Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care

YiChuan Wang
Raymond J. Harbert
College of Business, Auburn University
yzw0037@auburn.edu

LeeAnn Kung
Raymond J. Harbert
College of Business, Auburn University
hzt0007@auburn.edu

Chaochi Ting
IBM Global Service
tingchaochi@hotmail.com

Terry Anthony Byrd
Raymond J. Harbert
College of Business, Auburn University
byrdter@auburn.edu

Abstract

To date, the health care industry has paid little attention to the potential benefits to be gained from big data. While most pioneering big data studies have adopted technological perspectives, a better understanding of the strategic implications of big data is urgently needed. To address this lack, this study examines the development, architecture and component functionalities of big data, and identifies its capabilities, including traceability, the analysis of unstructured data and patterns of care, and its predictive capacity to support healthcare managers seeking to formulate more effective big-data-based strategies. Our findings will help healthcare organizations respond strategically to the challenges they face in today’s highly competitive healthcare market.

1. Introduction

To cope with the high volume digital flood of information that is being generated at ever-higher velocities and varieties, healthcare entities are seeking a breakthrough in data handling that will help them to improve their healthcare services and create new, more effective, business models. A likely concept of such a breakthrough is information technology (IT)-enabled transformation [1], which consists of a series of sequential changes leading to the operational improvement and internal integration of IT functionalities and the transformation of IT capabilities to create competitive advantage and improve financial performance through business redesign [2]. With this in mind, healthcare managers believe that the adoption and innovation of medical technologies and healthcare information systems holds great promise for reducing care delivery costs and improving service quality and clinical performance [3]. However, IT-related challenges such as inadequate integration of healthcare technologies and systems, poor healthcare information management are seriously hampering efforts to transform IT value to business value in the U.S. healthcare sector. The consequences are unnecessary increases in medical costs and time for both patients and healthcare service providers. There is therefore an urgent need for healthcare industries to seek effective IT artifacts that will enable them to consolidate organizational resources and deliver a high quality patient experience and improve organizational performance.

Big data, an overwhelming phenomenon which has been addressed through various new and old data management technologies in terms of healthcare informatics, data analytics and business intelligence (BI), hold the key to healthcare transformation. Traditional data management in healthcare provides BI tools and applications on a stand-alone system along with limited clinical data processing ability [4]. New data management tools such as MongoDB, MarkLogic and Apache Cassandra for data integration and retrieval, allow data being transferred between traditional and new big data management systems. To store the big volume and various formats of data, there are Apache HBase and NoSQL systems. Big data platforms and tools such as Hadoop/MapReduce, on the other hand, are ideal for all kinds of data repository in their inherent business object formats and are capable of processing an immense volume, variety and velocity of data across a wide range of healthcare platforms. Such functions facilitate information integration and analysis capabilities, and provide fresh business insights to help healthcare organizations meet future market needs and trends [5].

However, healthcare organizations have paid little attention to the potential benefits to be gained from big data. To understand big data, and how to adopt it successfully in healthcare, managers have to identify the strategic and business value of big data, rather than merely concentrating on a technological understanding of its implementation [3, 6]. In this paper, we begin by providing the historical context and current status of big data in healthcare, and then move on to define big data capabilities in healthcare from various
perspectives and develop a big data architecture. We have conducted a content analysis of 26 big data implementation cases in healthcare to identify five major capabilities that will help healthcare organizations to acquire a comprehensive understanding of the potential values of big data. Finally, we present several strategies for introducing big data solution to healthcare practitioners.

2. Big data: the historical context

The history of big data is inextricably linked with that of data science. The term “big data” was used for the first time in 1997 by Michael Cox and David Ellsworth in a paper presented at an IEEE conference to explain the visualization of data and the challenges it posed for computer systems [7]. By the end of the 1990s, the rapid improvements in technology had begun to generate a great deal of data but this was accompanied by little in the way of usable information. The resulting development of the concept of business intelligence has emphasized the importance of the collection, integration, analysis, and interpretation of business information and how this can help a business to make more appropriate decisions and acquire a better understanding of market behaviors and trends.

