Business Horizons (2017) xxx, xxx—xxx

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY

www.elsevier.com/locate/bushor

Addressing barriers to big data

Abdulkhaliq Alharthi [a], Vlad Krotov [b,*], Michael Bowman [b]

[a] *College of Business, Abu Dhabi University, P.O. Box 59911, Abu Dhabi, U.A.E.*
[b] *Murray State University, 109 Business Building, Murray, KY 42071, U.S.A.*

KEYWORDS
Big data;
Big data infrastructure;
Data privacy;
Organizational culture;
Competitive advantage;
Hadoop

Abstract Increasingly, big data is viewed as the most strategic resource of the 21st century, similar in importance to that of gold and oil. While sitting on these vast pools of data, many organizations are simply not ready to take advantage of this new strategic resource. Embracing big data requires addressing a number of barriers that fall into the domains of technology, people, and organization. A holistic, socio-technical approach is required to overcome these barriers. This article introduces the specific tactics we recommend for addressing big data barriers, which involve changes to technology infrastructure, a focus on privacy, promotion of big data and analytic skills development, and the creation of a clear organizational vision related to big data.

1. The era of big data

Recent advancements in information and communications technologies (ICT), as well as the ever-increasing affordability and ubiquity of networks and electronic devices, have resulted in a massive volume of data from various sources and in different formats. The volume of data available today is measured in zettabytes (ZB)—a measure equal to one trillion gigabytes (GB) and equivalent to the data storage capacity of about 250 billion DVDs. The world wide web alone was estimated to contain 0.5 ZB of data in 2009 (Fan & Bifet, 2013). This amount of data is available from more than one trillion web pages currently accessible on the web. The total amount of digital data in the whole world reached 1.8 ZB in 2011 and is predicted to grow to approximately 90 ZB by 2020 (Jeon, 2012). Currently, about 90% of the digital data available was created in the last 2 years (Gobble, 2013). These massive amounts of recently created digital data are often referred to as *big data*.

Participants of the 2012 World Economic Forum in Davos, Switzerland, declared that big data has become a strategic economic resource, similar in significance and liquidity to currency and gold (Johnson, 2012). While organizations today across the globe realize that big data analytics will be an

* Corresponding author
E-mail addresses: alharthi2000@yahoo.com (A. Alharthi), vkrotov@murraystate.edu (V. Krotov), michael.bowman@murraystate.edu (M. Bowman)

0007-6813/$ — see front matter
http://dx.doi.org/10.1016/j.bushor.2017.01.002

important source of competitive advantage in the future, a number of barriers prevent many organizations from fully exploiting the opportunities big data can offer. These barriers include outdated IT infrastructure, the inherent complexity and messiness of big data, lack of data science skills within organizations, privacy concerns, and organizational cultures that are not conducive to data-driven operations or data-driven decision making. This article proposes specific tactics for removing these barriers, namely: (1) the utilization of commodity hardware and specialized big data software, (2) collaboration with educational institutions, (3) installation of policies and processes that would support individual privacy, and (4) development of a clear organizational vision in relation to big data.

2. What is big data?

The term big data is used to describe the massive volume of digital data produced by human activity that is very difficult to manage using conventional data analysis tools. While scholars and practitioners may have different views on what precisely is meant by big data, the Gartner Group's definition is widely used. According to the definition (Gartner, 2017), big data is characterized by the 3 Vs:

- *Volume* refers to the vast quantity of structured and unstructured data that is hard to collect, manage, and analyze with the existing IT infrastructure and tools; thus, these massive data sets require new and innovative tools and approaches for capturing, storing, and analyzing data.
- *Variety* refers to the fact that the data comes from various sources such as spreadsheets, traditional databases, text documents, and digital data streams.
- *Velocity* refers to the fact that these big data sets are often comprised of and continuously expanded by real-time data streams.

The term big data refers not only to data, but also to the tools and practices for analyzing, processing, and managing these massive, complex, and rapidly evolving data sets. Therefore, the terms big data and big data analytics are used in this article interchangeably.

