Managing Big Data for Firm Performance:  
a Configurational Approach
Emergent Research Forum papers

LeeAnn Kung  
Rowan University  
kung@rowan.edu

Hsiang-Jui Kung  
Georgia Southern University  
hjkung@georgiasouthern.edu

Allison Jones-Farmer  
Miami University  
farmerl2@miamioh.edu

YiChuan Wang  
Auburn University  
yzwo037@auburn.edu

Abstract

Big data are challenging organizations to find a thoughtful, holistic approach to data, analysis and information management to facilitate timely and sound decisions making, and in turn to gain competitive advantages. Managing big data is not a simple technical issue, but a complex managerial and strategic one. To achieve the vast potential of big data not only will enterprise IT architectures need to change, firms also need a new strategy, a new mind set, and a capability to deal with unexpected environmental turbulences. In this paper, we present a conceptual model and a novel analysis method, fuzzy set Qualitative Comparative Analysis to model and interpret interdependent non-linear relationships among elements and the outcome, performance. We posit that data management strategy, big data competence, IT capability and organization improvisational capability are interdependent and mutual reinforcing that form a network of nonlinear influential factors for firm decision quality and in turn, performance.

Keywords

Big data competence, Data management strategy, organization improvisational capability, configuration theory, fsQCA (fuzzy set Qualitative Comparative Analysis).

Introduction

Unlike other disruptive forces in the digital business ecosystem, big data are challenging organizations to find a thoughtful, holistic approach to data, analysis and information management to facilitate timely and sound decisions making, and in turn to gain competitive advantages. To achieve the vast potential of big data not only will enterprise IT architectures need to change, but almost every department within a company will also undergo adjustments (Davenport et al., 2010). Managing big data to get value is not a simple technical issue per se, but a complex managerial (Mcafee and Brynjolfsson, 2012) and strategic one. There are multiple factors involved such as organization capability in dealing with unexpected challenges, big data competence and IT to achieve high organization performance. To be able to find optimal combination of these interdependent and mutually reinforced factors would bring nonparallel advantages. This study thus addresses this research question: What are some possible configurations for firm performance combining big data as assets, IT capability, data management strategy and organization improvisational capability (OIC).

Gartner predicts that enterprise data in all forms will grow 650 percent over the next five years. No matter how big the data are, how various or how fast they are being created, they still go through a life cycle, from inception to dissimilation, and organizations still need to manage them using IT to harness the insights. Research has shown that IT has become a part of core business value (Byrd and Davidson, 2003). Big data technologies are seen as one of the major innovations in the last decade. It is virtually impossible to manage big data without IT. It is also imperative to have a (big) data management strategy to guide the IT and processes through the data life cycle.
Information systems (IS) researchers have also acknowledged that IT capabilities as one of the fundamental drivers that promote a firm's business agility and improvisation, leading to distinctive competitive advantages (Lu and Ramamurthy 2011; Pavlou and El Sawy, 2010; Sambamurthy et al., 2003). Considerable research from both strategic and organizational management fields has emphasized the importance of organizational improvisational capability to handle extreme competition, cope with changing circumstances, and pursue potential business opportunities (e.g., Akgun et al., 2007; Barrett 1998; Bergh and Lim, 2008; Hadida and Tarvainen, 2014; Pavlou and El Sawy, 2010; Weick, 1998). Improvisational capability has been regarded as a means of responding to unexpected turbulences and a complementary source of dynamic capabilities in volatile business environment (Moorman and Miner, 1998). Such “spontaneous” capabilities enable organizations to make effective and real-time decisions in response to unprecedented threats under highly turbulent circumstances.

Using big data is a management revolution (Mcafee and Brynjolfsson, 2012). The biggest challenge of big data is not technical, but managerial and cultural (Brynjolfsson et al., 2012; LaValle et al., 2011; Wang et al., 2014). To get the optimal benefits from big data, organizations have to transform into having a big-data-mindset, finding best ways to harvest insights. In other words, organizations need new type of strategy. This is evident in the competition between Blockbuster and Netflix. Blockbuster took advantage of its huge video rental chain for distribution which caused Netflix’s stock price tumble in 2004. Netflix improvised to address this treats by integrating the potential of e-commerce to provide a wide selection of movies based on data-driven recommended options at a lower cost that brick and mortar retailers could not compete with. Few years later, Blockbuster is in tatters while Netflix flourishes approaching 50 million global subscribers with a 32.4% video streaming market share in the US in 2014. Netflix’s improvisational capability and change of strategy and their business model acted as a life saver in a highly turbulent environment. A secret behind the success of Netflix is the adoption of big data analytics (Bharadwaj et al., 2013). More and more organizations like Netflix do recognize there is value in the flood of data, from improved business forecasts to reduced uncertainty in decision-making and improved competitive positioning.