The period from 2001 to 2008 was the evolutionary stage for big data development. Big data was first defined in terms of the volume, velocity, and variety of its data (3Vs), after which it became possible to develop more sophisticated software to suit the needs of the information explosion. Software and application developments like Extensible Markup Language (XML) Web services, database management systems, and Hadoop added analytics modules and functions to core modules that focused on enhancing usability for end users, and enabled users to process huge amounts of data across and within organizations collaboratively and in real-time. At the same time, healthcare organizations were starting to digitize their medical records and aggregate clinical data in huge electronic databases. This development made the health data storable, usable, searchable, and actionable, and helped healthcare providers practice more effective medicine.

At the beginning of 2009, big data entered the revolutionary stage [8]. Not only had big-data computing become a breakthrough innovation for business intelligence, but researchers were predicting that data management and its techniques were about to shift from structured data into unstructured data and from a static terminal environment to a ubiquitous cloud-based environment. Big data computing pioneer industries such as banks and e-commerce were beginning to have an impact on improving business processes and workforce effectiveness, reducing enterprise costs and attracting new customers. As of 2011, stored health care data had reached 150 exabytes (1 EB = 10^18 bytes) worldwide, mainly in the form of electronic health records [9]. However, most of the potential for value creation is still in its infancy, because predictive modeling and simulation techniques for analyzing healthcare data as a whole have not yet been adequately developed.

The recent trend has been towards the use of cloud based big data, which emphasizes how enterprises have increasingly adopted a “big data in the cloud” solution as software-as-a-service (SaaS) that offers an attractive alternative at a lower cost. According to the Gartner’s IT trend prediction [10], taking advantage of cloud computing services for big data management systems that support a real-time analytic capability and cost-effective storage will become a preferred IT solution by 2016. The main trend in the healthcare industry is a shift in data type from structure-based to semi-structured based (e.g., home monitoring, telehealth, sensor-based wireless devices) and unstructured data (e.g., transcribed notes, images, and video). The increasing use of sensors and remote monitors are key factors supporting the rise of home healthcare services, meaning that the amount of data being generated from sensors will continue to grow significantly. This will in turn improve the quality of healthcare services through more accurate big data analysis and prediction.

3. Big data architecture in health care

To identify the potential benefits offered by big data, it is necessary to understand its architecture and component functionalities. For formulating the best practice of big data architecture in healthcare, we invited four IT experts (two practitioners and two academics) to participate in the five-round brainstorming and discussion. Big data architecture consists of a data-based logical framework that starts with data capture, proceeds via data transformation, and culminates with data consumption [3]. Figure 1 depicts a best practice big data architecture that is loosely comprised of five major architectural layers: (1) data, (2) data aggregation, (3) analytics, (4) information exploration, and (5) big data governance. These logical layers make up the big data components that perform specific functions, and will therefore enable healthcare managers to understand how to transform the healthcare data from various sources into meaningful clinical information through big data implementations.

Data layer. This layer includes all the data sources necessary to provide the insights required to support
daily operations and solve business problems. The data is divided into structured data such as traditional EHRs records, semi-structured data such as the logs of health monitoring devices, and unstructured data such as clinical images. These data are collected from various locations inside the hospital or external units, and will be stored immediately into appropriate databases, depending on the content format.

