ability to achieve these objectives.

Limitations. While we have focused on the potential benefits of big data analytics, there also exists a potential to cause great harm. Three important issues arise concerning big data collection, storage and analysis: privacy, security, and quality of the data. The persons whose data is obtained may perceive such actions to be an invasion of privacy. Further collecting such data may be considered unlawful in some regions and not in others. Data obtained through the internet further complicates this issue. Boundaries are not clear regarding appropriate types of information that should be collected, the means for collecting it, and what the information should or should not be used for. Protecting personal information is another concern associated with big data. Interconnected systems and increased sharing or access to large volumes of personal information can pose vulnerabilities that can result in unauthorized access and distribution of such data. Identity theft and unauthorized transactions are more common today than ever before and companies need to be prepared to protect the data they house along with contingency plans in the event of a security breach. Again, policies for protecting personal information may differ from region to region, further adding to the threat of exposure. And finally, the quality of the data being analyzed is crucial as it can make the difference between advantageous and detrimental strategic decisions and operations. Internal data needs to be integrated in a way such that it is meaningful and appropriate. Too much data can be confusing and too little may not appropriately inform the user. It is important to align company objectives, measures, and targets with architectures that implement big data analytics. This issue becomes even more of concern when obtaining third party data and methods for ensuring the quality of data are underdeveloped.

We recommend that both researchers and practitioners focus on the development of appropriate regulations, policies, and practices to address privacy, security, and quality of data issues associated with the collection, storage, and use of big data. Of course these are complicated issues to address since regulations and policies vary from region to region, and practices are likely to be adapted to specific needs. However these issues pose ethical issues and serious threats to both customers and companies alike. As companies progress toward the creation of digital ecosystems, more effort is needed to identify practical and responsible applications in the emerging world of big data analytics.

5 References

Agievich, V., & Skripkin, K. (2014). Enterprise architecture migration planning using the matrix of change. *Procedia Computer Science, 31*(2nd International Conference on Information Technology and Quantitative Management, ITQM 2014), 231-235. doi:10.1016/j.procs.2014.05.264

Al-Araki, M. (2013). SWOT ananysis revisited throught PEAK-framework. *Journal of Intelligent & Fuzzy Systems, 25*(3), 615-625. doi:10.3233/IFS-120668

Amazon.com, Inc. (2014). *2013 amazon.com annual report*. Seattle: Amazon.com, Inc.

Amazon.com, Inc. (2015). *Amazon*. Retrieved from Amazon.com: www.amazon.com

Bertels, H. M., Koen, P. A., & Elsum, I. (2015). Business models outside the core. *Research Technology Management, 58*(2), 20-29. doi:10.5437/08956308X5802294

Brynjolfsson, E., & Renshaw, A. A. (1997). The matrix of change. *Sloan Management Review, 38*(2), 37-54.

Brynjolfsson, E., & Saunders, A. (2010). *Wired for Innovation: How Information Technology Is Reshaping the Economy*. Cambridge: The Massachusetts Institute of Technology Press.

Callado, A. A., & Jack, L. (2015). Balanced scorecard metrics and specific supply chain roles. *International Journal of Productivity & Performance Management, 64*(2), 288-300. doi:10.1108/IJPPM-05-2014-0071

Chen, P. C., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences, 275*, 314-347. doi:10.1016/j.ins.2014.01.015

Clemen, R. T., & Reilly, T. (2014). *Making Hard Decisions with Decision Tools* (3rd ed.). Mason: South-Western, Cengage Learning.

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review, 84*(1), 98-107.

Davenport, T. H. (2009). Make better decisions. *Harvard Business Review, 87*(11), 117-123.

Davenport, T. H. (2014). How strategists use 'big data' to support internal business decisions, discovery and production. *Strategy & Leadership, 42*(4), 45-50.

Deighton, J., & Kornfeld, L. (2013). *Amazon, Apple, Facebook, and Google*. Boston: Harvard Business School Press.

Dewey, J. P. (2014). *Big Data*. Salem Press Encyclopedia.

Dutta, D., & Bose, I. (2014). Managing a big data project: The case of Ramco Cements Limited. *International Journal of Production Economics, 165*, 293-306. doi:10.1016/j.ijpe.2014.12.032

Elattar, A. A. (2014). A holistic framework for transitional management. Orlando, Fla: University of Central Florida, 2014.

Evans, J. R., & Lindsay, W. M. (2014). *Managing for Quality and Performance Excellence* (9th ed.). Mason: South-Western, Cengage Learning.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2014). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal Of Production Economics, 165*, 234-246. doi:10.1016/j.ijpe.2014.12.031

George, G., Haas, M. R., & Pentland, A. (2014, April). Big data and management. *Academy of Management Journal, 57*(2), 321-326. doi:10.5465/amj.2014.4002.

Gobble, M. M. (2013). Big data: The next big thing in innovation. *Research Technology Management, 56*(1), 64-66. doi:10.5437/08956308x5601005

Hamel, G. (2002). *Leading the Revolution: How to Thrive in Turbulent Times by Making Innovation a Way of Life*. New York: Penguin Group.

Harasim, J., & Klimontowicz, M. (2013). Payment habits as a determinant of retail payment innovations diffusion: The case of Poland. *Journal of Innovation Management, 1*(2), 86-102.

Harvard Business School. (2006). *SWOT Analysis I: Looking Outside for Threats and Opportunities*. Boston: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (2001). *The Strategy Focused Organization: How balanced Scorecard Companies Thrive in the New Business Environment*. Boston: Harvard Business Review Press.

Kusiak, A. (2015). Break through with big data: industrial engineers play a key role in mining numbers for dramatic improvements. *Industrial Engineer, 47*(3), 38-42.

Manyika, J., Chui, M. B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, 1-156.

Martin, M. W., & Schinzinger, R. (2005). *Ethics in Engineering* (4th ed.). New York: McGraw-Hill.

Massachusetts Institute of Technology and the Center for Coordination Science. (2015, 04 01). *Matrix of Change*. Retrieved from MIT Center for Coordination Science. Accessed 28th April 2016. http://ccs.mit.edu/moc/index.html.

Matsudaira, K. (2014). Capturing and structuring data mined from the Web. *Communications of the ACM, 57*(3), 10-11.

Morese, L. C., & Babcock, D. L. (2014). *Managing Engineering and Technology* (6th ed.). Upper Saddle River: Pearson Prentice Hall.

Nair, M. (2004). *Essentials of Balanced Scorecard*. Hoboken: John Wiley & Sons, Inc.

Nicola, S., Pinto Ferreira, E., & Pinto Ferreira, J. J. (2014). A quantitative model for decomposing & assessing the value for the customer. *Journal of Innovation*

Management, 2(1), 104-138.

Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken: John Wiley & Sons.

Pousttchi, K., & Hufenbach, Y. (2014). Engineering the value network of the customer interface and marketing in the data-rich retail environment. *International Journal Of Electronic Commerce, 18*(4), 17-41.

Rabelo, L., Eskandari, H., Shaalan, T., & Helal, M. (2007). Value chain analysis using hybrid simulation and AHP. *International Journal of Production Economics, 105*(2), 536-547.

Ritala, P., Golnam, A., & Wegmann, A. (2014). Coopetition-based business models: The case of Amazon.com. *Industrial Marketing Management, 43*(2), 236-249. doi:10.1016/j.indmarman.2013.11.005

Sahoo, S., Jayapandian, C., Garg, G., Kaffashi, F., Chung, S., Bozorgi, A., . . . Zhang, G. Q. (2014). Heart beats in the cloud: distributed analysis of electrophysiological 'Big Data' using cloud computing for epilepsy clinical research. *Journal Of The American Medical Informatics Association, 21*(2), 263-271.

Yuanhong, C., Zengbiao, Y., & Lin, T. W. (2015). How Zysco uses the balanced scorecard. *Strategic Finance, 97*(1), 27-36.

JIM 4, 1 (2016) 69-91
HANDLE: http://hdl.handle.net/10216/83250

Perspectives to Definition of Big Data: A Mapping Study and Discussion

Ossi Ylijoki[1*], Jari Porras[1]

[1]School of Business and Management, Lappeenranta University of Technology, Finland
ossi.ylijoki@phnet.fi, jari.porras@lut.fi

Abstract. Big data is an emerging research area where common terminology is still evolving. Different perspectives to the research area and terminology exist, but a common definition for big data does not exist. We have performed a systematic mapping study in order to identify different big data definitions and their perspectives. As a result, we present a state-of-the-art review of the current status in big data definitions, discuss the shortcomings of the current definitions, and propose possible solutions for the shortcomings. The paper contributes to the emerging big data research by analyzing current definitions of big data from different perspectives, suggesting enhancement to the terminology as well as pointing out new research avenues. In addition, the article helps new researchers and practitioners to understand what big data is, and bridges the knowledge between theory and practice.

Keywords. Analytics, Big Data, Big Data Definition, Business Model, Datafication, Digitization, Knowledge Management.

1 Introduction

Digitization is a current megatrend, meaning that digital technologies are integrated into our everyday life. The use of digital technologies enables the connection of different services and automation of many processes. Although digitization itself is an important technological (r)evolution, it enables even more fundamental change: datafication. An increasing number of devices and sensors are constantly connected to the Internet. Cameras, mobile phones, tablets, various applications and services running on them produce wide varieties of digital data. This data generation phenomenon is called datafication (Mayer-Schönberger and Cukier, 2013). Lycett (2013) defines datafication as a "sense-making process", which emphasizes the value generation aspect. Digitization and datafication make it possible to capture different situations, actions, or even series of events in the form of data. A vague term "big data" describes the data resulting from datafication. This phenomenon has widespread effects.

As an example, let us consider quadcopters. Amazon and DHL[1], among others, are prototyping these small flying devices for delivering goods to customers. Quadcopters

1 Amazon Prime Air. http://www.amazon.com/b?node=8037720011. Accessed 28th April 2016.
DHL: http://www.theguardian.com/technology/2014/sep/25/german-dhl-launches-first-commercial-drone-delivery-service. Accessed 28th April 2016.

ISSN 2183-0606
http://www.open-jim.org

gather vast volumes of different types of data in real-time (e.g. sensor readings, video and geolocation data) in order to be able to perform their tasks autonomously. They analyze and use the data in many tasks, such as avoiding collisions and orienting their way to the destination. In addition, they synthetize and distribute data. Sending data, such as location and altitude to the command center is essential for the fleet management.

As the fleet of a firm might contain thousands of quadcopters, this represents a real-world big data problem. In general, there are numerous *technical challenges* to conquer for organizations that wish to benefit of big data, see e.g. (Ma et al., 2013; Chen et al., 2014; Kambatla et al., 2014). So far, humans supervise most quadcopter experiments, but due to rapid technical advances, it is obvious that in the near future these little flying machines will become autonomous. Hardware and software vendors are investing heavily in their offerings, so this area is progressing rapidly.

Big data resulting from digitization is seen as a *significant opportunity*, see e.g. (Manyika et al., 2011; Mayer-Schönberger and Cukier, 2013; Schmarzo, 2013; Davenport, 2014). Big data is considered as a key enabler that can be used to generate value in private companies and public organizations. Governments have initiated big data strategies[2]. Examples of the benefits include creating new business opportunities, boosting R&D activities, and supporting decision making (Amatriain, 2013; Mehta et al., 2013; Lee et al., 2014). Quadcopters, among other technological solutions, can be used to save costs and even enable new business models for many organizations, both in the private and the public sector. There are naturally also questioning voices that criticize the big data paradigm and value proposals (see e.g. (boyd and Crawford, 2012; Fox and Do, 2013).

Datafication and big data are disruptive technologies that have widespread implications on the society. Technology vendors, the public sector, private companies, consumers, and policy makers, among others, have interests in the field. Moreover, as the number of stakeholders and parties increases, common understanding of the terminology and concepts becomes more and more important. Unfortunately, big data is a volatile term now. Different definitions of big data can be found in the literature, as well as among practitioners. A (theoretical) definition is a proposal for understanding the meaning of a term. It should be observable, clear (i.e., unambiguous) and complete. Good definitions improve the quality of communication significantly and enable common understanding among participants from different backgrounds. To put it simply, a good definition equals clarity.

The purpose of this article is – considering the broad implications of big data on the

---

2 E.g. European Big Data Value Strategic Research & Innovation Agenda. http://www.nessi-europe.eu/Files/Private/EuropeanBigDataValuePartnership_SRIA__v099%20v4.pdf. Accessed 28th April 2016.
U.S. Big data initiative: https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf. Accessed 28th April 2016.

society, organizations, and individuals – to shed light on the definition of big data. As the method, we use a systematic mapping study. According to Kitchenham (2007), mapping studies are designed to give a broad overview of a research area. Mapping studies have typically broad research questions. Our research questions are:

- What kind of definitions of big data exist in research papers and among practitioners?
- How has the definition of big data evolved?
- How do the definitions reflect the different characteristics and perspectives of big data?

2 Literature Search

Our initial search covered three major reference databases: Scopus, ProQuest, and Web of Science. We considered this as a good starting point, as these databases index a broad range of papers, covering both technical and business fields. Figure 1 gives an overview of the search process. In addition to wide research questions, Kitchenham (2007) suggests that mapping studies should use rather loose search criteria. We searched the databases (title, keywords, abstract) by using ("big data" and "definition") as a search string. All papers written in English and indexed up to 02-Sep-2015 were included in the initial result set. No additional limitations were set. A total of 479 papers were identified. Next, we removed duplicate articles (117).

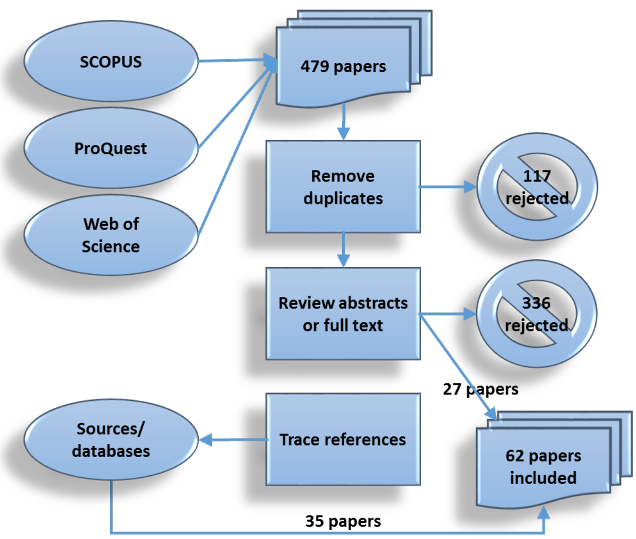


Fig. 1. Search process.