We posit that these aforementioned factors are interdependent and mutually reinforcing hence form a complex nonlinear relationship systems, which enhances an organization’s decision making quality and in turn organization performance. We present our research model (see Figure 1) and the proposed causal factors next.

**Research Model**

![Conceptual Model](image)

**Big Data Competence**

Big-data computing is perhaps the biggest innovation in computing in the last decade (Russom, 2011). According to a report from the U.S. Congress in 2012, big data are “large volumes of high velocity,
complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information". Big data competence such as their functionalities or values have been addressed through various new and old data management technologies in terms of informatics, data analytics and business intelligence (BI). Big data competence is claimed to hold the key to organizational transformation (Wang et al., 2015). However, measurement of big data competence is hard to find from the literature. To fill this gap, we propose to measure big data competence by evaluating the firm’s IT capability of handling the three primary characteristics of big data (volume, velocity, and variety; 3Vs) through data life cycle management phases (DLCM, see Figure 2).

In this study, 3Vs describe that data will be accumulated continuously, in real-time and at a rapid pace as well as in complicated data formats. DLCM is a comprehensive, policy-based approach to manage the flow of an information system’s data throughout their life cycle which involves procedures and practices as well as applications (DLM, 2010). All the data in an organization follow a data life cycle, starting with collection, through repository and process, and ending up with dissemination of data regardless of structure, as shown in Figure 2. Put 3Vs and DLCM together helps us to understand all the phases of big data life cycle, rather than merely focusing on analysis or modeling phases (Jagadish et al. 2014). For example, to make worthy predictions, managers must begin with identifying data that are relevant and valuable and therefore should be included for business analytics. Without capturing the “right” data, big data analytics is meaningless. Thus, we propose to conceptualize and measure big data competence in the level of competence handling the 3Vs of big data in each phase of the data life cycle. As a result, we define big data competence as a firm’s ability to acquire, store, process and analyze large amount of data in various forms, and deliver information to users that allows organizations to extract values from big data in a timely fashion.

**Figure 2: Data Life Cycle Management**

**Organization Improvisational Capability (OIC)**

OIC in this study is defined as “an organization’s learned ability to respond to unexpected environmental turbulences quickly by simultaneously forming and executing novel solutions by reconfiguring available resources” (Kung, 2015). Research from both strategic and organizational management fields has emphasized the importance of organizational improvisation to handle extreme competition, cope with changing circumstances, and pursue potential business opportunities (e.g., Akgun et al., 2007; Barrett 1998; Bergh and Lim, 2008; Hadida and Tarvainen, 2014; Weick, 1998). Organization improvisational capability plays a crucial role in building organizational agility to react to market changes. OIC also emphasizes the reuse of available resources. Combined with big data analysis, OIC’s “spontaneous” capability enables organizations to make effective and real-time decisions in response to turbulences without having to go through formal planning channel.

**IT Capability**

A firm’s IT capability is defined as its ability to mobilize and deploy IT-based resources in combination or copresent with other resources and capabilities (Bharadwaj, 2000). Research has shown that IT has become a part of core business value (Byrd and Davidson, 2003). This research paradigm was grounded
on the simplistic idea that IT provides the tools necessary to transform inputs to outputs effectively (Orlikowski and Iacono, 2001). Arguments have been made that research on IT business value should investigate the effects of IT on business processes (Kim et al., 2011; Ray et al., 2005). Proponents point out that it is the process (e.g., a better way of doing things) rather than the product where IT makes a true impact (McAfee and Brynjolfsson, 2008). The IT capability literature recognizes that competence in mobilizing and deploying IT-based resources is a source of competitive advantage and differentiates firm performance (Bharadwaj, 2000; Piccoli and Ives, 2005; Wang et al., 2013).

**Strategy and IT**

Research has shown that IT has become a part of core business value. Organization management theory has been highlighting the importance of aligning strategy and IT. It is not IT per se that will increase firm performance, it is the management of IT, i.e., IS strategy, that will bring about the competitive advantage. However, this traditional linear relationship is not adequate for new big data era. More recent research has shown that IT and competitive strategy are interdependent and should co-evolve for business to gain competitive advantages as early as possible.