Data aggregation layer. This layer is responsible for handling data from the various data sources. In this layer, data will be intelligently digested by performing three steps: data acquisition, transformation, and storage. The primary goal of data acquisition is to read data provided in various communication channels, frequencies, sizes, and formats. This step is often a major obstacle in the early stages of implementing big data, because these incoming data characteristics might vary considerably. Here, the cost may well exceed the budget available for establishing new data warehouses, and extending their capacity to avoid workload bottlenecks. During the transformation step, the transformation engine must be capable of moving, cleaning, splitting, translating, merging, sorting, and validating data. For example, structured data such as that typically contained in an eclectic medical record might be extracted from healthcare information systems and subsequently converted into a specific standard data format, sorted by the specified criterion (e.g., patient name, location, or medical history), and then the record validated against data quality rules. Finally, the data are loaded into the target databases such as Hadoop distributed file systems (HDFS) or in a Hadoop cloud for further processing and analysis. The data storage principles are based on compliance regulations, data governance policies and access controls, and data storage methods can be implemented and completed in batch processes or in real time.

Analytics Layer. This layer is responsible for processing all kinds of data and performing appropriate analyses. In this layer, data analysis can be divided into three major components: Hadoop Map/Reduce, stream computing, and in-database analytics, depending on the type of data and the purpose of the analysis. Mapreduce is the most commonly used programming model in big data analytics and provides the ability to process large volumes of data in batch form cost-effectively, as well as allowing the analysis of both unstructured and structured data in a massively parallel processing (MPP) environment. Stream computing can support high performance stream data processing in near real time or real time. With a real time analysis, users can track data in motion, respond to unexpected events as they happen and quickly determine next-best actions. For example, in the case of healthcare fraud detection, stream computing is an important analytical tool that assists in predicting the likelihood of illegal transactions or deliberate misuse of customer accounts. Transactions and accounts will be analyzed in real time and alarms generated immediately to prevent myriad frauds across healthcare sectors. In-database analytics refers to a data mining approach built on an analytic platform that allows data to be processed within the data warehouse. This component provides high-speed parallel processing, scalability, and optimization features geared toward big data analytics, and offers a secure environment for confidential enterprise information. However, the results provided from in-database analytics are neither current nor real time and it is therefore likely to generate reports with a static prediction. Typically, this analytic component in healthcare organizations is useful for supporting preventative healthcare practice and improving pharmaceutical management.

Information exploration layer. This layer generates outputs such as the various visualization reporting options, real-time monitoring of information, and meaningful business insights derived from the analytics platforms to users in the organization. Similar to traditional business intelligence platforms, the reporting is a critical big data feature that allows data to be visualized in a useful way to support users’ daily operations and help managers to make faster, better decisions. However, the most important output for healthcare may well be its real-time monitoring of information such as alerts and proactive notifications, real time data navigation, and operational key performance indicators (KPIs). This information is analyzed from sources such as smartphones and personal medical devices and can be sent to interested users or made available in the form of dashboards in real time for monitoring patients’ health and preventing accidental medical events. The analytics layer also provides exceptional support for evidence based medical practices by analyzing EHRs, patterns of care, care experience, and individual patients’ habits and medical histories.

Big data governance layer. Big data governance is the pillar of big data architecture that affects all of the logical layers. This layer is comprised of master data management (MDM), data life-cycle management, and data security and privacy management that emphasizes how to harness data in the organization. The first component of data governance, master data management, is regarded as the processes, governance, policies, standards, and tools for managing data. Data is properly standardized, removed, and incorporated in order to create the immediacy, completeness, accuracy, and availability of master data for supporting data analysis and decision making. The Second component, data life-cycle management, is the process of managing
business information throughout its lifecycle, from archiving data, through maintaining data warehouse, testing and delivering different application systems to deleting and disposing of data. By managing data effectively over its lifetime, firms are better equipped to provide competitive offerings to meet market needs and support business goals with lower timeline overruns and cost. The third component, data security and privacy management, is the platform for providing enterprise-level data activities in terms of discovery, configuration assessment, monitoring, auditing, and protection [11]. It is essential to implement rigorous data rules and control mechanisms for highly sensitive healthcare data to prevent security breaches and protect patient privacy. By adopting suitable policies, standards, and compliance requirements to restrict users’ permissions will ensure the new system satisfies healthcare regulations and creates a safe environment for the proper use of patient information.