After removing the duplicates, we read the abstracts, and where necessary, the whole text of each of the resulting papers. We categorized the papers by using the following inclusion/exclusion criteria: If the paper contains a definition of big data, include it, otherwise reject it. Due to the loose search criteria, a number of papers defining other things than big data were included in the initial search. Papers that obviously did not meet the eligibility criteria were rejected. If the decision was not clear, we performed a full text review, and the paper was either included or excluded on the basis of the review. Additional 17 papers were excluded because they were either commercial, high-price reports or they could not be found. As a result of this phase, 27 papers were included in the result set.

In the reference-tracking phase, we searched for additional papers on the basis of citations in the included papers (backward snowballing). Possibly interesting references were checked in the article context, and if still promising, they were tracked from databases or web sources, including Google Scholar and various web pages. If the article met the eligibility criteria, it was included. We identified additional 35 papers in this phase.

At the end of the search process phase, we had identified 62 papers that contained a definition of big data. The year-wise distribution of these papers is presented in Figure 2. It seems that although the first definition was presented more than 10 years ago, the discussion of the definition of big data started only a few years back. These papers and their definitions were examined further.

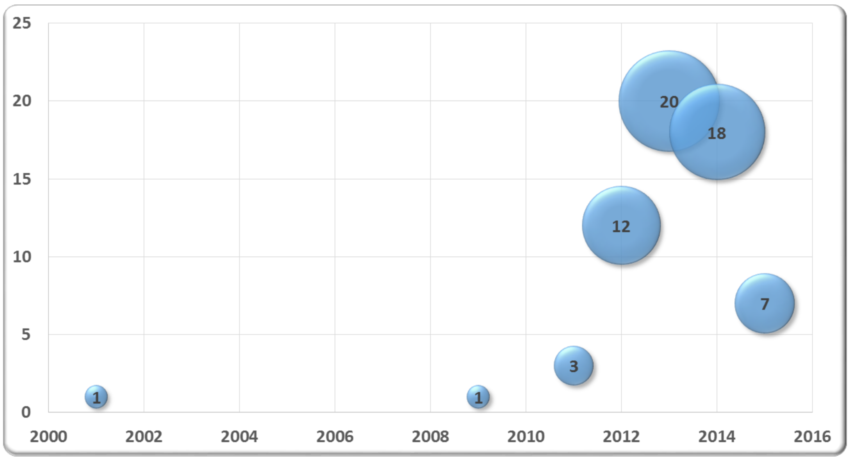


Fig. 2. Year-wise distribution of found papers.

3 Analysis of the Definitions

The first part of the analysis covers the evolution of the definition of big data. Definitions, their existence in time, as well as similarities and differences are presented. This analysis reveals what perspectives (or components) various participants have added to the definition over time. The second part of our analysis identifies gaps between the current definitions and big data value propositions, in order to find out what perspectives are still missing.

3.1 Evolution of the Term Big Data

The term "big data" is not new. It has been used both in research and non-research papers for quite a long time. Back in 1997 it was used in the context of visualizing large data sets (Cox and Ellsworth, 1997). In 1998 it was used in a hardware-related presentation (Mashey, 1998) and also in the data mining context (Weiss and Indurkhya, 1998), and 2003 in combination with statistics (Diebold, 2003). In the beginning, big meant the size and all these sources recognized and referenced big data with the increasing volumes of data. However, year 2001 can be considered as a major milestone in the definition of big data. Laney (2001) described three essential dimensions of big data: *volume*, *velocity* and *variety*. Operating with a swarm of autonomous quadcopters requires the management of high-volume, high-velocity (real-time) data that have many types (variety).

During the following decade, trailblazers like Google and Amazon developed practical big data solutions. These solutions have proved to add value to their businesses. In fact, the trailblazers build their business models on big data solutions. An article published in 2008 in the Wired magazine (Anderson, 2008) aroused public interest in the use of big data and its effects in science. The next significant milestone was 2011, when McKinsey Global Institute and IDC published reports (Gantz and Reinsel, 2011; Manyika et al., 2011) that drew wide public attention to the potential value of big data. Since then a number of newspaper articles, scientific big data papers and books have been published.

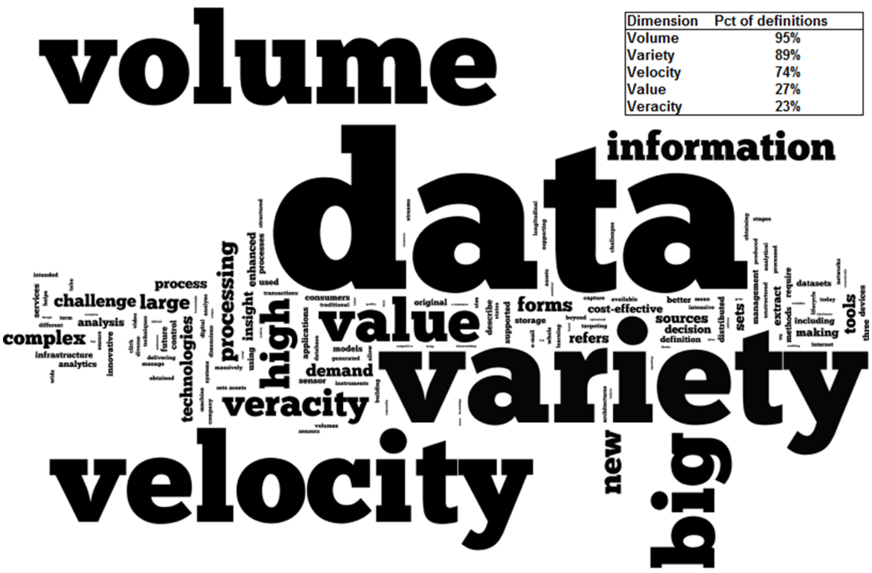


Fig. 3. Most common big data characteristics.

We considered Laney (2001) to be the one to offer the first real definition, although the term big data had been used earlier. In our analysis of the studies, we could not identify references to earlier papers. Naturally, Laney must have been influenced by earlier work, but his paper was the first to introduce the three big data dimensions: volume, variety and velocity. Most of the definitions rely at least partly on the 3V definition by Laney (2001). Figure 3 shows the most common characteristics used in the definitions of big data (see Appendix 1 for details of the definitions). 95% (59 occurrences out of 62) of the papers identified volume as a key characteristic of big data. In addition, the papers considered variety (55 occurrences) and velocity (46) to be typical big data factors. Value (17) and veracity (14) had also caught attention. These five dimensions dominate the current definitions of big data.

The included 62 papers (see Appendix 1 for details) were arranged by their publishing date, and each paper was inspected against previously published definitions. If the paper contained a new definition or added some new elements to the existing definitions, it was considered to be a new definition. This analysis resulted in 17 different definitions. These 17 definitions have similarities in the sense that many of them aim to widen the 3V definition to cover technical and especially business aspects. This is quite a natural consequence with regard to the big data value proposal. However, wide definitions can be problematic, and some essential aspects of big data are still lacking. We will discuss these aspects below. The rest of the papers (45) contained definitions essentially covered in earlier papers. Appendix 1 presents details of the definitions.

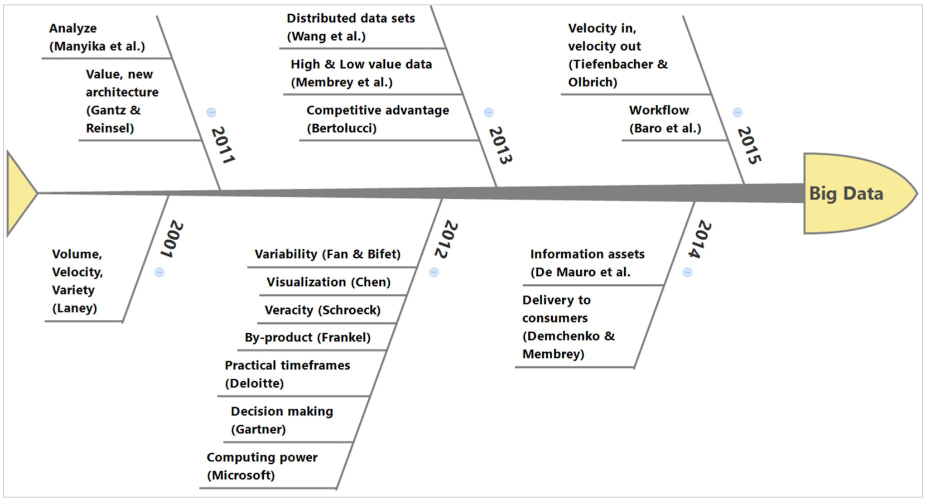


Fig. 4. Evolution of the definition of big data.

The fishbone diagram in figure 4 gives an overview of the evolution. The bones show essential additions of all 17 different definitions, i.e. new aspects or components that each definition adds. Laney (2001) presented the original, so-called 3V definition of big data. The Vs come from volume, velocity and variety. Volume refers to ever-increasing amounts of data. Velocity indicates the need to capture and analyze high-speed or bursts of data in (near) real-time, or else the value may be lost. Variety is related to different types of data, be it structured or non-structured, such as social media posts or a video.

The 3V definition was the de-facto big data standard until 2011, when both Manyika et al. (2011) and Gantz and Reinsel (2011) published their reports. Manyika et al. (2011) emphasize the potential value of big data, but curiously enough, their definition focuses on data volume including only a hint ("analyze") of the value. Also, when compared to Laney (2001), Manyika et al. (2011) have left out velocity and variety. Gantz and Reinsel (2011) include the three Vs, and add value extraction and new architectures. They have also decided to define big data technologies instead of big data. This approach allows them to balance the definition between data, technology and business components with a reasonable logic.

The big data hype was at its peak in the years 2012 and 2013. Several aspects of big data were discussed, such as privacy, security, (business) value, and veracity. We identified seven definitions from 2012 that were either completely new, like the one by Microsoft (2012), or added new components to existing definitions (Gartner, 2012; Schroeck et al., 2012; Fan and Bifet, 2013), and three from the year 2013. After that date we identified four more additions. Two of these later definitions (Demchenko, DeLaat, et al., 2014; Baro et al., 2015) note the importance of delivering the results to consumers. This analysis showed that the evolution of the definition started with data and especially data volumes, and then the discussion shifted to infrastructure topics, followed by the (business) value of data. Finally, more fine-grained aspects, like data delivery and collaboration, appeared.

3.2 Definitions vs. Big Data Value Chain

An interesting question is how the 17 different big data definitions reflect the significant value proposal of big data? Several frameworks explain how data adds value. One of the first of such models is the Virtual Value Creation (VVC) framework presented by Rayport and Sviokla (1995). This framework describes five steps that are required to create value from data: gather, organize, select, synthesize, and distribute (see Figure 5). The steps gather and organize are data-related, and they cover aspects like data acquisition from sensors, integration with other data, and data storing. The steps select, synthesize and distribution depend on data usage. They are activities like filtering data for analysis, or represented as artifacts like analytical models, data visualization, and information delivery tools. Value is expected to increase as data items from various sources are combined to form meaningful information chunks in the VVC process.

A quadcopter reads its current location from the GPS sensor and combines it with the destination information (gather, organize). Based on the analysis, it may take a decision to change its direction (select, synthetize). At frequent intervals, the copter sends data (e.g. location) to the command center (distribute). This simple VVC process adds value, as it enables the copter to work autonomously. However, taking a helicopter view by looking at the whole fleet instead of one quadcopter, it becomes clear that much more value is available. The command center systems gathers data from each of the copters and other sources, e.g. from delivery orders (gather, organize). An analytical model calculates the routes (select, synthetize) and sends instructions (like pick-up and delivery addresses) to each of the copters (delivery). This automated VVC process creates value from the data by producing optimal routes, maximizing the number of deliveries and increasing efficiency.

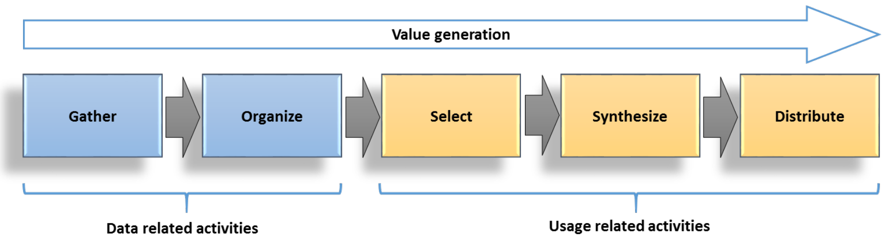


Fig. 5. Virtual value creation process.

Table 1 maps the 17 different definitions to the big data value chain. Together the current big data definitions cover all phases of the value chain. However, most of the definitions cover only parts of the chain. There are two definitions that consider all five phases, those of Demchenko, DeLaat, et al. (2014) and Baro et al. (2015).

Note that the table shows which phases of the value chain the new perspective of each definition emphasizes. This is for clarity: many of the definitions cover also other phases, e.g. Demchenko, DeLaat, et al. (2014) have also covered other steps. However, the new perspective of their definition is the delivery aspect, and therefore only the distribute phase is included in Table 1.

Table 1. Mapping the new perspectives of the big data definitions to the value chain.

Authors	New perspective	Gather (data)	Organize (data)	Select (usage)	Synthesize (usage)	Distribute (usage)
(Laney, 2001)	Volume, velocity, variety	x	x			
(Manyika et al., 2011)	Analyze			x	x	
(Gantz and Reinsel, 2011)	Value, new architecture		x	x	x	
(Microsoft, 2012)	Computing power		x	x	x	
(Gartner, 2012)	Decision making			x	x	
(DeloitteConsulting, 2012)	Practical timeframes		x	x	x	
(Frankel, 2012)	By-product	x				
(Schroeck et al., 2012)	Veracity	x	x	x	x	
(Chen et al., 2012)	Visualization				x	
(Fan and Bifet, 2013)	Variability	x				
(Wang et al., 2013)	Distributed data sets		x			
(Membrey et al., 2013)	High & low value data		x	x		
(Bertolucci, 2013)	Competitive advantage				x	
(Demchenko, DeLaat, et al., 2014)	Delivery to consumers					x
(De Mauro et al., 2015)	Information assets		x			
(Tiefenbacher and Olbrich, 2015)	Velocity in, velocity out	x	x	x	x	
(Baro et al., 2015)	Workflow					x

4 Discussion

As can be seen in the definitions and analysis, big data can mean different things, depending on the selected viewpoint. Some perceive big data as a technical challenge, others view it as a vehicle to increase efficiency or profits. In this section we will show that combining data and its intended usage leads to vague definitions, and consider how the disruptive nature of big data should be taken into account.