**Data Management Strategy and Big data**

In the ever-competitive world of business, data strategy is becoming the next big competitive advantage. According to Gartner Group, “By tapping a continual stream of information from internal and external sources, businesses today have an endless array of new opportunities for: transforming decision-making; discovering new insights; optimizing the business; and innovating their industries.” Managing big data is not just about holding lots of information; it is how to best use big data as assets in combination with other organization assets and capabilities to get insights. To manage and to get the best out of big data, traditional IT such as relational database is not adequate any more. On the same token, organizations need to reevaluate their data management strategy to better handle big data.

**Decision Making Quality**

For many organizations there is no problem in handling big data through the first three phases of data life cycle, i.e., collection, storage and process. Instead, a key factor that determines whether organizations sink or swim in today’s rapidly evolving business environment is their ability to harness the potential and power of data by gleaning useful insights for decision-making and innovation (LaValle et al., 2011). There is no value created without use. “All the analytics in the world won’t help unless we use them to make and execute better decision” (p.175, Davenport et al., 2010, citing analytic user). One of the values of big data analytics has always been the ability to make better predictions and to enhance managerial decision quality. We, therefore, use the decision making quality as the primary outcome of the combinations of elements, and then link it to the organization performance.

**Holistic Proposition.** Different combinations of interconnected big data competence, OIC, IT and data management strategy can help organizations achieve high performance through increased decision quality. In other words, under different circumstances, organizations use different combinations of these elements to gain high performance.

**Research Methods**

**Data Collection Approach**

We will use a web survey which includes quantitative and qualitative questions to collect data from various industries. We will collect data from executives (CIOs and CEOs) and IT managers whose roles should make them knowledgeable of their firms’ strategy, IT, and performance. We then will match data collected with a secondary data source for firm size, IT department size and revenue. One advantage of applying configuration method is that only a small number of cases is needed for analysis.
Measurement of Constructs

Efforts were made to find empirically tested and validated scales for the constructs (elements). We were able to find measurements for IS quality, data management strategy, OIC and decision making efficiency. We developed big data competence measurement for this study. Big data competence is measured by evaluating the level of competence of managing the 3Vs of big data in each process of the data life cycle. For example, to assess a firm’s ability of collecting (big) data, questions about the ability to collect big volume of data in various formats at a near real-time high speed were developed. Organization Improvisational capability scale was developed and tested by one of the authors (Kung, 2015).

Data Analysis Methods: fuzzy-set Qualitative Comparative Analysis (fsQCA)

Since we conceptualized interdependent non-linear relationships among causal factors and outcome, we use configurational method as our inquiring system instead of traditional linear path model. A configurational approach assumes complex causality and nonlinear relationships (Meyer et al., 1993). Fundamental premise of configurational approaches is that “patterns of attributes will exhibit different features and lead to different outcomes depending on how they are arranged” (Fiss, 2007 p.1181). Configurational analysis stresses the concept of equifinality, which refers to “a system can reach the same final state, from different initial conditions and by a variety of different paths” (Katz and Kahn, 1978, p.30). Comparing to more dominant quantitative methods, configurational approach provides the opportunities to model the complex and nonlinear relationships among casual factors and the outcome.

From its inception, qualitative comparative analysis (QCA) was aimed at the “middle ground” between quantitative and qualitative methodologies (Ragin, 2000, p.22). Fuzzy sets complement QCA as a methodological tool to translate categorical concepts into measurable conditions, drawing on the notion that cases can hold degrees of membership in a given set (Ragin and Fiss, 2008). The building block of fuzzy-set QCA is “fuzzy” membership of cases in a set of cases with a given characteristic. The typical results of fsQCA are multiple configurations of the factors that produce the same level of the outcome, in this case, firm performance.

Sample Results

Figure 3 shows an example of fsQCA results, configurations for high market value growth. The configurations are expressed by the notation systems from Ragin and Fiss (2008). The dark shaded circles indicate the presence of an element, crossed-out circles indicate the absence of an element, large circles indicate core elements, and small circles indicate peripheral elements. Blank space indicates a “don’t care” situation, which means that the causal element may be either present or absent. Consistency measures the degree to which a relation of necessity or sufficiency between a causal condition (or combination of conditions) and an outcome is met within a given data set (Ragin et a., 2006). It resembles the notion of significance in statistical models (Thiem, 2010). Coverage provides a measure of empirical relevance (Legewie, 2013). The analogous measure for coverage in statistical models would be $R^2$.