4. Toward a definition of big data capability

In the era of data explosion, a collection of data sets can easily become so large and complex that it becomes difficult to process using traditional database management tools or data processing applications. Big data management systems have had to be developed to address this dilemma. According to the resource-based view (RBV) [12, 13], big data management systems could be logically expected to generate big data-specific IT capabilities and several definitions of big data capability have therefore been developed. Big data capability refers to the ability to manage a huge volume of disparate data to allow real-time analysis and reaction [14]. Bharadwaj et al. [15] also define big data capability as the ability to capture, curate, manage, and process the data within a specified elapsed time. Big data capability may offer several benefits in managing customer service over traditional information systems, including the ability to gather unstructured or semi-structured data from current and former customers to gain useful knowledge to support better decision-making [16], to predict customer behavior via predictive analytics software, and to retain valuable customers by providing real-time offers [17].

5. Research Methodology

5.1. Cases collection

Our cases were drawn from materials on current and past big data projects from multiple sources such as practical journals, print publications, case collections, and companies', vendors', consultants' or analysts'
reports. The absence of academic discussion in our case collection about the utilization of big data is due to the incipient nature of such in the field of healthcare. The following case selection criteria were applied: (1) the case presents an actual implementation of big data; (2) it clearly describes the software they introduce and benefits obtaining from big data. We were able to collect 26 big data cases specifically related to the healthcare industry. Of these sources, we classified 15 sources (58%) as material released by vendors or companies, 2 sources (8%) as originating from journal databases, and 9 sources (34%) as print publications, including healthcare institute reports and case collections. Categorizing by region, 16 cases were collected from Northern America, 8 cases from Europe, and others from Asia-Pacific region. The cases we used are listed in Appendix 1.

5.2. Research approach and process

Content analysis is a method for extracting various themes and topics from text, and it can be understood as, “an empirically grounded method, exploratory in process, and predictive or inferential in intent” [18, p. xvii]. Specifically, this study followed inductive content analysis, because the knowledge about big data implementation in healthcare is fragmented [4]. A three-phase research process for inductive content analysis (i.e., preparation, organizing, and reporting) that reported by Elo and Kyngäs [19] was performed in order to ensure a better understanding of big data capabilities and benefits in the healthcare context.

The preparation phase starts with selecting the “themes” (informative and persuasive nature of case material), which can be sentences, paragraphs, or a portion of a page [19]. For this study, themes from case materials were captured by a senior consultant who has over 15 years working experience with a multinational technology and consulting corporation headquartered in the United States, and currently is involved in several big data projects. The senior consultant manually highlighted the textual contents that completely describe how a big data solution and its functionalities create the big-data-enabled IT capabilities and potential benefits while reading through all 26 big data cases for a couple of times. Subsequently, a total of 136 statements directly related to the research objectives were obtained and recorded in a Microsoft Excel spreadsheet.

The second phase is to organize the qualitative data emerged from phase one through open coding, creating categories and abstraction [19]. In the process of open coding, the 136 statements were analyzed by one of the authors, and then grouped into preliminary conceptual themes based on their similarities. The purpose is to reduce the number of categories by collapsing those that are similar into broader higher order generic categories [20, 21, 22]. In order to increase the interrater reliability, the second author went through the same process independently. The two coders agreed on 84% of the categorization. Most discrepancies occurred between the two coders on the categories of analytical capability. Ensuring interrater reliability led to much discussion and debate. Once conflict occurred, the two coders reassessed each case and eventually arrived at a consensus. After consolidating the results of the coding, two coders named each generic category of big data capabilities using content-characteristic words.

Overall, the five generic categories of big data capabilities we identified in our review of the cases are traceability, unstructured data analytical capability, analytical capability for patterns of care, decision support capability, and predictive capability.