4.1 Separate Data and Its Usage

Our analysis revealed that several definitions have logical incoherencies. Value, for example, must be derived from the data by using analytics, there is no value in plain data as such (Ackoff, 1989). Value is also case-dependent. A certain piece of information may be worthless to one company but highly valued by some other firm or in another situation. For example, quadcopter flight details are much more valuable in case of an accident than in a normal situation. This is emphasized by Mayer-Schönberger and Cukier (2013) who state that the value of big data is in the secondary uses of the data. For veracity, analytics is required to determine whether the data is relevant for the planned usage. As important factors as value and veracity are in practice, they do not define the characteristics of big data, but instead they reflect the *usage* of the data. Vague definitions are typically hard to understand as they raise questions that cannot be answered coherently. This will lead to different interpretations and misunderstandings.

The original 3V definition (Laney, 2001) leaved the business effects out. This is one of the main reasons why many new definitions have emerged. Both technology vendors and enterprises have an interest to add a value proposition. Companies see big data as a vehicle to gain value, vendors naturally like to justify the costs of their offerings with potential benefits. A natural tendency would be to add a value component to the definition. However, as discussed above, value is not a characteristic of data. Definitions should be clear and unambiguous. Therefore, adding data usage to the definition is not a good idea, as the definition would become unambiguous, and coherency would be lost.

Our suggestion to the problem is that the *data and its usage should be separated*. Data is similar to oil: when combined with data management and analytics processes it provides organizations with value. Analytics and data usage are of course essential elements in successful big data exploitation. However, from the definition point of view, combining data and its usage is like combining oil and engine into one single definition. Separating big data from its intended usage clarifies the inconsistencies of the definitions and helps us to understand the plain characteristics of big data. As the purpose of data usage is to realize the potential value of the data, we propose the term *big data insights* to be used in any context in data usage -related activities (see also figure 5).

4.2 Other Perspectives – Big Data as a Phenomenon

In addition to technical and value aspects, scholars have focused on several other perspectives to big data, such as privacy, security (Altshuler, 2011; Berghel, 2013; Lu et al., 2014), and policy-making (Keen et al., 2013; Blume et al., 2014; Truyens and Van Eecke, 2014). None of the current definitions of big data consider these. These aspects are not characteristics of big data; we do not suggest that these aspects should be included in the definition. Instead, they are aspects that help to understand big data as a phenomenon. Moreover, these perspectives are important, as failing to consider them can drive an organization to difficulties.

Another, even more important aspect is that the current definitions neglect the disruptive nature of big data. On the basis of the literature it seems obvious that in the future, big data will have significant impacts on businesses (Manyika et al., 2011; Schmarzo, 2013; Davenport, 2014). Big data is seen as a technology that can have huge impacts on most industries and enterprises. Data-driven companies can achieve significant benefits (McAfee et al., 2012), but transformational business changes (Dehning et al., 2003) are required to achieve full competitive advantage from big data. The impact of big data will be significant, but the nature of the change is even more important. The effects of big data on firms, ecosystems and industries will be disruptive (Earley, 2014; Fan and Gordon, 2014; Kim et al., 2014). Industry structures are changing, and new business opportunities are emerging. On the other hand, this means that also competitors may be able to invent new business models, not to speak of new entrants, which will increase the turbulence effectively. The impacts of big data may – and will – be positive for some organizations, negative for others. Due to the disruptive nature of big data, companies must review their business models in order to reveal possible threats and opportunities. Moreover, as the disruptive drivers are technological by nature, these technologies and their potential effects must be linked with strategy.

We suggest that a *new definition for big data as a phenomenon should be considered*. For clarity and coherency, the definition of big data should cover only data and data management aspects (like the 3V definition). The phenomenon of big data is a broad concept that deserves a definition of its own. Instead of defining big data, the definition should consider several important perspectives of it. In our opinion, this definition should include the disruptive nature and strategic importance of the phenomenon. Adding these elements would help managers to understand the importance of the matter. This opens a new research avenue. Discussing and defining the nature and relations between various perspectives would build understanding of the broader context of big data, big data as a phenomenon.

5 Conclusions

Our aim was to shed light to the concept of big data, especially from the following viewpoints:

- What kind of definitions of big data exist in research papers and among practitioners?

- How has the definition of big data evolved?
- How do the definitions reflect the different characteristics and perspectives of big data?

A systematic mapping study was conducted in order to find answers to these questions. We made a search in major reference databases, search engines, and web sources containing both technical and business topics. A total of 62 sources were included in the result set. With regard to our research questions, we chose a broad search strategy in order to cover a wide range of possible sources. We identified 17 different definitions of big data that together presented a clear picture of the current situation and evolution of the definition, thus providing answers to our first and second research questions. We also compared the current definitions with various characteristics of big data. We found that the current definitions do not cover several perspectives that are discussed among big data scholars and practitioners, which answers our third research question. In addition, we identified several logically incoherent definitions. This clouds the matter further, as these definitions raise new questions, which will typically lead to ambiguous answers.

5.1 Results

This study revealed 17 different big data definitions from 62 relevant source papers. Each of the papers was analyzed against previously published definitions. If the paper contained a new definition or added some new elements to the existing definitions, it was considered as a new definition. The key contributions of this study are:

- Although there are various opinions on what big data is, the 3V definition by Laney (2001) contains three dimensions (volume, velocity, variety), which are common to most definitions. In addition to these dimensions, many definitions include technical parts and components related to the intended usage of the data, such as analysis or decision-making.
- Many of the definitions are logically inconsistent, which is one reason for the vagueness of the term big data. A typical flaw is to include both the data and its intended usage in the definition. We suggest that they should be separated. The term big data should cover data-related aspects, whereas a new term *big data insights* should be used when discussing data usage-related activities.
- The current definitions do not consider several important aspects of the big data phenomenon, such as security and privacy, or its disruptive nature. These are not characteristics of big data, but they are important factors of the big data phenomenon that both scholars and practitioners must consider. We suggest that a new definition for big data as a phenomenon should be developed.

In addition, this study bridges the knowledge between theory and practice. We have presented the history and the state-of-the-art of the definition of big data. This will help new researchers and practitioners to understand the different perspectives of big data, as well as the limitations of the current definitions. Therefore, we hope that this paper will also stimulate discussion about the terminology and help parties coming from different backgrounds to understand each other and communicate their reasoning clearly.

5.2 Limitations

We recognize that an uncountable number of various definitions of big data exist in the “Internet jungle”, e.g. in blog postings and discussion forums. However, due to limited resources, identifying and analyzing all or even most of them would be impossible, and therefore we have filtered blogs and forums out. Another limitation is that we have excluded all non-English language sources.

5.3 Suggestions for Further Studies

There are several possible topics for further studies, including the following. It is clear that there is a need to develop the terminology and taxonomy further (including related terms, such as big data analytics, big data phenomenon, and veracity) in order to create common understanding of the key concepts and their relationships in the area of big data. Another interesting research avenue would be to investigate the effects of big data on organizations' business models or decision-making processes, organizational structures, and culture.

6 References

Ackoff, R.L., 1989. From Data to Wisdom. Journal of Applied Systems Analysis 16, 3–9.

Akerkar, R., 2014. Analytics on Big Aviation Data: Turning Data Into Insights. International Journal of Computer Science and Applications 11, 116–127.

Altshuler, M. Y.; Aharony N.; Pentland A.; Elovici Y.; Cebrian, 2011. Stealing Reality: When Criminals Become Data Scientists (or Vice Versa). IEEE Intelligent Systems 26, 22–30.

Amatriain, X., 2013. Beyond data: from user information to business value through personalized recommendations and consumer science. Proceedings of the 22nd ACM international conference on information & knowledge management 2201–2208.

Anderson, C., 2008. The end of theory: The data deluge makes the scientific method obsolete. Wired Magazine, 23-Jun-2008.

Ashraf, J., Hussain, O.K., Hussain, F.K., 2015. Making sense from big RDF data: OUSAF for measuring ontology usage. Software: Practice and Experience 45, 1051–1071.

Balar, A., Malviya, N., Prasad, S., Gangurde, A., 2013. Forecasting consumer behavior with innovative value proposition for organizations using big data analytics. Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on 1–4.

Baro, E., Degoul, S., Beuscart, R., Chazard, E., 2015. Toward a Literature-Driven Definition of Big Data in Healthcare. BioMed Research International 2015, 1–9.

Benjamins, V.R., 2014. Big Data: from Hype to Reality? Proceedings of the 4th

International Conference on Web Intelligence, Mining and Semantics (WIMS14) 2.

Berghel, H., 2013. Through the PRISM darkly. Computer 46, 86–90.

Bertolucci, J., 2013. Big Data: A Practical Definition. Informationweek - Online.

Blume, G., Scott, T., Pirog, M., 2014. Empirical Innovations in Policy Analysis. Policy Studies Journal 42, S33–S50.

boyd, danah, Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society 15, 662–679.

Cackett, D., 2013. Information Management and Big Data: A Reference Architecture. Oracle Corporation. Accessed 28th April 2016. http://www.oracle.com/technetwork/topics/entarch/articles/info-mgmt-big-data-ref-arch-1902853.pdf.

Chen, H., Chiang, R.H., Storey, V.C., 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly 36, 1165–1188.

Chen, M., Mao, S., Liu, Y., 2014. Big data: A survey. Mobile Networks and Applications 19, 171–209.

Collins, E., 2014. Intersection of the cloud and big data. IEEE Cloud Computing 1, 84–85.

Cox, M., Ellsworth, D., 1997. Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th conference on Visualization'97 235–ff.

Cuzzocrea, A., Song, I.-Y., Davis, K.C., 2011. Analytics over large-scale multidimensional data: the big data revolution! Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP 101–104.

Davenport, T., 2014. Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press.

De Mauro, A., Greco, M., Grimaldi, M., 2015. What is big data? A consensual definition and a review of key research topics. AIP Conference Proceedings 1644, 97–104.

Dehning, B., Richardson, V.J., Zmud, R.W., 2003. The value relevance of announcements of transformational information technology investments. Mis Quarterly 27, 637–656.

DeloitteConsulting, 2012. The insight economy: Big data matters—except when it doesn't. Deloitte. Accessed 28th April 2016. http://www2.deloitte.com/content/dam/Deloitte/global/Documents/Deloitte-Analytics/dttl-analytics-us-ba-insight-economy-10012012.pdf.

Demchenko, Y., DeLaat, C., Membrey, P., 2014. Defining architecture components of the Big Data Ecosystem. Collaboration Technologies and Systems (CTS), 2014 International Conference on 104–112.

Demchenko, Y., Grosso, P., De Laat, C., Membrey, P., 2013. Addressing big data issues in scientific data infrastructure. Collaboration Technologies and Systems (CTS), 2013 International Conference on 48–55.

Demchenko, Y., Gruengard, E., Klous, S., 2014. Instructional Model for Building Effective Big Data Curricula for Online and Campus Education. Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on 935–941.

Demchenko, Y., Ngo, C., Laat, C. de, Membrey, P., Gordijenko, D., 2014. Big Security for Big Data: Addressing Security Challenges for the Big Data Infrastructure, in: SecureDataManagement. Springer, pp. 76–94.

Diebold, F.X., 2003. Big Data'Dynamic factor models for macroeconomic measurement and forecasting, in: Advances EconomicsEconometricsTheory Applications, Eighth World CongressEconometricSociety(edited M. Dewatripont, LP HansenS.Turnovsky). pp. 115–122.

Earley, S., 2014. The Digital Transformation: Staying Competitive. IT Professional 16, 58–60.

Emani, C.K., Cullot, N., Nicolle, C., 2015. Understandable Big Data: A survey. Computer Science Review 17, 70–81.

EMC, 2012. Big Data-as-a-Service. EMC Solutions Group. Accessed 28th April 2016. http://www.emc.com/collateral/software/white-papers/h10839-big-data-as-a-service-perspt.pdf.

Fan, W., Bifet, A., 2013. Mining big data: current status, and forecast to the future. ACM sIGKDD Explorations Newsletter 14, 1–5.

Fan, W., Gordon, M.D., 2014. The power of social media analytics. Communications of the ACM 57, 74–81.

Ferrando-Llopis, R., Lopez-Berzosa, D., Mulligan, C., 2013. Advancing value creation and value capture in data-intensive contexts. Big Data, 2013 IEEE International Conference on 5–9.

Fox, S., Do, T., 2013. Getting real about Big Data: applying critical realism to analyse Big Data hype. International Journal of Managing Projects in Business 6, 739–760.

Frankel, D.A., 2012. Big Data and Risk Management. Risk Management.

Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35, 137–144.

Gantz, J., Reinsel, D., 2011. Extracting value from chaos (No. 1142), IDC iview. IDC. Accessed 28th April 2016. http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf.

Gardner, E., 2013. The HIT approach to big data. Health data management 21, 34–36.

Gartner, 2012. Gartner IT Glossary - Big Data. Accessed 28th April 2016. http://www.gartner.com/it-glossary/big-data.

Gerhardt, B., Griffin, K., Klemann, R., 2012. Unlocking value in the fragmented world of big data analytics, Cisco Internet Business Solutions Group, June. Cisco Internet Business Solutions Group (IBSG).

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., Khan, S.U., 2015. The rise of "big data" on cloud computing: review and open research issues.

Information Systems 47, 98–115.

Hu, H., Wen, Y., Chua, T.-S., Li, X., 2014. Toward scalable systems for big data analytics: A technology tutorial. Access, IEEE 2, 652–687.

Jacobs, A., 2009. The pathologies of big data. Communications of the ACM 52, 36–44.

Jin, X., Wah, B.W., Cheng, X., Wang, Y., 2015. Significance and challenges of big data research. Big Data Research 2, 59–64.