Big data analytics is similar to the concept of business intelligence (BI). Both concepts are used to denote usage of data management technologies and computer-based analytical tools for discovering actionable business knowledge and facilitating decision making based on organizational data. Yet these two concepts have three major differences related to the type of data used by BI and big data. These differences are closely related to the 3 Vs discussed previously (see Table 1). BI architecture is often conceptualized as an organizational data architecture that relies on a finite set of highly structured data sources (e.g., various internal and external databases) that are accessed in an offline mode (e.g., via a specialized data mart extracted from an organizational data warehouse). In contrast, big data movement aims to develop data management technologies and analytical tools that can handle an infinite number of data sets comprised of highly complex and 'messy' data in virtually real-time formats.

Table 1. Big data vs. business intelligence

Characteristics	Big data	Business intelligence
Volume	Infinite	Finite
Velocity	Real time	Offline
Variety	Unstructured	Structured

3. Big data opportunities

Organizations across the globe increasingly have come to the realization that the ability to analyze and use big and complex data sets will be the most important source of competitive advantage in the 21st century. Big data has the potential to deliver better customer experience, enhance internal efficiency, and, ultimately, improve profitability and competitiveness of organizations across all industries. Organizations can use big data to get smarter and innovative in ways that were not possible before the advent of the 'zettabyte era' (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Some examples of successful big data initiatives are discussed in the next sections.

3.1. Etihad Airways

Based in Abu Dhabi, United Arab Emirates, Etihad Airways flies to more than 89 destinations across the globe and carries 10 million passengers annually. Since the company's founding in 2003, Etihad began using hundreds of sensors on every plane to generate data about its fleet. This arrangement is similar to the Internet of Things (IoT) concept, in that a number of electronic (e.g., computers) and non-electronic objects (e.g., an airplane tire equipped with a wireless pressure sensor) are connected into one single network and exchange information in real time. The network generates vast volumes of digital data. This data is analyzed by Etihad and used to control the entire fleet of airplanes. One specific application of Etihad's IoT is predictive maintenance. The company monitors the state and location of its fleet in real time and strives to preempt problems. This results in substantial cost savings for the company (Van Rijmenam, 2014).

3.2. Walt Disney

Every year, nearly 100 million people visit Walt Disney parks around the world. In 2013, the company introduced a park entry pass in the form of a bracelet equipped with a radio-frequency identification (RFID) sensor. RFID sensors are similar to barcodes in their purpose, with the difference being that RFID sensors are small wireless computer devices that can transmit digital information about the object that they are attached to automatically and wirelessly, over a substantial distance, and without line of sight required. These RFID-equipped bracelets, called MagicBands, offer visitors benefits such as jumping the lines and pre-booking rides. While offering these benefits, the MagicBand also records the visitor's movement and facilitates automatic collection of large volumes of complex and valuable data about visitors and their interactions with Disney's parks. Disney uses these data sets, in conjunction with several big data tools, to analyze customer behaviors, including purchase history, waiting time, and preferences. This information helps Disney provide better experiences for park visitors and improve its marketing effectiveness. All this helps the company attract more customers and increase their revenues per customer (Van Rijmenam, 2014).

3.3. United Parcel Service (UPS)

UPS tracks data on 16.3 million packages per day for 8.8 million customers. The company uses telematics sensors in over 46,000 UPS delivery vehicles to generate vast amounts of data, including the delivery truck speed, location, and overall state of the vehicle. Use of big data analytics in UPS optimizes truck movement and results in savings in fuel consumption by cutting millions of miles off daily routes. The company is currently working on a similar project in hopes that it can optimize the daily flights of the company's aircraft fleet (Davenport & Dyché, 2013).

3.4. Dublin City Council

The Dublin City Council is responsible for providing housing, water, and transport services to the 1.2 million residents of Ireland's capital. Nearly 1,000 public buses must be managed to ensure smooth operations of the city's public transportation system. The city council equipped all buses with GPS sensors and collects the geospatial data from these sensors in real time. The data collected is analyzed to optimize bus routing and utilization. Optimized bus routes help the council save money on fuel and decrease the level of air pollution emitted from the transportation system (IBM, 2014b).