6. Capability profile of big data in health care

Prior to adopting big data, healthcare organizations need to identify the inherent capabilities provided by big data solutions that will help them acquire organizational agility and operational optimization, thereby enhancing their competitive advantage and creating business values and profits. These capabilities are derived from the various design principles and functionalities of big data and are confirmed by the real-world use of big data in healthcare contexts. These are described in turn below.

6.1. Traceability

Traceability is the ability to track output data from all the system’s IT components throughout the organization’s service units. Healthcare-related data such as activity and cost data, clinical data, pharmaceutical R&D data, patient behavior and sentiment data are commonly collected in real time or near real time from payers, healthcare services, pharmaceutical companies, consumers and stakeholders outside healthcare [23]. Traditional methods for harnessing these data are insufficient when faced with the volumes experienced in this context, which results in unnecessary redundancy in data transformation and movement, and a high rate of inconsistent data. Using big data algorithms, on the other hand, enables authorized users to gain access to large national or local data pools and capture patient records simultaneously from different healthcare systems or devices. This not only reduces conflicts between different healthcare sectors, but also decreases
the difficulties in linking the data to healthcare workflow for process optimization.

The primary goal of traceability is to make data consistent, visible and easily accessible for analysis. Traceability in healthcare facilitates monitoring the relation between patients’ needs and possible solutions through tracking all the datasets provided by the various healthcare services or devices. For example, the use of remote patient monitoring and sensing technologies has become more widespread for personalized care and home care in U.S. hospitals. Big data, with its traceability, can track information that is created by the devices in real time, such as the use of Telehealth Response Watch in home care services. This makes it possible to gather location, event and physiological information, including time stamps, from each patient wearing the device. This information is immediately deposited in appropriate databases (e.g., NoSQL and the Hadoop distributed file system), with excellent suitability and scalability for review by medical staff when needed. Similarly, incorporating information from radio frequency identification devices (RFID) into big data systems enables hospitals to take prompt action to improve medical supply utilization rates and reduce delays in patient flow. A case study at Brigham and Women’s Hospital (BWH) provides a typical example of the use of in-depth traceability in large longitudinal healthcare databases to identify drug risk. By integrating big-data algorithms into the legacy IT systems, medical staff can automatically monitor drug safety by tracking warning signals triggered by alarm systems.

6.2. Unstructured data analytical capability

Analytical capability refers to the analytical techniques typically used in a big data management system to process data with an immense volume (from terabytes to exabytes), variety (from text to graph) and velocity (from batch to streaming) via unique data storage, management, analysis, and visualization technologies [17, 24]. An analytical process in a big data management system starts by acquiring data from both inside and outside the healthcare sectors, storing it in distributed database systems, filtering it according to specific discovery criteria, and then analyzing it to integrate meaningful outcomes for the data warehouse, as shown in Figure 2. After unstructured data has been gathered across multiple healthcare units, it is stored in a Hadoop distributed file system and NoSQL database that maintain it until it is called up in response to users’ requests. NoSQL databases support the storage of both unstructured and semi-structured data multiple sources in multiple formats in real time. The core of the analytic process is the MapReduce algorithms implemented by Apache Hadoop. MapReduce is a data analysis process that captures data from the database and processes it by executing “Map” and “Reduce” procedures, which break down large job objective into a set of discrete tasks, iteratively on computing nodes. After the data has been analyzed, the results will be stored in a data warehouse and made visually accessible for users to facilitate decision-making and taking appropriate action.

The main difference in analytical capability between big data management systems and traditional management systems is that the former has a unique ability to analyze semi-structured or unstructured data.