Kambatla, K., Kollias, G., Kumar, V., Grama, A., 2014. Trends in big data analytics. Journal of Parallel and Distributed Computing 74, 2561–2573.

Keen, J., Calinescu, R., Paige, R., Rooksby, J., 2013. Big data+ politics= open data: The case of health care data in England. Policy & Internet 5, 228–243.

Kim, G.-H., Trimi, S., Chung, J.-H., 2014. Big-data applications in the government sector. Communications of the ACM 57, 78–85.

Kitchenham, B., 2007. Guidelines for performing systematic literature reviews in software engineering (No. EBSE-2007-01), Technical report, Ver. 2.3 EBSE Technical Report. EBSE. Keele University.

Lamont, J., 2012. Big data has big implications for knowledge management. KM World 21, 8–11.

Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. META Group Research Note 6, 70.

Lee, J., Kao, H.-A., Yang, S., 2014. Service innovation and smart analytics for industry 4.0 and big data environment. Procedia CIRP 16, 3–8.

Li, T.Z., Wang, S.H., Ma, J., 2014. Study on Fair Definitions and Application Modes of Big Data, in: AppliedMechanics Materials. Trans Tech Publ, pp. 606–613.

Lin, P.P., 2014. What CPAs Need to Know about Big Data. The CPA Journal 84, 50.

Lu, R., Zhu, H., Liu, X., Liu, J.K., Shao, J., 2014. Toward efficient and privacy-preserving computing in big data era. Network, IEEE 28, 46–50.

Lycett, M., 2013. Datafication': Making sense of (big) data in a complex world. European Journal of Information Systems 22, 381–386.

Ma, M., Wang, P., Chu, C.-H., 2013. Data management for internet of things: challenges, approaches and opportunities. Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing 1144–1151.

Madden, S., 2012. From databases to big data. IEEE Internet Computing 16, 4–6.

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.

Mashey, J.R., 1998. Big Data... and the Next Wave of InfraStress. Accessed 28th April 2016. http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf.

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., Davenport, T.H., Patil, D., Barton, D., 2012. Big data: The Management Revolution. Harvard Business Review 90, 61–67.

Mehta, S., Pimplikar, R., Singh, A., Varshney, L.R., Visweswariah, K., 2013. Efficient multifaceted screening of job applicants. Proceedings of the 16th International Conference on Extending Database Technology 661–671.

Membrey, P., Chan, K.C., Demchenko, Y., 2013. A disk based stream oriented approach for storing big data. Collaboration Technologies and Systems (CTS), 2013 International Conference on 56–64.

Meng, L., Meng, L., 2014. Application of Big Data in Higher Education, in: 2ndInternationalConference TeachingComputationalScience. Atlantis Press.

Microsoft, 2012. The Big Bang: How the Big Data Explosion Is Changing the World. Accessed 28th April 2016. http://news.microsoft.com/2013/02/11/the-big-bang-how-the-big-data-explosion-is-changing-the-world/.

O'Leary, D.E., 2013. Artificial intelligence and big data. IEEE Intelligent Systems 28, 96–99.

Pandey, S., Tokekar, V., 2014. Prominence of mapreduce in big data processing. Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on 555–560.

Pospiech, M., Felden, C., 2013. A Descriptive Big Data Model Using Grounded Theory. Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on 878–885.

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

Rayport, J.F., Sviokla, J.J., 1995. Exploiting the virtual value chain. Harvard business review 73, 75.

Richards, G., 2014. Let us clarify what we mean by Big Data; We first need a clear definition. The Ottawa Citizen, 02-Aug-2014.

Sagiroglu, S., Sinanc, D., 2013. Big data: A review. Collaboration Technologies and Systems (CTS), 2013 International Conference on 42–47.

Schmarzo, B., 2013. Big Data: Understanding how data powers big business. John Wiley & Sons.

Schneider, R.D., 2012. Hadoop for dummies, Special Edition, John Willey & sons. John Wiley & Sons Canada, Ltd.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P., 2012. Analytics: The real-world use of big data, IBM Global Business Services, Somers. IBM. Accessed 28th April 2016. https://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf.

Spiess, J., T'Joens, Y., Dragnea, R., Spencer, P., Philippart, L., 2014. Using big data to improve customer experience and business performance. Bell Labs Technical Journal 18, 3–17.

Stonebraker, M., Robertson, J., 2013. {Big Data is'Buzzword Du Jour;'CS Academics'Have the Best Job'}. Commun. ACM 56, 10–11.

TataConsultancyServices, 2013. The Emerging Big Returns on Big Data. Tata Consultancy Services. Accessed 28th April 2016. http://www.tcs.com/SiteCollectionDocuments/Trends_Study/TCS-Big-Data-Global-Trend-Study-2013.pdf.

TheIrishTimes, 2013. Get to grips with the streams of data you generate. Irish Times.

Tiefenbacher, K., Olbrich, S., 2015. Increasing the Value of Big Data Projects–Investigation of Industrial Success Stories. System Sciences (HICSS), 2015 48th Hawaii International Conference on 294–303.

Truyens, M., Van Eecke, P., 2014. Legal aspects of text mining. Computer Law & Security Review 30, 153–170.

Vossen, G., 2014. Big data as the new enabler in business and other intelligence. Vietnam Journal of Computer Science 1, 3–14.

Wang, W., Zhou, X., Zhang, B., Mu, J., 2013. Anomaly detection in big data from UWB radars. Security and Communication Networks 8, 2469–2475.

Ward, J.S., Barker, A., 2013. Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.

Weiss, S.M., Indurkhya, N., 1998. Predictive data mining: a practical guide. Morgan Kaufmann.

Wyber, R., Vaillancourt, S., Perry, W., Mannava, P., Folaranmi, T., Celi, L.A., 2015. Big data in global health: improving health in low-and middle-income countries. Bulletin of the World Health Organization 93, 203–208.

Xin, N.Y., Ling, L.Y., 2013. How we could realize big data value. Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013 2nd International Symposium on 425–427.

Zhang, J., Chen, Y., Li, T., 2013. Opportunities of innovation under challenges of big data. Fuzzy Systems and Knowledge Discovery (FSKD), 2013 10th International Conference on 669–673.

Appendix 1 – Included Papers

Definitions of big data sorted by date. The Definition column contains either a direct quotation from the paper or essential parts of the given/referenced definition. The New perspective column indicates what new component or aspect the definition has added to the previous ones.

Authors	Date	Definition	New perspective
(Laney, 2001)	2001-Feb	"E-commerce, in particular, has exploded data management challenges along three dimensions: volumes, velocity, and variety."	volume, velocity, variety
(Jacobs, 2009)	2009-Aug	“data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time.”	-
(Manyika et al., 2011)	2011-May	“Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze."	analyze
(Gantz and Reinsel, 2011)	2011-Jun	"Big Data technologies describe a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis."	value, new architecture
(Cuzzocrea et al., 2011)	2011-Oct	”enormous amounts of unstructured data produced by high-performance applications falling in a wide and heterogeneous family of application scenarios”	-
(Microsoft, 2012)	2012-Feb	"Big data is the term increasingly used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information."	computing power
(Lamont, 2012)	2012-Apr	Volume, variety, velocity	-
(Madden, 2012)	2012-May	”...it means data that’s too big, too fast, or too hard for existing tools to process.”	-
(Gartner, 2012)	2012-Jun	"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."	decision making
(Gerhardt et al., 2012)	2012-Jun	Volume, variety, velocity	-
(EMC, 2012)	2012-Jul	"Big data refers to new scale-out architecture that address these needs. [quickly processing more	-

Authors	Date	Definition	New perspective
		varied, more complex, and less structured data] Big data is fundamentally about massively distributed architectures and massively parallel processing, using commodity building blocks to manage and analyze the data."	
(Schneider, 2012)	2012-Sep	Volume, variety, velocity	-
(DeloitteConsulting, 2012)	2012-Oct	"Big data generally refers to datasets so large and complex they create significant challenges for traditional data management and analysis tools in practical timeframes."	practical timeframes
(Frankel, 2012)	2012-Oct	”the volumes of structured and unstructured data produced as a by-product of operating a company”	by-product
(McAfee et al., 2012)	2012-Oct	Volume, variety, velocity	-
(Schroeck et al., 2012)	2012-Oct	"...characterizing three dimensions of big data – “the three Vs:” volume, variety and velocity. And while they cover the key attributes of big data itself, we believe organizations need to consider an important fourth dimension: veracity."	veracity
(Chen et al., 2012)	2012-Dec	big data and big data analytics have been used to describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies.	visualization
(Fan and Bifet, 2013)	2012-Dec	Volume, Velocity, Variety, Value, Variability	variability
(Cackett, 2013)	2013-Feb	Volume, velocity, variety, value	-
(Gardner, 2013)	2013-Mar	”Volume, velocity, variety”	-
(O’Leary, 2013)	2013-Mar	”Big data isn’t just volume, variety, and velocity, though; it’s volume, variety, and velocity at scale.”	-
(Provost and Fawcett, 2013)	2013-Mar	For this article, we will simply take big data to mean datasets that are too large for traditional data-processing systems and that therefore require new technologies.	-
(TataConsultancyServices,	2013-Mar	Volume, variety, velocity	-

Authors	Date	Definition	New perspective
2013)			
(Wang et al., 2013)	2013-Mar	“ Big Data refers to large, diverse, complex, longitudinal, and distributed data sets generated from instruments, sensors, Internet transactions, e-mail, video, click streams, and other digital sources available today and in the future”	distributed data sets
(Demchenko et al., 2013)	2013-May	”...we intend to propose wider definition of Big Data as 5 Vs: Volume, Velocity, Variety and additionally Value and Veracity.”	-
(Membrey et al., 2013)	2013-May	"Extensions to the (3V) model that take Value into account are then proposed and discussed. … However recording the data does not bring any value to the company. It only becomes valuable once that data is used or processed. ... High Value Data (HVD) is data that has a known benefit from its storage. ... Low Value Data (LVD) is data that is stored in the anticipation that value will be drawn from it in the future."	High & low value data
(Sagiroglu and Sinanc, 2013)	2013-May	Volume, variety, velocity	-
(Zhang et al., 2013)	2013-Jul	Volume, velocity, variety, value	-
(Bertolucci, 2013)	2013-Aug	Big data is about "building new analytic applications based on new types of data, in order to better serve your customers and drive a better competitive advantage."	competitive advantage
(Ward and Barker, 2013)	2013-Sep	"Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning."	-
(Stonebraker and Robertson, 2013)	2013-Sep	In summary, big data can mean big volume, big velocity, or big variety.	-
(Ferrando-Llopis et al., 2013)	2013-Oct	Volume, velocity, variety, veracity	-
(TheIrishTimes, 2013)	2013-Nov	”A simple definition is that it gives organisations insights into data which they don't already have and does that in a way that helps them improve their operational efficiency and helps them make better decisions."	-
(Vossen, 2014)	2013-Nov	Volume, velocity, variety, veracity	-

Authors	Date	Definition	New perspective
(Balar et al., 2013)	2013-Dec	Volume, variety, velocity	-
(Xin and Ling, 2013)	2013-Dec	Volume	-
(Pospiech and Felden, 2013)	2013-Dec	Volume, variety, velocity	-
(Chen et al., 2014)	2014-Jan	Volume, variety, velocity	-
(Spiess et al., 2014)	2014-Feb	Volume, variety, velocity	-
(Ashraf et al., 2015)	2014-Apr	Volume, velocity, variety, value, veracity	-
(Pandey and Tokekar, 2014)	2014-Apr	Volume, variety, velocity	-
(Blume et al., 2014)	2014-May	Volume	-
(Collins, 2014)	2014-May	Volume, variety, velocity	-
(Demchenko, DeLaat, et al., 2014)	2014-May	"Big Data (Data Intensive) Technologies are targeting to process high-volume, high-velocity, high-variety data (sets/assets) to extract intended data value and ensure high-veracity of original data and obtained information that demand cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control; all of those demand (should be supported by) new data models (supporting all data states and stages during the whole data lifecycle) and new infrastructure services and tools that allow obtaining (and processing) data from a variety of sources (including sensor networks) and delivering data in a variety of forms to different data and information consumers and devices."	delivery to consumers
(Demchenko, Ngo, et al., 2014)	2014-May	Volume, velocity, variety, value, veracity	-
(Benjamins, 2014)	2014-Jun	Volume, variety, velocity	-
(Hu et al., 2014)	2014-Jun	Volume, velocity, variety, value	-

Authors	Date	Definition	New perspective
(Li et al., 2014)	2014-Jun	" Big Data refers to large, diverse, complex, longitudinal, and distributed data sets generated from instruments, sensors, Internet transactions, e-mail, video, click streams, and other digital sources available today and in the future"	-
(Lu et al., 2014)	2014-Jul	Volume, variety, velocity	-
(Meng and Meng, 2014)	2014-Jul	Volume, velocity, variety, value	-
(Richards, 2014)	2014-Aug	Big Data is commonly defined as data that cannot be processed by standard database systems.	-
(De Mauro et al., 2015)	2014-Sep	"Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value."	Information assets
(Lin, 2014)	2014-Nov	Volume, variety, velocity	-
(Akerkar, 2014)	2014-Dec	"Using big volume, big velocity, big variety data asset to extract value (insight and knowledge), further confirm veracity (quality and trustworthiness) of the original data and the acquired information, that demand cost-effective, novel forms of data and information processing for enhanced insight, decision making, and processes control. Additinally, those demands are supported by new data models and new infrastructure services and tools which are able to procure and process data from a variety of sources and deliver data in a variety of forms to several data and information consumers and devices."	-

Authors	**Date**	**Definition**	**New perspective**
(Demchenko, Gruengard, et al., 2014)	2014-Dec	"Big Data (Data Intensive) Technologies are targeting to process high-volume, high-velocity, high-variety data (sets/assets) to extract intended data value and ensure highveracity of original data and obtained information that demand cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control; all of those demand (should be supported by) new data models (supporting all data states and stages during the whole data lifecycle) and new infrastructure services and tools that allow obtaining (and processing) data from a variety of sources (including sensor networks) and delivering data in a variety of forms to different data and information consumers and devices."	-
(Hashem et al., 2015)	2015-Jan	Volume, velocity, variety, value	-
(Tiefenbacher and Olbrich, 2015)	2015-Jan	Volume, variety, velocity (in), visibility, veracity, virtue (= value), velocity (out)	velocity in, velocity out
(Baro et al., 2015)	2015-Feb	"Volume: Log(n * p) ≥ 7" (n=statistical individuals, p=nbr of variables) Properties: "Great variety, High velocity, Challenge on veracity, Challenge on all aspects of the workflow, Challenge on computational methods, Challenge on extracting meaningful information, Challenge on sharing data, Challenge on finding human experts"	workflow
(Jin et al., 2015)	2015-Feb	Volume, velocity, variety, value, veracity	-
(Wyber et al., 2015)	2015-Mar	Volume, velocity, variety, veracity	-
(Gandomi and Haider, 2015)	2015-Apr	Volume, variety, velocity	-
(Emani et al., 2015)	2015-May	Volume, velocity, variety, veracity	-

JIM 4, 1 (2016) 114-136
HANDLE: http://hdl.handle.net/10216/83250

Big Data: Fueling the Next Evolution of Agricultural Innovation

Steven T. Sonka

Emeritus Professor, Agriculture and Consumer Economics, University of Illinois at Urbana-Champaign, Illinois, USA

ssonka@illinois.edu

Abstract. Application of Big Data in agriculture could both contribute to economic gain and to reduction of environmental impact. Especially at the farming level, the high cost of measuring actual operations as they occurred historically restrained decision making in the sector. Currently several sensing technologies associated with Big Data are being evaluated and adopted within the sector. Their adoption offers the opportunity to alter that historic benefit/cost relationship. Combined with advanced analytics, measurement and analysis of diverse sources of data promise to create value for sector decision makers and society. While consumers likely will continue to be the ultimate beneficiaries of such advances, the pattern by which value is captured by entities in the sector remains uncertain. Factors such as organizational collaboration and the application of rules associated with intellectual property will have significant impact upon the evolution of Big Data's implementation within agriculture.