3.5. The Daimler Group

The Daimler Group is the company behind Mercedes-Benz Cars, Daimler Trucks, Mercedes-Benz Vans, Daimler Buses, and Daimler Financial Services. Daimler produces around 10,000 cylinder heads daily that are used in car engine production. During the manufacturing process, Daimler continuously collects more than 500 data attributes of cylinder heads such as physical dimensions and cylinder temperature. The data is collected and analyzed by a comprehensive solution provided by IBM that is based on SPSS, the widely-used statistical analysis and data mining package. IBM's SPSS solution helps the company predict errors and correct variances before cylinder head defects can occur. Two years after the initiative began, Daimler cylinder production increased by about 25% (IBM, 2014a).

4. Barriers to big data

In a study investigating the organizational and technology management practices at 330 public companies in North America, it was discovered that many organizations were not ready to embrace the leveraging of big data for improving organizational performance (McAfee & Brynjolfsson, 2012), as it requires overcoming a number of barriers (or challenges) in relation to big data. These barriers range from developing new employee skills and upgrading IT infrastructure, to instilling new management practices or a new organizational culture across the entire organization (Manyika et al., 2011). These challenges are discussed next.

4.1. Infrastructure readiness

The development of IT infrastructure for big data analytics requires significant investments in software and hardware to support the analysis of hundreds of millions of records in real time. Most of the existing technologies utilized were not designed to meet the growing requirements of big data analytics. Cloud and heterogeneous computing architectures (i.e., architectures combining different types of processors over the internet) are believed to offer solutions for big data. But these technologies often fail, either from a technical or economic point of view, when a large amount of data is processed. For example, 1,000 processing nodes connected over a cloud would require approximately 750 days to process one petabyte of data (one petabyte equals one million gigabytes). Moreover, this system would cost more than $6 million to assemble (Trelles, Prins, Snir, & Jansen, 2011). Thus, cloud and heterogeneous computing may not help organizations avoid data processing bottlenecks that result from the use of big data.

4.2. Complexity

An organization wishing to leverage the power of big data may face significant problems related to data complexity and its inherent messiness. Today, digital data is often stored in many different formats, including unstructured databases and discrete text files (Douglas, 2013). Moreover, the volume of data is increasing day by day, making the task of handling data from various sources and in different formats even more challenging (Johnson, 2012). The specific organizational and technological factors that contribute to the challenges associated with data complexity are described as follows.

4.2.1. Rate of data growth

Most organizations do not have a plan in place to address the problem of the large rate of data growth at reasonable cost. Many organizations simply opt to delete old data instead of trying to accommodate data growth.

4.2.2. Multiple sources of data

The challenge of combining data from different sources arises from the different ways data is organized in various sources (the so-called 'semantic conflict' between various data sources). For example, the term earnings may mean profit in one data set but it may represent revenue or sales in another.

4.2.3. Multiple formats of data

Today's data is frequently stored in many different formats, both structured and unstructured. Examples of unstructured data include text documents, SMS, images, videos, audio files, emails, massages, and transaction files. Structured data is organized typically in relational databases. The challenge of managing and analyzing different data formats is often beyond the capability of many organizations.

4.3. Lack of skills

Currently, the lack of employees with big data or general analytics skills is one of the major challenges facing organizations seeking to embrace big data. According to the U.S. Department of Labor, the shortage of people with big data skills in the U.S. alone is predicted to be between 120,000 and 190,000 by 2018 (Douglas, 2013). Lack of data analytics skills among existing employees may increase data entry errors that could result in placing information in the wrong record, losing valuable information, and limiting the value a business can derive from the data that it captures (Hoffman & Podgurski, 2013). To make things worse, many organizations do not have the technology to recognize and recover missing or erroneous data (Schouten, 2013).

4.4. Privacy

Concerns over privacy often hinder adoption and use of big data analytics within organizations. Big data analytics often employs personal data (i.e., the data that was gathered for specific purposes with the consent of the individual) that was collected for an entirely different purpose (Douglas, 2013). Combining personal information with other data sources can create numerous legal and ethical challenges, such as the potential to reveal and leak private information about the individual (e.g., medical records, financial situation, embarrassing behavior, family relationships). What makes the problem even more cumbersome is that many organizations are not explicit about how they use their customer data. For example, it has been argued repeatedly that Twitter, Facebook, and WhatsApp are not that clear in their privacy policies regarding their current and future activities in relation to the data about consumers they collect. At one point, Twitter was suspected of selling a large number of tweets to a big data dealer. Similarly, WhatsApp is facing the accusation of using customer data for commercial purposes and violating customer rights to privacy (Van Rijmenam, 2014). Facebook is also known to store permanently all data that a user generates. The public is wondering what would be the potential uses of data accumulated by Facebook over the years. Some describe this data as a ticking privacy bomb.