Unstructured and semi-structured data in healthcare refer to information that can neither be stored in a traditional relational database nor fit into predefined data models, such as XML-based electronic healthcare records (EHRs), clinical images, medical transcripts, and lab results. Most importantly, the ability to analyze
unstructured data plays a pivotal role in the success of big data in healthcare settings since 80% of health data is unstructured. According to a 2011 investigation by the TDWI research [25], the benefits of analyzing unstructured data capability are illustrated by the successful implementation of targeted marketing, providing revenue-generating insights and building customer segmentation. In the case of health care, Leeds Teaching Hospitals in the UK analyze approximately one million unstructured case files per month, and have identified 30 distinct scenarios where there is room for improvement in either costs or operating procedures by taking advantage of natural language processing (NLP). This unstructured data analytical capability enables Leeds to improve efficiency and control costs through identifying costly healthcare services such as unnecessary extra diagnostic tests and treatments.

6.3. Analytical capability for patterns of care

Analytical capabilities in healthcare can be used to identify patterns of care and discover associations from massive healthcare records, thus providing a broader view for evidence-based clinical practice. Healthcare analytical systems provide solutions that fill a growing need and allow healthcare organizations to parallel process large data volumes, manipulate real-time, or near real time data, and capture all patients’ visual data or medical records. In doing so, this analysis can identify previously unnoticed patterns in patients related to hospital readmissions and support a better balance between capacity and cost. Interestingly, analyzing patient preference patterns also helps hospitals to recognize the utility of participating in future clinical trials and identify new potential markets.

For example, PatientsLikeMe is a patient social networking website that enables patients to share information in terms of individual experience, diet, and medications and connect with other patients who receive the same treatments. Through taking advantage of social media resources such as comments and feedback, hospitals can gain access to more information on current patient treatment conditions (e.g., side effects and hospitalization), and thus detect rapidly increasing interest in specific markets that will enable them to develop high quality health care that meets the needs of their patients.

6.4. Decision support capability

Decision support capability emphasizes the ability to produce reports about daily healthcare services to aid managers’ decisions and actions. In general, this capability yields sharable information and knowledge such as historical reporting, executive summaries, drill-down queries, statistical analyses, and time series comparisons. Such information can be utilized to provide a comprehensive view to support the implementation of evidence-based medicine, to detect advanced warnings for disease surveillance, and to develop personalized patient care. Some information is deployed in real time (e.g., medical devices’ dashboard metrics) while other information (e.g., daily reports) will be presented in summary form.

The reports generated by the analytics engines of big data systems are distinct from transitional IT systems, showing that it is often helpful to assess past and current operation environment across all organizational levels. Big data reports are created with a systemic and comprehensive perspective and the results evaluated in the proper context to enable managers to recognize feasible opportunities for improvement, particularly regarding long-term strategic decisions. For example, Premier Healthcare Alliance collects data from different departmental systems and sends it to a central data warehouse. After near-real-time data processing, the reports generated are then used to help users recognize emerging healthcare issues such as patient safety and appropriate medication use.

6.5. Predictive capability

Predictive capability is the ability to apply diverse methods from statistical analysis, modelling, machine learning, and data mining to both structured and unstructured data to determine future outcomes [16, p. 289]. Wessler [26, p. 21] defines predictive capability as the process of using a set of sophisticated tools to develop models and estimations of what the environment will do in the future. Both these definitions focus on the importance of predicting future trends and insights for organizations and individuals by identifying gaps between current and future states. Predictive analysis makes it possible to cross reference current and historical data to generate context-aware recommendations that enable managers to make predictions about future events and trends. This capability relies on predictive analytical engines that incorporate a data warehouse, a predictive platform with predictive algorithms (e.g., decision trees, neural networks, and logistic regression), and a predictive interface that provides feedback and recommendations to users. Predictive capabilities in healthcare assess current healthcare service situations to help managers disentangle the complex structure of clinical cost, identify best clinical practices, and gain a broad understanding of future healthcare trends based on an
in-depth knowledge of patients’ lifestyles, habits, disease management and surveillance [23].