Keywords. Farming Systems, Information Technology and Data Processing, Knowledge Economy, Agriculture-Industry Relationships.

1 Introduction

Agriculture[1] is a vitally important sector affecting the global economy, societal well-being, and the vitality of natural ecosystems. Access to safe, nutritious, and affordable food is a goal for the citizens of all nations. In many developing countries, agricultural production employs the majority of the labor force. In more developed nations, an effective food and agricultural sector typically is a key component of the economy. Since man first tilled the soil to raise crops, agriculture has affected its supporting natural resource systems. Producing food (and other products) for current and expected future population levels are stressing those natural systems and developing means to reduce that stress is of global interest.

Innovation, especially in the last 150 years, has been an important means by which food and agricultural systems have increased productivity and fed an ever increasing global population (Borlaug, 2000; Chakraborty and Newton, 2011; Reid, 2015). Mechanization of tillage practices fueled expansion of land available for production while reducing human drudgery and labor needs. Biology focused on crop breeding

[1] The term agriculture often is viewed as synonymous with the farming activity. However, in this paper agriculture is viewed more broadly to encompass the entire food and agricultural system from genetics to retail. The terms, production agriculture or farming, will be used when referring to that specific subsector in the system.

ISSN 2183-0606
http://www.open-jim.org

increased the amount of production available from a given amount of inputs. Science applied to mitigation of the pests that affect crops and livestock and to more effective preservation of agricultural produce after harvest further ensure that food availability could expand for much of the world's population. More recently, genetic advances through application of biotechnology have been successfully employed (albeit not without controversy) and offer future potential to further contribute to human wellbeing. To be effective, however, each of these innovations had to be understood, adopted, and adapted by farmers and other managers.

In just the last few years, another source of innovation, Big Data, has captured the attention of citizens and decision makers in both the public and private sectors. While some would assert that Big Data currently is riding the crest of its "hype cycle" (Zwilling, 2014), application of Big Data has been effectively applied in numerous diverse settings. And Big Data is perceived to be as relevant for agriculture as it is for the rest of the economy, even by non-aggies. Padmasree Warrior, Chief Technology and Strategy Officer for Cisco Systems (Kirkland, 2013), believes:

> *In the next three to five years, as users we'll actually lean forward to use technology more versus what we had done in the past, where technology was coming to us. That will change everything, right? It will change health care;* ***it could even change farming****. There are new companies thinking about how you can farm differently using technology; sensors connected that use water more efficiently, use light, sunlight, more efficiently.*

While such potentials are exciting, it is important to remember that Big Data won't have much impact unless it too is understood, adopted, and adapted by farmers and other managers.

The purpose of this article is to explore the potential implications for Big Data and its adoption in agriculture. Because of the article's perspective on the future, its findings are necessarily speculative. The article is comprised of the following five sections:

- Key analytical concepts
- Precision agriculture; precursor to Big Data
- Likely sources of value creation
- Understanding the potential for value capture
- Summary and implications

This article's perspective is that the tools and techniques associated with Big Data offer the potential for agriculture to become significantly more effective in the pursuit of both economic and societal goals. Big Data's application can remove one of the fundamental constraints limiting agricultural managers – farmers, private sector managers, and public sector decisions makers. The constraint that the cost of measurement of actual operations typically has been significantly higher than the resulting benefits. Therefore, decisions tended to be driven by general conditions and often had a heavy bias to repeating traditional methods. Learning from actual operations was limited. The capabilities associated with Big Data offer the potential to fundamentally fuel management innovation. As will be further detailed in the paper, fully exploiting Big Data capabilities likely will require development of novel relations between firms and sectors within agriculture. This evolution could contribute to fundamental strategic change in the sector.

2 Key analytical concepts

Big Data is a term that has received extensive exposure. However, as illustrated in Figure 1, that exposure is a relatively recent phenomenon. Prior to 2011, the Big Data term was barely of note. However, the term's usage literally exploded in 2012 and 2013. Therefore, while it is both appropriate and important to attempt to anticipate the potential impact of Big Data, that anticipation can't be based upon historic experience in the overall economy or in the agricultural sector itself. Instead, this analysis must intentionally be speculative in nature. Of course, as physicist Niels Bohr has said, "Prediction is extremely difficult. Especially about the future" (Ellis, 1970, p. 431).

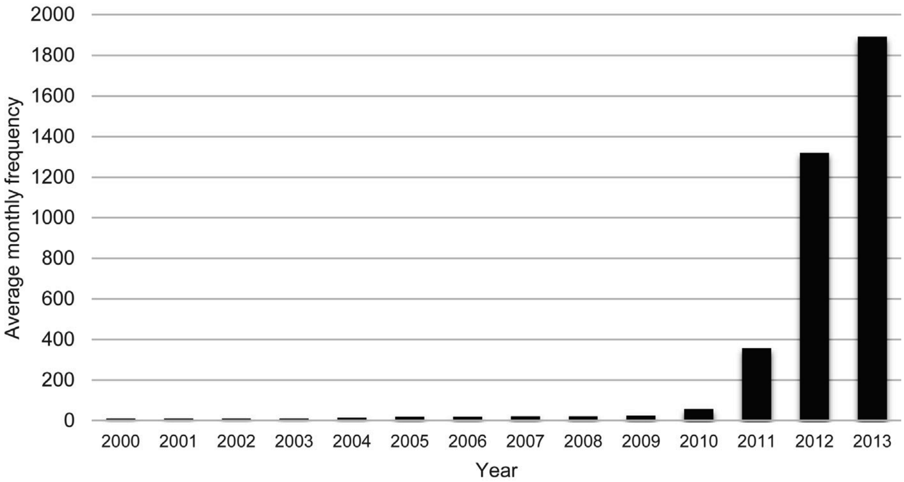


Fig. 1. Frequency distribution of documents containing the term "big data" in ProQuest Research Library (Gandomi and Haider, 2015).

The analysis presented here will employ three strategic concepts as a fundamental framework:

- The role of business models
- Value creation/value capture
- The resource-based theory of the firm.

Each concept will be described briefly in this section as they form the basis for the analysis presented later in the report.

2.1 The Role of Business Models

The term business model achieved extensive notoriety in the late 1990s as an outgrowth of the sudden surge of interest in e-commerce and the Internet as a business tool (Zott et al., 2011). While much of the media use of the term is not well structured, the term has important use as a means to understand the business and technological logic by which firms compete in their marketplace. As will be detailed later, the use of Big Data tools and approaches in agriculture likely will affect the nature of competition and of inter-firm relationships. Business models that have long existed in the sector therefore will be under pressure for change.

Although media use of the business model term tends to be unstructured, recent work in the academic literature does provide useful definitions:

- A firm's business model is "a system of interdependent activities that transcends the focal firm and spans its boundaries" (Zott and Amit, 2010, p.216).
- The business model is "the heuristic logic that connects technical potential with the realization of economic value" (Cheesebrough and Rosenbloom, 2002, p.529).
- Business models consist of four interconnected elements – customer value proposition, profit formula, key resources, and key processes (Johnson et al., 2008).

The nature of business models for firms in production agriculture (farms) and those firms which support the farm sector have to a large extent been dictated by the costs of capturing and communicating data (Sonka et al., 2000). Historically the costs of data management were high relative to the direct benefits of doing so. Therefore, transaction-based interactions (employing only price and quantity information) dominate the business models both at the farm and the agribusiness level. As will be detailed in later sections of the paper, that historic cost/benefit relationship will be fundamentally altered by the application of the technologies and methods associated with Big Data. This has the potential to reshape the dominant business model employed in the sector as well.

2.2 Value creation/value capture

To be successful, innovations need to provide value to users and to do that in a way that provides incentives and compensation to the inventors (as well as returns to the business entities employing the innovation to provide goods and/or services). The processes of value creation and value capture, therefore, are key to understanding adoption of innovations. Those processes, however, have differing dynamics that should be carefully understood.

From an economic perspective, innovations are judged based upon the value that their use can provide. That use actually can be further divided into two components; use value and exchange value (Bowman and Ambrosini, 2000). Exchange value is more easily measured as it is documented as the price users pay for the goods and/or services associated with use of the innovation. Profits are the difference between the exchange value and the cost of providing those goods and/or services.

Use value, however, is the perceived benefit received by the user. For business uses, use value often can be measured. For consumer innovations, the benefits exist but tend to be subjective in nature.

Value capture is the process by which the profits earned from use of innovations accrue to the various entities involved. Customers compare benefits from use of the innovation with existing and emerging alternatives which can address the same purpose. Value capture, the realization of the exchange value, is driven by the bargaining power of buyers and sellers (Bowman and Ambrosini, 2000).

In agriculture, the eventual beneficiaries of technological progress historically have been consumers. While innovations from Big Data may not change that outcome, the pattern by which actors in the sector are benefitted from their use is a dynamic and uncertain process.

2.3 Resource-based theory of the firm

A strategic concept, the resource-based theory of the firm, has proven useful in understanding and anticipating the dynamics of value capture in numerous settings (Barney, 1991). Relative to technology innovation, this approach focuses on the resource portfolios of effected firms.

Here the firm[2] is considered as a bundle of resources. Some of those resources can be complements essential for successful implementation of the innovation in question. Other resources can be competitive substitutes, which serve as forces to constrain innovation or which may be rendered obsolete by innovation. In the competitive marketplace, firms which excel are those who can integrate innovative technologies with existing resources in a manner which fosters sustainable competitive advantage. Such resources are identified as:

- Valuable,
- Rare,
- Hard to imitate, and
- Have weak substitutes.

The resource-based approach is particularly intriguing relative to Big Data applications in agriculture because of the likely need for complementary resources to fully exploit the benefits of Big Data innovations. These resources reside in firms and organizations at differing levels within the sector.

3 Precision agriculture; Precursor to Big Data

This section of the paper will provide a brief overview of the precision agriculture experience. It is not intended as comprehensive assessment. It is intended to provide a sense of the evolution of precision agriculture, identify the more popular technologies employed and discuss the admittedly scanty evidence as to the economic gains from use of these innovations.

It is important to note that precision agriculture and Big Data are not synonymous. As we'll see, the current tools and techniques of precision agriculture have existed largely without Big Data concepts. However, it is hard to foresee that Big Data approaches could have significant impact without employing precision agriculture technologies. Further, some of attributes of Big Data adoption likely are foretold by the precision agriculture experience.

---

[2] For simplicity, the term, firm, is used in this discussion, even though it might be more accurate to refer to economic actors. Such economic actors could include NGO, universities, and government research entities who are and have the potential playing key roles in the evolution of Big Data in agriculture. This might particularly be the case for Big Data application in developing agricultural settings.

Precision agriculture has several dimensions; indeed the concept itself is not precisely defined. A 1997 report of the National Research Council refers to precision agriculture, "*... as a management strategy that uses information technologies to bring data from multiple sources to bear on decisions associated with crop production*". Key technologies and practices included within precision agriculture are:

- Georeferenced information;
- Global positioning systems;
- Geographic information systems and mapping software;
- Yield monitoring and mapping;
- Variable-rate input application technologies;
- Remote and ground-based sensors;
- Crop production modeling and decision support systems; and
- Electronic communications.

The term, precision agriculture, primarily has been linked to crop production. However, precision practices (and Big Data techniques for that matter) are equally applicable in animal agriculture, where georeferencing can refer to both sub areas of a field and individual animals. The tracking processes and required tools may differ but the managerial goal is still to separately manage increasingly smaller units of observation.

Farmers and agribusiness managers played a significant role in the development of precision agriculture. For example, in the mid-1990s, a group of agribusiness professionals in Champaign County, Illinois, came together to explore the opportunities associated with two emerging technologies — site-specific agriculture and that strange thing called the Internet. This group, called CCNetAg, was part of an initiative co-sponsored by the local Chamber of Commerce and the National Center for Supercomputing Applications at the University of Illinois. A voluntary enterprise, CCNetAg provided a vehicle for farmers, agribusiness managers, and university researchers to jointly explore adoption of these tools. Figure 2 depicts their expectations of a <u>then</u> future precision agriculture.

Although created some time ago, the graphic continues to depict key elements of precision farming:

- The role of georeferencing is indicated by satellites linking to the farm field.
- On the field itself, key farming operations are being directed by and are capturing digital information on:
  - Soil characteristics,
  - Nutrient application,
  - Planting,
  - Crop scouting, and
  - Harvesting.
- The layers that underlie the farm field represent the notion that visual mapping would allow the farmer, and the farmer's advisors, to see meaningful correlations to inform future decisions.