4.5. Cultural barriers

Organizational culture—through its assumptions, values, norms, and symbols—has a strong impact on various aspects of an organization, such as strategy, structure, and processes. Top management will select the strategy and design the organization structure based on the way they perceive and understand organizational values and norms. Thus, many obstacles in relation to big data are likely to be related to organizational culture and not to data or technology. For example, some organizations lack the overall understanding of how big data can improve their business operations and, consequently, see little value in pursuing big data initiatives (LaValle et al., 2011). This can result in organizational resistance to big data.

5. Addressing barriers to big data

Leveraging big data as a vehicle for improving organizational performance requires that a company address the barriers related to people, technology, and organizational domains. Table 2 provides a summary of some generic tactics for addressing the barriers discussed earlier. The table is followed by a detailed discussion on our proposed remedies to big data barriers.

5.1. Technological barriers

5.1.1. Infrastructure readiness

Building a new and independent platform for big data analytics is the best option for new organizations. However, this option is not practical for existing organizations with legacy IT systems. In most cases, it is more feasible to find a solution that integrates modern big data platforms with the existing legacy IT systems (Davenport & Dyché, 2013).

Luckily, big data platforms can be designed and developed using low-cost commodity hardware (that also typically comprise legacy IT systems). Big data infrastructure should be able to store and process massive volumes of data in real time and must be protected against service interruption or failure. Use of a large number of commodity servers to store and process data simultaneously can meet these requirements. The commodity servers are classified either as slave nodes or as master nodes. The master nodes are used to manage data while the slave nodes are used to store and process data. Big data storage systems can be built upon

Table 2. Barriers to big data and potential solutions

Barrier	Literature sources	Possible solutions
1. Technological barriers		
Infrastructure readiness	Trelles et al., 2011	Commodity hardware should be used to enhance processing power and storage capacity.
Complexity of data	Douglas, 2013; Johnson, 2012	Specialized software tools and algorithms should be used, such as MapReduce and Hadoop, to store, manage, and analyze complex data in a more efficient, reliable, fast and economical manner (Van Rijmenam, 2014).
2. Human barriers		
Lack of skills	Douglas, 2013; Hoffman & Podgurski, 2013; Schouten, 2013	Organizations should collaborate with educational institutions to align their educational curriculum with the industry requirements for big data skills. Industry and academic institutions should collaborate on providing practical training to address missing skills in the field of data analytics and big data (Miller, 2014).
Privacy	Douglas, 2013; Van Rijmenam, 2014	Legislative bodies are required to endorse laws that protect individuals from the misuse of data (Schadt, 2012). Organizations need to accommodate this legislation as well as incorporate general best practices for handling sensitive customer data into their policies and operations.
3. Organizational barriers		
Culture	LaValle et al., 2011; McAfee & Brynjolfsson, 2012	Successful cultural change can be achieved by having a clear organizational vision in relation to big data (Rogers et al., 2006).

these commodity servers or nodes. This kind of architecture is cost efficient and scalable; it allows for an increase in the number of nodes to improve the processing power and storage capacity with relative ease (Davenport & Dyché, 2013).

While developing big data infrastructure based on commodity hardware, one should not forget that these commodity servers and storage systems are connected via ethernet or fiber networks. Network bandwidth is critical for data exchange across servers. Therefore, it is important that the network infrastructure support the high throughput and bandwidth associated with high volumes of data moving across servers.

5.1.2. Complexity of data

Due to increasing amounts of web data after the year 2000, Google and other internet-based organizations encountered computational difficulties in indexing its digital content. To address the issue, Google Labs published a paper in December 2004 on a new algorithm called MapReduce. MapReduce was defined as a model used for parallel processing of large data sets across large numbers of commodity computers (nodes). The algorithm helped Google to improve its search index systems and analyze internet content and users' online behavior in a more efficient, reliable, fast, and economical manner (Van Rijmenam, 2014).