In this example solution 1, turbulence is a “don’t care” condition, which means as long as an organization has the combination of very high level of outside-in IT capability, high level of spanning IT capability, somewhat-flat structure or entrepreneurial culture, adequate munificence and no complexity in the environment, it can achieve high market value growth regardless the level of turbulence. So, even if an organization is in a highly turbulent environment, it can achieve high market value growth if it follows solution 1.
Combine Big data and other capabilities for performance

<table>
<thead>
<tr>
<th>High Market Value Growth</th>
<th>Solution/Recipe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Outside-in IT Capability</td>
<td>●</td>
</tr>
<tr>
<td>Spanning IT Capability</td>
<td>●</td>
</tr>
<tr>
<td>Inside-out IT Capability</td>
<td></td>
</tr>
<tr>
<td>Organization Characteristics</td>
<td>●</td>
</tr>
<tr>
<td>Turbulence</td>
<td></td>
</tr>
<tr>
<td>Munificence</td>
<td>●</td>
</tr>
<tr>
<td>Complexity</td>
<td>❌</td>
</tr>
<tr>
<td>Raw Coverage</td>
<td>0.38</td>
</tr>
<tr>
<td>Unique Coverage</td>
<td>0.26</td>
</tr>
<tr>
<td>Consistency</td>
<td>0.86</td>
</tr>
<tr>
<td>Overall solution coverage</td>
<td>0.42</td>
</tr>
<tr>
<td>Overall solution consistency</td>
<td>0.85</td>
</tr>
</tbody>
</table>

**Figure 3. Configurations for achieving high market value growth**

**Conclusion**

This research intends to examine possible configurations of firm performance using the fsQCA method. We argue that big data’s impact on firm performance depends on the entire configuration of big data competence, IT capability, organization improvisational capability and firm’s data management strategy. We propose the concept and measurement of big data competence integrating big data’s 3Vs characteristics and data life cycle concept. Based on empirical field data, this research intends to identify multiple configurations that produce equally high organizational performance. While this study does not show empirical result, it still contributes to the big data studies, specifically treating big data as assets working side-by-side with other organizational capabilities for high quality decisions and firm performance. Following researchers such as Ragin and Fiss (2008), Park and El Sawy (2013), we advocate fsQCA as a new way to investigate multiple configurations and to show how it is best suited for explaining the complex dynamics between tightly interconnected elements.

**REFERENCES**


*Twenty-first Americas Conference on Information Systems, Puerto Rico, 2015*


Data life cycle management (DLM) Published: 02 Aug 2010; http://searchstorage.techtarget.com/definition/data-life-cycle-management; Posted by: Margaret Rouse; accessed 08/25/13)


Leadership Council for Information Advantage, Big Data: Big Opportunities to Create Business Value, report.

Legewie, N. 2013. “An Introduction to Applied Data Analysis with Qualitative Comparative Analysis”. In Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 14, paper 3.


**Appendix 1: Measurement of Big Data Competence**

**Big Data Competence**
BDC1: obtain
1. Able to receive vast amount of data (greater than petabyte)
2. Able to collect different formats of data
3. Able to access data from different sources
4. Able to collect data real-time (as they are generated)

BDC2: storage
1. Able to store amount of data (greater than petabyte) either internally or externally
2. Able to store different formats of data either internally or externally
3. Storage is scalable
4. Has high-capacity data warehouse

BDC3: process/analyze
1. Able to process data collected quickly
2. Able to process data stored quickly
3. Able to process different formats of data (includes structured, semi-structured and unstructured data)
4. Able to filter data
5. Able to determine data relevancy based on enterprise context.(Data quality)
6. Able to analyze collected data quickly
7. Able to analyze different formats of data quickly
8. Complex event processing ability
9. Able to report in a meaningful way
10. Able to query data
11. Offer forecast information to user
12. Able to find patterns
13. Offer based-on-data predictions

BDC4: delivery
1. Data are shared throughout the entire organization in near-real time
2. Able to distribute information quickly
3. Able to deliver information to every organization users in a short time
4. Able to visually present information according to user’s requirements/needs