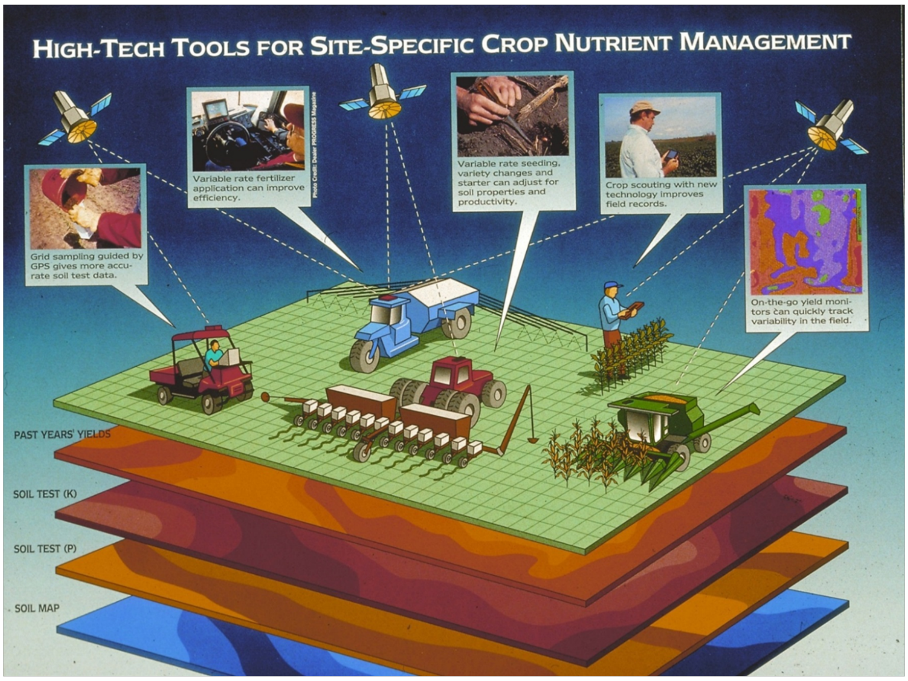


Fig. 2. A mid-1990s view of precision farming from the CCNetAg group (Sonka and Coaldrake, 1996).

Since 1997, technologies have advanced, although the general categories remain relevant. For example, auto-steer capabilities on farm machinery have become much more prevalent. And active, detailed measurement of the planting process (recording where "skips" occur) is now feasible. Further, the ability to monitor the status of farm machinery as it operates is now paired with electronic communications to signal when machine operations are out of acceptable bounds.

While there have been many publications describing precision agriculture, reports with independent evaluation of the economics of adoption are much less numerous. One means to assess whether there are net benefits of a technology is to monitor its marketplace adoption. For several years the Center for Food and Agricultural Business at Purdue University and CropLife magazine have surveyed agricultural input suppliers regarding the adoption of precision agriculture. Focused primarily on the Midwest and Southern regions, this work is a particularly useful assessment of the technology's application. From the 2015 report, Figures 3 and 4 provide evidence of adoption for key precision agriculture practices (Erickson and Widmar, 2015).

The crop input dealers who provided input for this study are uniquely well positioned to understand and report on adoption of these technologies. Their firms provide inputs (fertilizer, pesticides, and seeds) and services to producers evaluating and adopting precision agriculture.

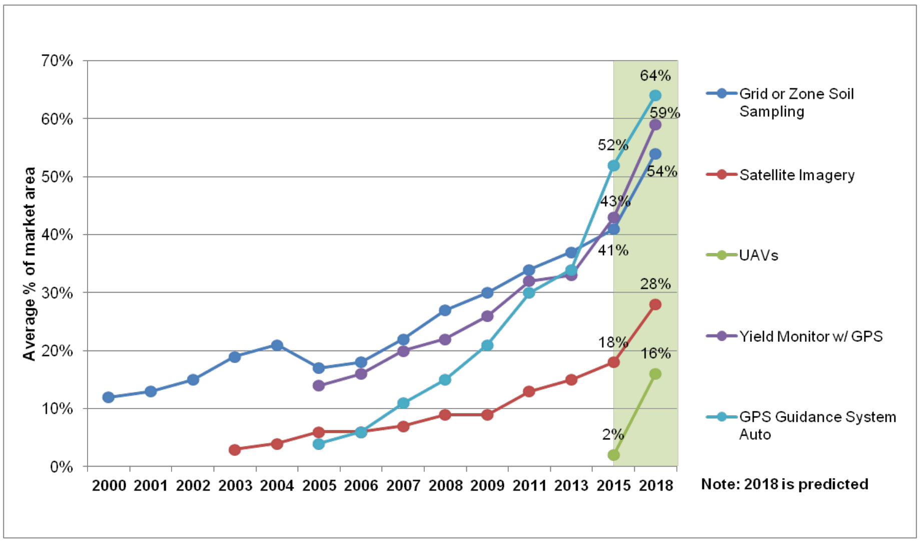


Fig. 3. Estimated market area using precision services over time (adapted from Erickson and Widmar, 2015).

Early interest in precision agriculture focused on site-specific application of inputs and on use of yield monitors. As shown in Figure 3, grid sampling, a practice associated with site-specific lime and fertilizer application, is currently employed on about 2 out of 5 crop acres. Increased coverage to a majority of acres is expected by 2018. Similar adoption rates (43% and 59%) are noted for GPS-assisted yield monitors. Over the last decade, use of GPS guidance systems has increased rapidly, to a current use estimated to exceed 50%. Continued strong growth to 2018 is expected. The use of satellite imagery and UAVs as tools to support crop production is more recent. Current use affects 18% and 2% of acreages, respectively. Interesting, acreage covered by UAVs is expected to increase eightfold, to 16%, in just three years.

Figure 4 describes a relatively consistent adoption pattern for VRT (variable rate technology) practices. In the early 2000s, adoption was at single digital levels. Since then, steady increases in the extent of acreage covered have occurred. However, the most utilized practice, application of lime, is only now achieving coverage on 41% of the total acreage. These patterns also are interesting because of the very different price regimes that existed for corn and soybeans over these 15 years. When output prices were low prior to 2008, the driver for adoption likely was cost reduction. Possibly, increasing yields was a more significant factor in recent years when prices were higher.

Media and marketing attention sometimes blur distinctions between precision agriculture and Big Data. Some communications seem to suggest that Big Data is just an updated buzzword for precision agriculture practices. That is not the case.

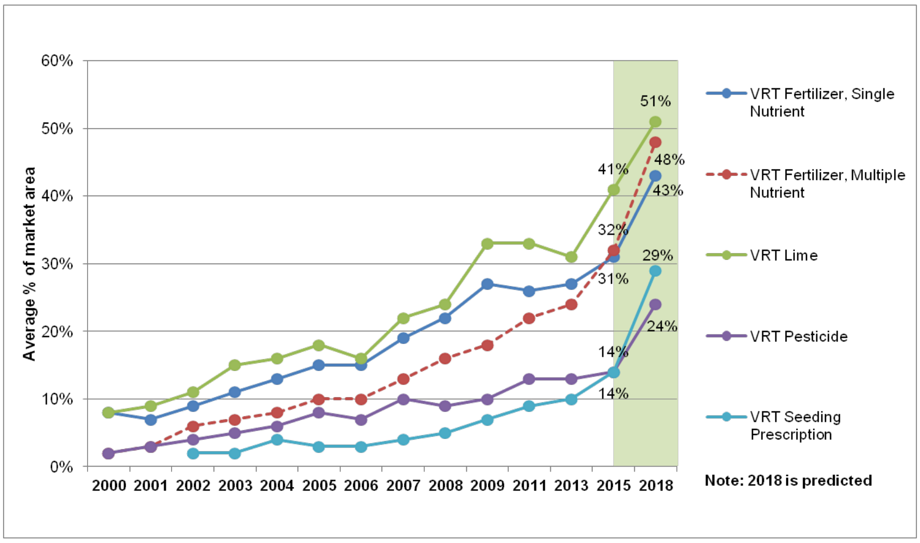


Fig. 4. Estimated market area using VRT technology over time (adapted from Erickson and Widmar, 2015).

Figure 2 above can be used to identify key differences:

- While a useful picture, that graphic does focus our attention on the individual field. The volume characteristic of Big Data requires observations from many, many farm fields to be effective. Discerning the interrelated effects of soil type, several nutrients, and seed variety requires data dispersed over time and space.
- While the farmer has several types of precision data from each field, additional sources of data naturally reside and originate beyond the fencerow. Achieving the Big Data's variety characteristic requires access to that broader set of information.
- Precision agriculture employs comparisons across field map layers as its dominant method of analysis. The effect of a single factor, such as a blocked tile line or a buried fencerow, often is observable from a map. However, identifying complex interactions across several production factors and multiple years requires much more sophisticated tools. Analytics is a major differentiating feature of Big Data.
- As noted previously, precision agriculture has had 20+ years of experience. Aggregating all the digital information collected from yield monitors and site-specific input operations would result in an extremely large set of data. However, that data currently is located on innumerable thumb drives, disk drives, and desktop computers. Effective analysis won't be possible unless/until that data can be accessed and aggregated. The associated organizational issues of doing that will be discussed in a later section of this article.

Both precision agriculture and Big Data arise from the advent and application of information and communication technologies. As noted previously, they are not synonymous. That said, it is hard to foresee that Big Data approaches will have significant impact without employing the data generated by precision agriculture practices.

4 Likely sources of value creation

Big Data generally is referred to as a singular entity. It is not! In reality, Big Data is much more a capability than it is a thing. It is the capability to extract information and insights where previously it was economically, if not technically, not possible to do so. Advances across several technologies are fueling the growing Big Data capability. These include, but are not limited to computation, data storage, communications, and sensing. The growing ability of analysts and managers to exploit the information provided by the Big Data capability is equally important.

Although of relatively recent origin, numerous attempts have been made to define Big Data. For example:

- The phrase "big data" refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future (The National Science Foundation, 2012).
- Big Data shall mean the datasets that could not be perceived, acquired, managed and processed by traditional IT and software/hardware tools within a tolerable time (Chen et al., 2014)
- Big Data is where the data volume, acquisition velocity, or data representation [variety] limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing (Cooper and Mell, 2012).
- Big Data is high-volume, -velocity, and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making (Gartner IT Glossary, 2012).

The purpose of this section of the paper is to move beyond those definitions to explore how application of Big Data could foster the creation of value in agriculture. Three pathways to value creation are identified. Application of tools to measure and monitor agricultural activities – at extremely low cost – is the first. Data analytics which can integrate data from diverse sources to generate novel insights is the second. The third factor focuses on external pressures to better monitor agricultural activities which, in so doing, create sources of data that potentially can lead to strategic change.

4.1 Dimensions of Big Data

Three dimensions (Figure 5) often are employed to describe the Big Data phenomenon: Volume, Velocity, and Variety (Manyika et al., 2011). Each dimension presents both challenges for data management and opportunities to advance business decision-making. These three dimensions focus on the nature of data. However, just having data isn't sufficient. Analytics is the hidden, "secret sauce" of Big Data. Analytics, discussed later, refers to the increasingly sophisticated means by which useful insights can be fashioned from available.

"*90% of the data in the world today has been created in the last two years alone*" (IBM, 2012). In recent years, statements similar to IBM's observation and its emphasis on volume of data have become increasingly more common.

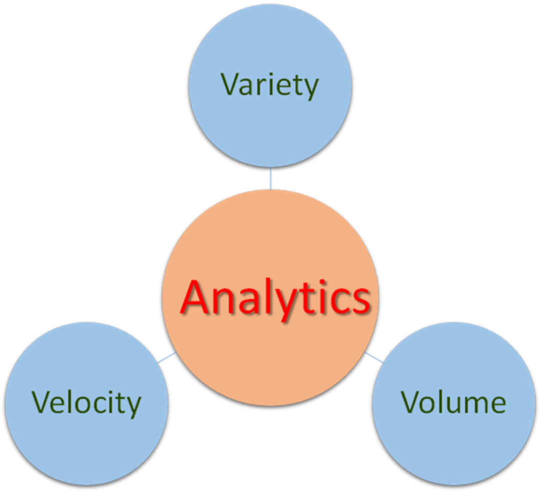


Fig. 5. Dimensions of Big Data.

The Volume dimension of Big Data is not defined in specific quantitative terms. Rather, Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective; with no single standard of how big a dataset needs to be to be considered big. And that standard can vary between industries and applications.

An example of one firm's use of Big Data is provided by GE — which now collects 50 million pieces of data from 10 million sensors everyday (Hardy, 2014). GE installs sensors on turbines to collect information on the "health" of the blades. Typically, one gas turbine can generate 500 gigabytes of data daily. If use of that data can improve energy efficiency by 1%, GE can help customers save a total of $300 billion (Marr, 2014)!

The Velocity dimension refers to the capability to acquire, understand, and respond to events *as they occur*. Sometimes it's not enough just to know what's happened; rather we want to know what's happening. We've all become familiar with real-time traffic information available at our fingertips. Google Map provides live traffic information by analyzing the speed of phones using the Google Map app on the road (Barth, 2009). Based on the changing traffic status and extensive analysis of factors that affect congestion, Google Map can suggest alternative routes in real-time to ensure a faster and smoother drive.

For analysts interested in retailing, anticipating the level of sales is important. Brynjolfsson and McAfee (2012) report on an effort to monitor mobile phone traffic to infer how many people were in the parking lots of a key retailer on Black Friday — the start of the holiday shopping season in the United States — as a means to estimate retail sales.

Variety, as a dimension of Big Data, may be the most novel and intriguing of these three characteristics. For many of us, data referred to numbers meaningfully arranged in rows and columns. For Big Data, the reality of "what is data" is wildly expanded. The following are just some of the types of data available to be converted into information:

- Financial transactions
- The movement of your eyes as you read this text
- "Turns of a screw" in a manufacturing process
- Tracking of web pages examined by a customer
- Photos of plants
- GPS locations
- Text
- Conversations on cell phones
- Fan speed, temperature, and humidity in a factory producing motorcycles
- Images of plant growth taken from drones or from satellites
- Questions

4.2 Data variety requires low cost measurement

"You Can't Manage What You Don't Measure!" is a phrase attributed to both Peter Drucker and W. Edwards Deming. This phrase is as applicable to farmers as it is to managers at Toyota or Amazon (Brynjolfsson and McAfee, 2012). The relationship between measurement and the ability to make improved decisions is critically important in understanding the potential for Big Data to affect agricultural management.