Doug Cutting, an employee of Yahoo! at that time, took a great interest in the MapReduce algorithm publicized by Google. He decided to create his own version of the technology that would allow applications based on the MapReduce algorithm to run on commodity hardware. Cutting named the technology Hadoop after his son's toy elephant. Hadoop, now synonymous with big data, was adopted initially by Yahoo! and later by a number of other prominent organizations. In 2008, Hadoop was released as an open-source tool. Today, it is managed by the non-profit Apache Software Foundation (Van Rijmenam, 2014). Hadoop's architecture is scalable (i.e., new nodes can be added as needed), flexible (i.e., can absorb any type of data from any source), cost effective (i.e., uses commodity computers), and reliable (i.e., the system redirects work to another node when a particular node fails). These characteristics make it low-cost and practical compared to proprietary big data products (Davenport & Dyché, 2013; Van Rijmenam, 2014).

Besides the open-source version of Hadoop (e.g., Apache Hadoop), several commercial distributions of Hadoop are available in the marketplace. A number of organizations have created their own versions. With these proprietary Hadoop versions, the customer receives additional software features, training, documentation, and other services. The most recognized enterprise versions of Hadoop include Cloudera, MapR, and Hortonworks. In addition to the mentioned distributions, IBM and EMC have introduced their own Hadoop distributions. Well-known software companies like Microsoft, Teradata, and Oracle provide support for such Hadoop-based big data platforms as Cloudera and Hortonworks.

5.2. Human barriers

On the people side, lack of big data and general analytical skills, as well as privacy concerns among the public, are considered to be the most challenging barriers for many organizations when pursuing big data initiatives. Each of these two barriers is discussed in detail below.

5.2.1. Lack of skills

Given the growing importance of big data, organizations will have an unmet demand for data scientists, data analysts, and big data technology experts. Organizations can collaborate with educational institutions to align their curriculum with the industry requirements in relation to big data skills. Most likely, a dedicated chief data officer (CDO) would help to manage this kind of collaboration as well as other initiatives related to big data.

Some universities, working to meet industry needs, have already started training students in big data technologies and analytics skills (Miller, 2014). The key skill required for big data is, primarily, knowledge in big data technologies and platforms (e.g., Hadoop, MapReduce, In-Memory DB, NoSQL). Secondary skills, which include advanced skills in math and statistics, machine learning, predictive analytics, decision-making models, and data visualization, should also be incorporated into training programs (Davenport & Harris, 2007; McAfee & Brynjolfsson, 2012; Miller, 2014; Van Rijmenam, 2014).

5.2.2. Privacy

The rapid advancements in big data technologies have been closely associated with challenges related to privacy and the protection of personal information. By definition, big data involves the collection of an immense volume of data. At least some of this data is likely to be personal or sensitive in nature. As personal and sensitive information becomes more exposed in the era of big data, the need to protect individuals' privacy becomes essential to avoid legal or ethical controversies and to ensure customer buy-in of big data initiatives.

Although technology glitches may lead to privacy or security breaches, it is the behavioral side of privacy and security that is often most problematic. In a way, it does not matter how strong or advanced the technical dimension of security is as long as humans are in charge of the data. For example, many well-known security breaches (e.g., the ones documented by WikiLeaks) involved employees simply copying and distributing data to which they had access.

Currently, several behavioral solutions are proposed to enable individuals to maintain control over their data and avoid privacy and security incidents. One example of a proposed practice in this area includes the right of individuals to delete their historical data, the right to specify expiration dates for data, and the right of individuals to have ownership of information about their social connections. Solidifying these practices may require the creation and adoption of laws that formalize these ideas (Schadt, 2012). These laws and policies will not only protect consumer privacy, but also encourage individuals to share their personal information in a way that benefits both the consumers and the organizations relying on this data for improving their performance.