Predictive capabilities can reduce the degree of uncertainty, enabling managers to make better decisions faster and hence supporting preventive care [17, 27]. The Texas Health Harris Methodist Hospital Alliance, for example, analyzes information from medical sensors to predict patients’ movements and thus provide needed services more efficiently. It also monitors patients’ actions throughout their hospital stay to improve the prevention of medical risk. Healthcare entities with superior big data predictive capabilities should be able to leverage helpful predictive reports to improve decision-making and optimize existing operations and provide high quality healthcare services.

I+Plus, an advanced analytical solution used in an Australian healthcare organization, consists of powerful analytical tools for three levels (claims, aggregated, and admission) of analysis [28]. It provides claim-based intelligence to facilitate customers’ claim governance, balance cost and quality, and evaluate payment models. Specifically, through these analytical patterns managers can review a summary of cost and profit related to each healthcare service, identify any claim anomalies based on comparisons between current and historical indicators, and thus make proactive (not reactive) decisions by utilizing productive models.

7. Formulating big data strategies in healthcare

Healthcare transformation through implementing big data is still in the very early stages, and most of the big data studies focusing on the realm of technology. Attention is sorely needed for research to formulate appropriate big data strategies that will enable healthcare organizations to move forward to leverage big data most efficiently and effectively. We consider that the following five strategies might provide useful directions for those seeking to develop big data implementations in healthcare organizations.

7.1. Implementing (big) data governance

Big data governance is an extension of IT governance that focuses on leveraging enterprise-wide big data resources to create business value. Big data is a double-edged sword for IT investment, potentially incurring huge financial costs for healthcare organizations due to poor governance. On the other hand, with appropriate data governance, big data has the potential to equip organizations with the tools they need to harness the mountains of heterogeneous data, information, and knowledge that they routinely gather, disentangle intricate customer networks and develop new portfolio of business strategies for products and services.

Success big data governance requires a series of organizational changes in business processes. Several issues should be taken into consideration when developing big data governance for a healthcare organization. The first step is to formulate the governance mission, with clearly focused goals, execution procedures, governance metrics, and performance measures. In other words, a strong data governance protocol should be defined that provides clear guidelines for data availability, criticality, authenticity, sharing, and retention that enable companies to harness data effectively from the time it is acquired, stored, analyzed, and finally used. This allows healthcare organizations to ensure the appropriate use of big data and build sustainable competitive advantages. Second, healthcare organizations should review the data they gather within all their units and realize their value. Once the value of these data has been defined, managers can make decisions on which datasets to be incorporated in their big data framework, thereby minimizing cost and complexity. Finally, information integration is the key to success in big data implementation, because the challenges involved in integrating information across systems and data sources within the enterprise remain problematic in many instances. In particular, most healthcare organizations encounter difficulties in integrating the data from legacy systems into big data frameworks. Managers need to develop robust data governance before introducing big data in their organization.

7.2. Developing an information sharing organizational culture

A prerequisite for implementing big data successfully is that healthcare organizations’ employees have an information sharing mindset and culture. This is critical for reducing any resistance to new information management systems from physicians and nurses. Without an information sharing culture, data collection and delivery will be limited, with consequent adverse impacts on the effectiveness of the big data analytical and predictive capabilities. To address this issue, healthcare organizations should engage data providers from the earliest stage of the big data transition process and develop policies that encourage and reward them for collecting data and meeting standards for data delivery. This will significantly improve the quality of data and the accuracy of analysis and prediction.
7.3. Training key personnel to use big data

In addition to contributing to data analysis, decision support is another crucial capability of big data management systems due to its ability to create meaningful predictive reports. The key to use these reports effectively is to equip managers and employees with relevant professional competencies, such as the skills of making an appropriate interpretation of the results and critical thinking. According to American Management Association that 64% of organizations in the United States fail to meet all of their expected analyzing data skills needed in the workplace [29]. In this regard, incorrect interpretation of the reports generated could lead to serious errors of judgment and questionable decisions.