The author of this paper grew up on a small farm in the Midwest region of the United States and, throughout his career, has learned extensively from farmers in the United States and globally. With apologies for a small digression, let me use personal experience to focus on the linkage between measurement and management. Growing up on a farm, the linkage between what could be measured and our ability to improve performance was straightforward. In those days, we had to carry the, hopefully, full milking machine from the cow to the milk tank and it was fairly easy to know which cows were producing more. And because there were less than 20 cows in the herd, it also was possible to remember those higher producing cows and give them an extra portion of grain.

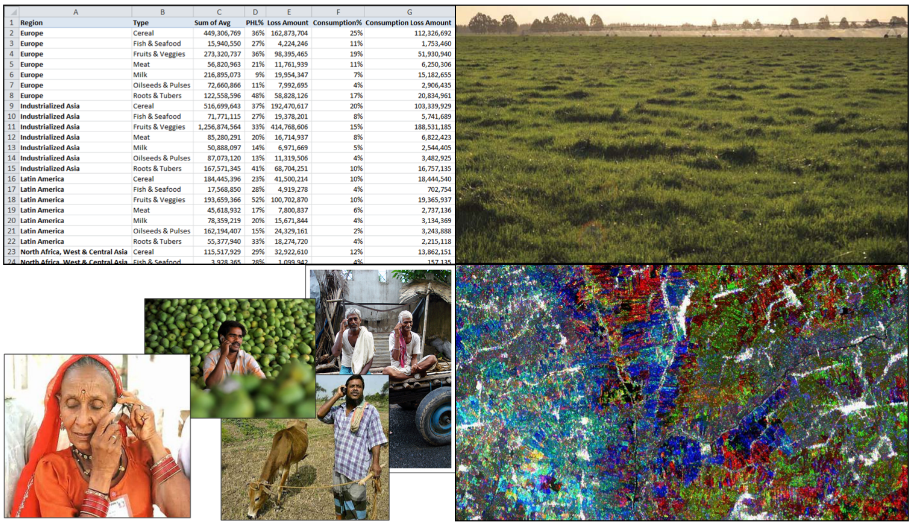
On this same farm, about 120 egg producing chickens were housed in a building, with ample room to roam outdoors as well. Eggs were collected twice a day. Performance of the entire of group was observable. Knowledge that could lead to improved performance of individual birds, however, was not observable. Technically, it might have been possible to establish a production system where measurement of individual bird performance would have been available. However, the economics of egg production at that time didn't justify the costs of such a system.

The important point to stress here is that the desire to link measurement of outcomes and management actions in farming is not new. However, the economics of measurement (the cost of measurement versus the benefits of doing so), given the available technology, inhibited my father and other farmers from capturing and exploiting more data.

4.3 Variety as a key

Suddenly (at least in agricultural measurement terms), the "what is data" question – the variety dimension of Big Data – has new answers. Figure 6 provides a visual illustration of the change. In its upper left hand corner, we see data as we are used to it – rows and columns of nicely organized numbers. The picture in the upper right hand corner is of a pasture in New Zealand. Pasture is the primary source of nutrition for dairy cows in that country and supplemental fertilization is a necessary economic practice. The uneven pattern of the forage in that field is measured by a sensor on the fertilizer spreader to regulate how much fertilizer is applied – as the spreader goes across the field. In this situation, uneven forage growth is now data.

The lower left hand corner of Figure 6 shows the most versatile sensor in the world – individuals using their cell phone. Particularly for agriculture in developing nations, the cell phone is a phenomenal source of potential change – because of both information sent to those individuals and information they now can provide. And as illustrated in the lower right hand quadrant of Figure 6, satellite imagery can measure temporal changes in reflectivity of plants to provide estimates of growth (RIICE, 2013). The picture is focused on rice production in Asia.



	A	B	C	D	E	F	G
1	Region	Type	Sum of Avg	PHL%	Loss Amount	Consumption%	Consumption Loss Amount
2	Europe	Cereal	449,306,769	36%	162,873,704	25%	112,326,692
3	Europe	Fish & Seafood	15,940,550	27%	4,224,246	11%	1,753,460
4	Europe	Fruits & Veggies	273,320,737	36%	98,395,465	19%	51,930,940
5	Europe	Meat	56,820,963	21%	11,761,939	11%	6,250,306
6	Europe	Milk	216,895,073	9%	19,954,347	7%	15,182,655
7	Europe	Oilseeds & Pulses	72,660,866	11%	7,992,695	4%	2,906,435
8	Europe	Roots & Tubers	122,558,596	48%	58,828,126	17%	20,834,961
9	Industrialized Asia	Cereal	516,699,643	37%	192,470,617	20%	103,339,929
10	Industrialized Asia	Fish & Seafood	71,771,115	27%	19,378,201	8%	5,741,689
11	Industrialized Asia	Fruits & Veggies	1,256,874,564	33%	414,768,606	15%	188,531,185
12	Industrialized Asia	Meat	85,280,291	20%	16,714,937	8%	6,822,423
13	Industrialized Asia	Milk	50,888,097	14%	6,971,669	5%	2,544,405
14	Industrialized Asia	Oilseeds & Pulses	87,073,120	13%	11,319,506	4%	3,482,925
15	Industrialized Asia	Roots & Tubers	167,571,345	41%	68,704,251	10%	16,757,135
16	Latin America	Cereal	184,445,396	23%	41,500,214	10%	18,444,540
17	Latin America	Fish & Seafood	17,568,850	28%	4,919,278	4%	702,754
18	Latin America	Fruits & Veggies	193,659,366	52%	100,702,870	10%	19,365,937
19	Latin America	Meat	45,618,932	17%	7,800,837	6%	2,737,136
20	Latin America	Milk	78,359,219	20%	15,671,844	4%	3,134,369
21	Latin America	Oilseeds & Pulses	162,194,407	15%	24,329,161	2%	3,243,888
22	Latin America	Roots & Tubers	55,377,940	33%	18,274,720	4%	2,215,118
23	North Africa, West & Central Asia	Cereal	115,517,929	29%	32,922,610	12%	13,862,151
24	North Africa, West & Central Asia	Fish & Seafood	3,928,365	28%	1,099,942	4%	157,135

Fig. 6. A few sources of data3.

While satellite imagery is one source of remotely sensed data, recent years have seen a pronounced increase in the capabilities and interest in Unmanned Aerial Systems (UAS) as a source of data for agriculture. There are numerous ongoing efforts to transform UAS technology originally focused on military purposes to applications supporting production agriculture. "Universities already are working with agricultural

[3] Graphics courtesy of: agrioptics.co.nz; T. Abdelzaher, Champaign, IL.; Mock, Morrow & Papendieck; International Rice Research Institute.

groups to experiment with different types of unmanned aircraft outfitted with sensors and other technologies to measure and protect crop health" (King, 2013). Example applications include:

- Monitoring of potato production (Oregon State University)
- Targeting pesticide spraying on hillside vineyards (University of California, Davis)
- Mapping areas of nitrogen deficiency (Kansas State University)
- Detecting airborne microbes (Virginia Polytechnic Institute and State University)

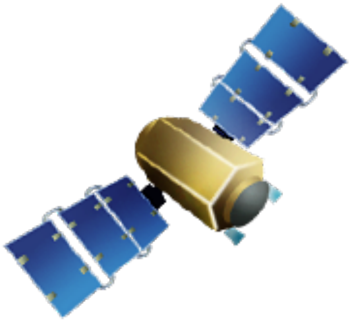


Fig. 7. Unmanned Aerial Systems offer low cost data acquisition[4].

Those specific examples are only a few of the numerous experiments and demonstrations being conducted to identify cost effective means to employ UAS technology (Figure 7). UAS capabilities offer flexibility and potentially lower cost relative to the use of even small manned aircraft. Development efforts are being conducted globally; however, it is likely that initial commercial application will occur where higher value crops dominate.

4.4 Analytics

Access to lots of data, generated from diverse sources with minimal lag times, sounds attractive. Managers, however, quickly will ask, what do I do with all this stuff? Without similar advances in analytic capabilities, just acquiring more data is unlikely to have significant impact within agriculture.

Analytics and its related, more recent term, data science, are key factors by which Big Data capabilities can actually contribute to improved performance in the agricultural sector. Data science refers to the study of the generalizable extraction of knowledge from data (Dhar, 2013). Tools based upon data science are being developed for implementation in the sector, although these efforts are at their very early stages.

The associated concept of analytics similarly is maturing and its use refined (Davenport, 2013; Watson, 2013). Analytic efforts can be categorized as being of one of three types:

- Descriptive efforts focus on documenting what has occurred,

---

[4] Graphic courtesy of: Microsoft Corporation

- Predictive efforts explore what will occur, and
- Prescriptive efforts identify what should occur (given the optimization algorithms employed).

One tool providing predictive capabilities was recently unveiled by the giant retailer, Amazon (Bensinger, 2014). This patented tool will enable Amazon managers to undertake what it calls "anticipatory shipping", a method to start delivering packages even before customers click "buy". Amazon intends to box and ship products it expects customers in a specific area will want but haven't yet ordered. In deciding what to ship, Amazon's analytical process considers previous orders, product searches, wish lists, shopping-cart contents, returns, and even how long an Internet user's cursor hovers over an item. Analytics and its related, more recent term, data science, are key factors by which Big Data capabilities actually can contribute to improved performance, not just in retailing, but also in agriculture.

In agriculture, as in most fields, descriptive efforts have been most common and even those are relatively infrequent. Within production agriculture, knowing what has occurred – even if very accurately and precisely – does not necessarily provide useful insights as to what should be done in the future.

Production agriculture is complex, where biology, weather, and human actions interact. Science-based methods have been employed to discern why crop and livestock production occurs in the manner in which they do. Indeed, relative to the Big Data topic, it might be useful to consider this as the "small data" process.

The process starts with lab research employing the scientific method as a systematic process to gain knowledge through experimentation. Indeed the scientific method is designed to ensure that the results of an experimental study did not occur just by chance (Herren, 2014). However, results left in the lab don't lead to innovation and progress in the farm field. In the United States, the USDA, Land Grant universities, and the private sector have collaborated to exploit scientific advances. A highly effective, but distributed, system emerged where knowledge gained in the laboratory was tested and refined on experimental plots and then extended to agricultural producers.

In agriculture, therefore, knowledge from science will need to be effectively integrated within efforts to accomplish the goals of predictive and prescriptive analytics. Even with this additional complication, the potential of tools based upon emerging data science capabilities offers significant promise to more effectively optimize operations and create value within the agricultural sector.

4.5 Public pressures to better monitor agriculture

Beyond its direct economic impact, society has intense interest in the social and environmental effects of the agricultural sector. Food safety and security are of public interest in every society. Interest in mitigating negative environmental impacts of agricultural operations is increasingly of interest and that interest is not constrained to just citizens in developed nations. In addition to public sector interest, some consumer segments express interest and concern regarding the practices and methods employed to produce food. Therefore, in addition to public sector-based regulation, documentation as to practices employed is increasingly being required by the private sector by food manufacturers and retailers.

Interestingly, technological innovations, such as those noted previously, have potential to provide much better evidence as to these societal and environmental effects. These include both tools to more precisely measure and monitor as well as analytical methods to better understand and predict effects.

At first blush, managers tend not to welcome additional constraints, whether from public or private sources. However, there can be an interesting "unintended consequence" effect when information is captured digitally. That digital information, which might not have been captured otherwise, now becomes available for analysis. As we saw in the early days of the 1990s knowledge economy, unintended insights can be developed from digital data captured for other purposes (Sampler, 1997; Shapiro and Varian, 1999). Application of those insights can drive strategic change in effected industries.

5 Understanding the potential for value capture

To be attractive, prospective innovations have to display the potential to create value in the marketplace. The longer run economic effect of adopting innovations, however, is determined by value capture, the distribution of resulting benefits to consumers and among the firms within the value chain affected by the innovation. The prior section identified three interrelated pathways by which the technologies and application of Big Data can potentially create value for consumers, society and to the sector's economic entities. This section will explore the concept of value capture relative to the adoption of Big Data within agriculture.

Identification of potential value creation typically is more straightforward than is predicting the pattern and extent of value capture. Historically, food consumers have been the eventual beneficiaries of technology adoption in production agriculture. Even if that remains the likely long-run outcome, the allocation of net benefits among the sector's economic actors is of key interest. As noted in an earlier section of the paper, a strategic concept called resource-based theory of the firm has proven useful in understanding and anticipating the dynamics of value capture (Bowman and Ambrosini, 2000). Particularly in the context of Big Data in agriculture, the resources needed to create and capture value often will not reside within one firm. Therefore, new business models that enable collaboration across firm boundaries likely will be needed. Implementation of these business models could allow application of Big Data tools and techniques to be powerful and sustainable sources of competitive advantage.

From a manager's strategic perspective, therefore, implementation of effective Big Data based innovations is attractive. Within agriculture some of the data comprising these systems likely will come from external sources (for example, weather data, environmental regulatory filings, and futures market price movements). Other systems, however, will be based upon data generated from activities internal to the operations of firms in the food and agribusiness sector. Although that data often will be analyzed in combination with external data, firms will need access to internal data to effectively compete. Therefore, data access, based upon current operations, represents a resource of critical potential importance and is a starting point for this analysis.

Figure 8 provides a high level view of the key subsectors within agriculture that has proved useful for consideration of future competitive dynamics relating to Big Data. The genetics subsector is separately identified here because of its linkages with Big Data. A number of firms in that category have capabilities to operate as input suppliers as well. The input supply category refers to providers of equipment, seed, fertilizer, and chemicals to farmers as well as providers of financial and managerial services. The production agriculture segment is comprised of farming firms, which can range from low-resourced, smallholders to family corporations to subsidiaries of major corporations. The 1st handler segment refers to firms which aggregate, transport and initially process agricultural produce but do not directly market to consumers. The final segment relates to food manufacturers and retailers. These types of activities are combined here because of their common interest in employing Big Data tools to better understand consumers.