5.3. Organizational barriers

The cultural barriers related to big data are considered to be significant and challenging issues to overcome. Therefore, an organization must first modify its culture so that it is supportive of fact-based decision making in order to take full advantage of big data opportunities. Successful cultural change of this nature can be achieved by documenting, implementing, and communicating a clear organizational vision in relation to big data, ensuring top management commitment to this vision, and managing the drivers that influence organizational culture rather than trying to manage culture itself (Rogers, Meehan, & Tanner, 2006). Developing a clear vision of how big data fits in with the overall strategy of an organization should accelerate and solidify the acceptance of big data within the organization. Of course, once the vision is formulated, it has to be translated into specific organizational processes and broader initiatives that rely on big data to improve organizational performance.

6. Big data roadmap

Overcoming challenges and taking full advantage of the opportunities offered by the big data era requires a holistic, socio-technical approach. First, an organization needs to create and foster an organizational culture that is supportive of fact-based decision making and big data analytics. Developing a clear vision of how big data fits with the overall strategy of an organization should help accelerate and solidify the acceptance of this type of organizational culture. Once the vision is formulated, it has to be translated into specific organizational processes and initiatives that rely on big data to improve organizational performance. In the people domain, organizations should collaborate with educational institutions to acquire and develop the skills necessary for big data. Moreover, organizations should improve existing processes related to big data to include privacy protection measures. These measures should be supported by legislation. These collaboration and privacy protection initiatives should be managed by a CDO. The CDO should also be in charge of necessary technology improvements, the most important of which is developing an effective yet cost-efficient big data platform that consists of commodity hardware and specialized big data software capable of capturing, storing, and analyzing vast and unstructured data sets.

References

Davenport, T. H., & Dyché, J. (2013). Big data in big companies. *SAS Institute Inc*. Available at http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Brighton, MA: Harvard Business School Press.

Douglas, M. (2013). Big data raises big questions. *Government Technology, 26*(4), 12—16.

Fan, W., & Bifet, A. (2013). Mining big data: Current status and forecast to the future. *ACM SIGKDD Explorations Newsletter, 14*(2), 1—5.

Gartner. (2017). *Big data*. Available at http://www.gartner.com/it-glossary/big-data/

Gobble, M. M. (2013). Big data: The next big thing in innovation. *Research Technology Management, 56*(1), 64—66.

Hoffman, S., & Podgurski, A. (2013). Big bad data: Law, public health, and biomedical databases. *Journal of Law, Medicine, and Ethics, 41*(S1), 56—60.

IBM. (2014a). *Automotive manufacturer increases productivity for cylinder-head production by 25 percent* [White paper]. Available at http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=AB&htmlfid=YTC03659WWEN

IBM. (2014b). *Dublin City Council: Traffic flow improved by big data analytics used to predict bus arrival and transit times* [White paper]. Available at http://www.slideshare.net/UKBigDataStories/dublin-city-council

Jeon, Y.-H. (2012). Impact of big data: Networking considerations and case study. *International Journal of Computer Science and Network Security, 12*(12), 30—34.

Johnson, J. E. (2012). Big data + big analytics = big opportunity. *Financial Executive, 28*(6), 50—53.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics, and the path from insights to value. *MIT Sloan Management Review, 52*(2), 21—31.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011, May). *Big data: The next frontier for innovation, competition, and productivity*. Available at http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

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

Miller, S. (2014). Collaborative approaches needed to close the big data skills gap. *Journal of Organization Design, 3*(1), 26—30.

Rogers, P., Meehan, P., & Tanner, S. (2006). Building a winning culture [White paper]. *Bain and Company*. Available at http://growthtrilogy.com/bainweb/PDFs/cms/Public/BB_Building_winning_culture.pdf

Schadt, E. E. (2012). The changing privacy landscape in the era of big data. *Molecular Systems Biology, 8*(1), 1—3.

Schouten, P. (2013). Big data in health care solving provider revenue leakage with advanced analytics. *Healthcare Financial Management, 67*(2), 40—42.

Trelles, O., Prins, P., Snir, M., & Jansen, R. C. (2011). Big data, but are we ready? *Nature Reviews Genetics, 12*(3), 224.

Van Rijmenam, M. (2014). *Think bigger: Developing a successful big data strategy for your business*. New York: AMACOM.