To address this issue, it is important that healthcare organizations provide analytical training courses in areas such as basic statistics, data mining and business intelligence to those employees who will play a critical support role in the new information-rich work environment. Mentoring, cross-functional team-based training and self-study are also beneficial training approaches to help employees develop the big data analytical skills they will need according to a recent survey by the Institute for American Management Association (AMA) and Corporate Productivity (i4cp) [29]. Alternatively, healthcare organizations can adjust their job selection criteria to recruit prospective employees who already have the necessary analytical skills.

7.4. Incorporating cloud computing into the organization’s big data

Most hospitals are small and medium sized enterprises (SMEs), and often struggle with cost and data storage issues. Due to the rapid changes of technology, big data, and the general increase in data-intensive operations, healthcare organizations are facing some challenges: storage, analysis, and bottom line. The needs to store different formats of data and access to them for decision making have pushed healthcare organizations seeking better solutions other than traditional storage servers and processes. A typical model for the storage of big data is clustered network-attached storage (NAS), which is a costly distributed file system for SMEs. A usage-based charging model such as cloud computing services is an attractive alternative. Cloud computing is a network-based infrastructure capable of storing large scale of data in virtualized spaces and performing complex computing near real time. The combination of lower cost and powerful and timely processing and analyzing make cloud computing an ideal option for healthcare SMEs to fully take advantage of big data.

However, storing healthcare data in a public cloud raises two major concerns: security and patient privacy. Although the public cloud is a significant cost savings option, it also presents higher security risk and may lead to the loss of control of patient privacy since the access to data is managed by a third party vendor. A private cloud, on the other hand, provides a more secure environment and keeps the critical data in-house, but increase the budget for big data projects. Healthcare managers must strike a balance between the cost-effectiveness of the different cloud choices and patient information protection when adopting big data.

7.5 Generating new business ideas from big data

New idea generation is not only necessary for organizational innovation, but also can lead to changes in business operations that will increase productivity and build competitive advantages. This could be achieved through the use of powerful big data predictive analytics tools. These tools can provide detailed reporting and identify market trends that allow companies to accelerate new business ideas and generate creative thinking. In addition to using big data to answer known questions, managers should encourage users to leverage big data outputs to discover new ideas and market opportunities, and assess the feasibility of ideas.

8. Concluding comments

In pursuit of better healthcare with lower per capita cost, it is important to search an IT solution to boost healthcare innovation and transformation. Through analyzing big data best practice cases, we sought to better understand how healthcare organizations can leverage big data as a means of transforming IT to gain business value. These cases demonstrate big data infrastructure as an effective IT artifact to potentially create IT capabilities and business benefits. Specifically, we have identified five big data capabilities and formulated big data strategies that will help healthcare organizations to respond strategically to the challenges they face in the highly competitive healthcare market.

9. References


Appendix 1
The list of big data cases in this study:

- Material released by vendors or companies: Wissenschaftliches Institut der AOK (WIdO); Brigham and Women’s Hospital; The Norwegian Knowledge Centre for the Health Services (NOKC); Memorial Healthcare System; University of Ontario Institute of Technology; Centerstone Research Institute; Premier healthcare alliance; Bangkok Hospital; Rizzoli Orthopaedic Institute; Universität des Saarlandes; Fondazione IRCCS Istituto Nazionale dei Tumori (INT); Fraunhofer FOKUS; Leeds Teaching Hospitals; Beth Israel Deaconess Medical Center; Kaiser Permanente.
- Journal databases: Anonymous private health institution in Australia; University Hospitals Case Medical Center.
- Print publications: Texas Health Harris Methodist Hospital; United Healthcare; Mount Sinai Medical Center; Nevada Department of Health and Human Services; Newton Medical Center; Sharp Community Medical Group; Thundermist Health Center; Nice University Hospital; New York State Department of Health.