From a strategic perspective, it is important to stress that Big Data tools already are extensively employed, particularly at both "ends" of the sector. Firms at the food manufacturing and the food retailing levels expend considerable resources to continually develop a better understanding of consumers. Insights gained through application of Big Data analytics can allow managers both to anticipate and respond to consumer concerns. Far upstream in the sector, bioinformatics and other Big Data tools are employed to accelerate research and development processes, advancing genomic capabilities of the sector. Figure 8 identifies, at a general level, key interests that "naturally" reside within each subsector and have the potential to be important within Big Data applications.

Agricultural operations occur across time and space. Therefore, the logistics of providing inputs, production, and aggregating output consume considerable resources. Advances in information and communication technology combined with Big Data analytics offer the potential to reduce the amount of resources needed. Deadweight loss is a term that describes system inefficiencies that can be reduced by enhanced coordination within and between firms. Even in advanced agricultural settings, reduction of deadweight loss is perceived to be an attractive potential use of Big Data innovations.

In this context, deadweight loss refers to the processes by which inputs and outputs are delivered (when and where). A more intriguing issue for many is whether application of Big Data can fundamentally alter decision making as to "what" should be done. Can we further optimize the biology of agricultural production, especially in the context of the larger food and agricultural system? Earlier it was noted that new sensing technologies offer the potential to monitor and document what actually occurs as agricultural production takes place. The resulting data potentially would be available at never before levels of detail, in terms of time and space, and at low-cost. Further, analytic capabilities could combine diverse sources of data to discern previously unknown patterns and provide insights not available previously.

A result of application of these innovations would be optimization of agricultural production systems, simultaneously reducing its environmental impact and improving profitability. There are two interrelated factors that need to be addressed in considering the possible evolution of this optimization:

- Production agriculture involves biologic processes subject to considerable uncertainty. Therefore, even if one knows exactly what occurred in one production season and what actions would have optimized performance under those circumstances, that information may not be a good predictor of what actions should be undertaken in the next season. Agricultural science is devoted to discerning the why of agricultural production. That science will need to be integrated within Big Data techniques to truly optimize system performance.

- In most systems of agricultural production today, even the knowledge of what occurred doesn't necessarily reside within one organization. Further, as was noted for precision agriculture, individual entities at the production level typically don't have the scale to produce sufficient data nor to have the capabilities needed to analyze that data.

Because of these two factors, collaboration across organizational boundaries will be required to fully exploit the potential benefits of Big Data's application to agriculture. A host of factors, beyond technological effectiveness, will influence the speed and extent of this exploitation. These relate to intellectual property and competitive dynamics as well as the magnitude of economic benefits available. Such factors are not insurmountable and can be viewed as much as opportunities as they are impediments. How they are resolved, however, will have a major impact on Big Data's eventual contribution to performance within agriculture.

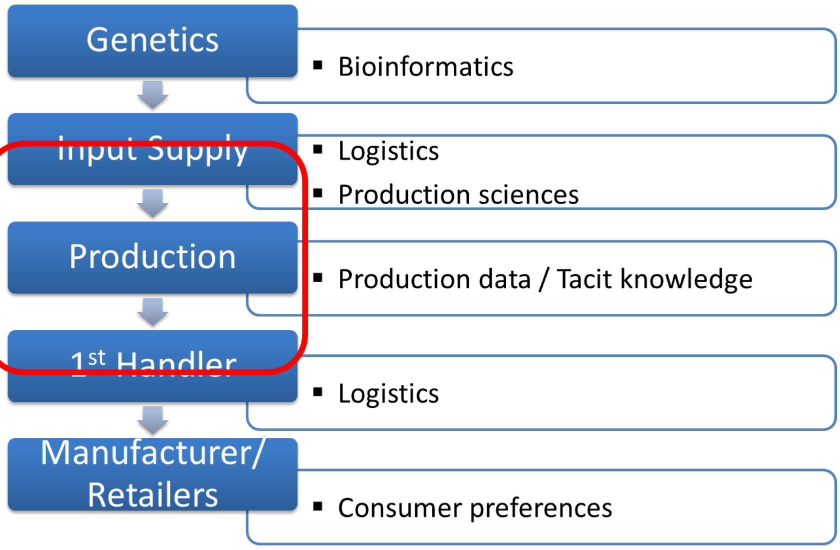


Fig. 8. Subsectors and their key strategic interests relating Big Data.

6 Summary and implications

Big Data capabilities have emerged in recent years as potential "game changers" that could affect economies and societies in profound, although somewhat uncertain, ways. Those potentials extend to economic, social, and environmental performance of food and agricultural systems as well. Although it is very early days in terms of Big Data adoption and agriculture, expectations already have been altered and investment in research, development, and testing of associated technologies is occurring.

Although necessarily speculative, this article explores the potential impact of Big Data in the context of the agricultural sector. While noting some of the technologies associated with potential Big Data implementation, a decision-making lens is adopted as the primary conceptual tool for this exploration. The reason for doing this is the belief that use of Big Data capabilities will have primary impact by altering decision-making processes relating to:

- Adoption and implementation of new technologies,
- Management of on-going operations, and
- Execution of existing and new relationships:
    - Among competing and collaborating firms
    - Between suppliers and customers
    - With customer and non-customer stakeholders.

To be economically attractive, innovations have to display the potential to create value in the marketplace. The longer run economic effect of adopting innovations, however, is determined by value capture, the distribution of resulting benefits to consumers and within the value chain affected by the innovation. Value capture is heavily influenced by the resource portfolios of effected firms. These strategic concepts, the business model, the resource-based theory of the firm, value creation and value capture, are employed here to frame the exploration of Big Data's potential effects.

Two interrelated questions are addressed in the context of potential strategic change driven by Big Data innovations. Specifically, if such change does occur:

- What would be the likely source of change?
- Who (in the context of economic entities) would be the likely change agents?

Historically, the geographic, time, and economic dimensions of agriculture have constrained the decision making capabilities of sector managers. Although managers desired to be able to measure the impact of their decisions and actions, typically the cost of measurement exceeded the benefits of doing so. Innovations, many of which are integral within a broad perspective of Big Data, now offer the potential to fundamentally alter that benefit/cost dynamic and in so doing foster the potential for value creation in the sector.

Three interrelated forces are identified as likely change agents driving value creation as (if) Big Data capabilities are applied in agriculture:

- Extensive implementation of low-cost sensor capabilities will allow managers to measure actual operation of systems and more effectively respond both in "real-time" and in planning future operations.
- The application of advanced analytics will provide insights that support improved decision making.
- Societal and business motivations will increasingly require more extensive monitoring in response to requirements imposed by the public sector or by customers. Because the associated data will be digital, prior experience indicates that additional use of that data can drive strategic change extending beyond the original intent.

The forces just identified offer the potential for value creation which can provide benefits to consumers, society and to the sector's economic entities. As is typically the case, identification of the forces for potential value creation is more straightforward than is anticipating the pattern of future value capture. Historically, consumers have been the eventual beneficiaries of adoption of technology in production agriculture. Even if that remains the likely long-run outcome, the allocation of net benefits among the sector's economic actors is of key interest.

Without attempting to predict that allocation, a number of key factors of interest can be detailed. It is important to note that Big Data capabilities already are being employed within the food and agribusiness sector. Firms at the retail and manufacturing level are aggressively monitoring social media and other data sources to better understand and serve consumers. Bioinformatics has become an essential tool for firms providing genetic resources for crops and livestock. In addition to direct application of the resulting information, linkages with associated partners at other levels of the sector offer the potential for further economic and social gains.

Firms operating at the input supply, production agriculture, and first handler levels of the sector are beginning to explore Big Data application. Employing Big Data for management and logistics purposes has the potential to reduce costs and to improve economic performance.

Optimization of the biology of production agriculture is a beguiling potential with extensive potential benefits. A few "farming" organizations do have the scale of operation which could justify development and application of Big Data capabilities. More generally, the information resources needed to move towards optimization reside within multiple organizations. The most numerous of these are individual farming operations. Typically, however, some combination of firms at the input supply, service provision and output handling/processor level also will have key elements of the needed information resources. Future decisions to shape effective business models for firms operating in these domains will determine the ultimate value capture dimensions of Big Data's application in agriculture.

Acknowledgements

The author wishes to express his sincere appreciation to Ms. Casey Cheng, University of Illinois, for her invaluable graphic design, editorial, and substantive contributions to this article. Over the last few years, the author has benefited greatly from numerous conversations and discussions with farmers, researchers, and agribusiness professionals from around the world regarding Big Data and its application. Although too numerous to list here, their contributions are greatly appreciate it. Any errors and omissions are, of course, the responsibility of the author.

7 References

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*. 17: 99-120.

Barth, D. (2009). The bright side of sitting in traffic: Crowdsourcing road congestion data. Retrieved February, 2, 2016, from Google's Official Blog Web site: https://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html

Bensinger, G. (2014). Amazon wants to ship your package before you buy it. *The Wall Street Journal*. Retrieved February, 2, 2016, from the Wall Street Journal Web site: http://blogs.wsj.com/digits/2014/01/17/amazon-wants-to-ship-your-package-before-you-buy-it/

Borlaug, N. E. (2000). Ending world hunger: The promise of biotechnology and the threat of antiscience zealotry. *Plant Physiology, 124*(2), 487-490.

Bowman, C., & Ambrosini, V. (2000). Value creation versus value capture: Towards a coherent definition of value in strategy. *British Journal of Management, 11*, 1-15.

Brynjolfsson, E. & McAfee, A. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60-68.

Chakraborty, S., & Newton, A. C. (2011). Climate change, plant disease and food security: An overview. *Plant Pathology, 60*(1), 2-14.

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

Cheesebrough, H. W., & Rosenbloom, R. S. (2002). The Role of the Business Model in Capturing Value from Innovation. *Industrial and Corporate Change*, *11*, 533-534.

Cooper, M., & Mell, P. (2012). Tackling Big Data [PDF document]. Retrieved February 2, 2016, from National Institute of Standards and Technology Web site: http://csrc.nist.gov/groups/SMA/forum/documents/june2012presentations/fcsm_june2012_cooper_mell.pdf

Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review, 91*(12), 64-72.

Dhar, V. (2013). Data Science and Prediction. *Communications of the ACM*, *56*(12), 64-73.

Ellis, A. K. (1970). *Teaching and learning elementary social studies*. Boston: Allyn & Bacon.

Erickson, B. & Widmar, D.A. (2015). *2015 Precision Agricultural Services Dealership Survey Results*. West Lafayette, Indiana: Purdue University. Retrieved February 2, 2016, from Center for Food and Agribusiness, Purdue University Web site: http://agribusiness.purdue.edu/resources/2015-precision-dealership-survey-results

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137-144.

Gartner IT Glossary. (2012). *Big Data*. Retrieved June, 5 2015, from Gartner Web site: http://www.gartner.com/it-glossary/big-data

Hardy, Q. (2014). G.E. opens its Big Data platform. Retrieved February 2, 2016, from The New York Times Web site: *http://bits.blogs.nytimes.com/2014/10/09/ge-opens-its-big-data-platform/?_r=2*

Herren, R. V. (2014). *The Science Of Agriculture A Biological Approach 4th Ed*. Independence: Cengage Learning.

IBM. (2012). *What is big data?* Retrieved November 22, 2015, from IBM Web site: https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Johnson, M. W., Christenson, C. C., & Kagermann, H. (2008). Reinventing Your Business Model. *Harvard Business Review*, *86*(12), 50-59.

King, R. (2013). Farmers Experiment with Drones. Retrieved February 2, 2016, from *The Wall Street Journal* Web site: http://blogs.wsj.com/cio/2013/04/18/farmers-experiment-with-drones/

Kirkland, R. (2013). Connecting everything: A conversation with Cisco's Padmasree Warrior. Retrieved February 2, 2016 from the McKinsey & Company Web site: http://www.mckinsey.com/Insights/High_Tech_Telecoms_Internet/Connecting_everything_A_conversation_with_Ciscos_Padmasree_Warrior?cid=disruptive_tech-eml-alt-mip-mck-oth-1305

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). *Big Data: The next frontier for innovation, competition, and productivity*. Retrieved February 2, 2016 from McKinsey & Company Web site: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Marr, B. (2014). How GE is using Big Data to drive business performance. Retrieved February 2, 2016, from SmartDataCollective Web site: http://www.smartdatacollective.com/bernardmarr/229151/how-ge-using-big-data-drive-business-performance

National Research Council. (1997). *Precision Agriculture in the 21st Century: Geospatial and Information Technologies in Crop Management*. Washington, D.C.: National Academies Press. Accessed October 30, 2015. http://www.nap.edu/catalog/5491/precision-agriculture-in-the-21st-century-geospatial-and-information-technologies

National Science Foundation. (2012). Core techniques and technologies for advancing big data science & engineering. Retrieved February 2, 2016, from the National Academies Press Web site: http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm

Reid, J. F. (2015). The impact of mechanization on agriculture. *The Bridge, 41*(3), 22-29. Retrieved from: https://www.nae.edu/File.aspx?id=52553

Remote sensing-based information and insurance for crops in emerging economics [RIICE]. (2013). About RIICE. Retrieved February 2, 2016, from International Rice Research Institute Web site: http://www.riice.org/about-riice/

Sampler, J. L., (1997). Redefining Industry Structure for the Information Age. *Strategic Management Journal*. 19:343-356.

Shapiro, C. & H. R. Varian. (1999). *Information Rules*. Boston: Harvard Business School Press.

Sonka, S. T., & Coaldrake, K. F. (1996). Cyberfarm: What does it look like? What does it mean? *American Journal of Agricultural Economics*, *78*(5), 1263-1268.

Sonka, S. T., Lins, D. A., Schroeder, R. C., & Hofing, S. L. (2000). Production Agriculture as a Knowledge Creating System. *International Food and Agribusiness Management Review*, *2*(2), 165-178.

Watson, H. J. (2013). The Business Case for Analytics. *Bized, 12*(3), 49-54.

Zott, C. & Amit, R. (2010). Designing your Future Business Model. *Long Range Planning*, *43*,216-226.

Zott, C., Amit, R., & Massa, L. (2011). The Business Model: Recent Developments and Future Research. *Journal of Management*, *37*(4), 1019-1042.

Zwilling, M. (2014, July 7). Where is your technology in the Gartner Hype Cycle? Retrieved February 2, 2016, from Startup Professionals Musings Web site: http://blog.startupprofessionals.com/2014/07/where-is-your-technology-in-gartner.html