Value Creation from Big Data: Looking Inside the Black Box

Abstract

The advent of big data is fundamentally changing the business landscape. We open the "black box" of the firm to explore how firms transform big data in order to create value and why firms differ in their abilities to create value from big data. Grounded in detailed evidence from China, the world's largest digital market, where many firms actively engage in value creation activities from big data, we identify several novel features. We find that it is not the data itself, or individual data scientists, that generate value creation opportunities. Rather, value creation occurs through the process of data management, where managers are able to democratize, contextualize, experiment, and execute data insights in a timely manner. We add richness to current theory by developing a conceptual framework of value creation from big data. We also identify avenues for future research and implications for practicing managers.

Key words: Big data, Resource management, Knowledge-based view, Open innovation, China

INTRODUCTION

Increasingly, companies are exploring how big data can be used to create and capture value (*McKinsey Global Institute*, 2011). 'Big data is about massive amounts of observational data, of different types, supporting different types of decisions and decision time frames (Goes, 2014: vii). The use of big data enables managers both to know fundamentally more about their businesses, and to translate that knowledge into better decision-making and improved performance (McAfee and Brynjolfsson, 2012). For example, Netflix is able to collect data by monitoring when customers pause, rewind or fast forward, from their browsing and scrolling behaviour, and identify customer preferences based on the ratings given to the content. By analysing these data, Netflix is able to predict the popularity of its content and purposefully design content based on the data insights generated from customers.

Like analytics that preceded it, big data analysis seeks to gather intelligence from data and translate that into business advantage, however, key characteristics mean that big data is more powerful than the analytics that were formerly used (McAfee and Brynjolfsson, 2012). These characteristics relate to the volume of data created; its velocity, i.e., the speed of data creation makes available real-time or close to real-time information; and its variety, big data is generated from a plurality of sources (George et al., 2014). McAfee and Brynjolfsson (2012) point out that while analytics brought rigorous techniques to decision making, big data is both simpler and more powerful. Added to this the steadily declining costs of computing, including storage, memory, processing and bandwidth, mean that previously expensive data-intensive approaches are becoming economical.

In seeking a technology-based competitive advantage, central to growth and success, exploitation of big data enables companies to create and capture value. Big data and increased computational power enable companies to be smarter and innovative, in ways they could not before (LaValle et al., 2011). The research findings of McAfee and Brynjolfsson (2012) show that the more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results. Similarly, the study by LaVale et al., (2011) clearly connects performance and the competitive value of analytics.

Although organizations are exploring ways of deploying and harnessing big data to create and capture value, many are struggling with how to create value from the significant amount of data they already have owing to its volume, velocity and variety (Goes, 2014; Lavalle et al. 2011). Indeed, merely possessing a valuable resource does not guarantee the development of value creation (Barney and Arikan, 2001; Priem and Butler, 2001a, 2001b; Sirmon et al. 2007). Lepark et al. (2007) define value creation as the relative amount of value that is subjectively realized by a target user and this subjective value realization must at least translate into the user's willingness to exchange a monetary amount for the value received. In order to realize value creation, firms must effectively accumulate, combine and exploit resources (Grant, 1991; Sirmon et al. 2007). However, the understanding of how managers manipulate resources in order to create value is largely underdeveloped (Barney and Arikan, 2001; Kraaijenbrink et al., 2010; Priem and Butler, 2001a; Sirmon et al. 2007). More recently, scholars and practitioners have noted the necessity for empirical investigation of the process of transforming big data to create value (Bharadwaj et al. 2013; Georg, et al, 2014; Goes, 2014; Munford, 2014). In response to this, our aim is to examine how firms transform big data to create value, and further to investigate why firms differ in their abilities to create value from big data.

We seek to make a contribution by exploring the value creation process in a Chinese context for three reasons. First, China is the world's largest digital market. Due to the scale of the country's economy and its engagement in the digital market, understanding how the leading companies in China transform data to create value is a timely undertaking. Second, the combinations of cultural heritage, rapidly developed infrastructure and digital market have created a distinct business phenomena embedded in a special context (Whetten, 2009). Research that focuses on the Chinese context therefore serves to make an important contribution to knowledge (Tsui, 2007; Whetten, 2009). Third, many Chinese companies were able to keep foreign competitors, such as Google, eBay and Amazon, at bay (Zeng and Glaister, 2016). Therefore, understanding the process of how firms in China manage big data to create value should stimulate novel insights into "local" organizational practices and foster improvements in borrowed theory (Tsui, 2007; Whetten, 2009).

Our study makes an important contribution by answering the call to investigate how managers create value from big data (Bharadwaj et al., 2013; George et al., 2014; Goes, 2014; Munford, 2014). Drawing on qualitative data, we observe that firms differ in their abilities to extract value from big data, both internally within the firm and externally across the extended data-sharing network. We highlight the process of data management where managers are able to democratize, contextualize, experiment, and execute data insights in a timely manner to create value. Further, we move beyond the value creation process at the firm level of analysis to the network-level of analysis. By sharing data with its partners, the firm is able to build a leadership position depending on its capability to establish, maintain and reinforce symbiotic relationships with its partners. We also contribute to the organization structure literature. We suggest that firms that combine improvisation with low-to-moderately structured rules to execute a variety of opportunities that emerge from big data are more flexible and able to respond to market shifts.

We structure the paper as follows. In the next section, we provide a review of the relevant literature. We then set out the study's methods, including the sampling strategy, the interview protocol, and the data analysis. We then discuss the findings, offer an overarching framework and develop theoretical implications. Finally, we suggest avenues for future research.

THEORETICAL BACKGROUND

Our focus here is on the value creation process from big data at the firm level, a topic that has so far received very limited attention from management scholars (George et al., 2014; Rai, 2016). We review four streams of literature that are related to the dynamic context in which our studies were carried out, namely, resource management, knowledge-based view, organization structure and network collaboration that highlights extracting value from big data. We review this literature by probing its applicability to the specific context of creating value from big data.

Resource management

The resource-based view (RBV) of strategic management postulates that the competitive advantage of a firm lies primarily in its application of the bundle of tangible and intangible assets, resources and capabilities that are rare, valuable, non-substitutable and difficult to imitate (Barney, 1991; Rumelt, 1984), which firms need to manage, adopt, and deploy in product markets in order to create value (Mahoney, 1995; Sirmon et al., 2007). Such superior resources result in greater value creation i.e. an advantage in competition (Barney, 1991; Hoopes et al., 2003). Simply owning resources does not guarantee the development of value creation (Barney and Arikan, 2001; Priem and Butler, 2001a, 2001b; Sirmon et al., 2007). Resource management is critical to value creation because using resources is at least as important as possessing or owning them (Hansen et al. 2004; Penrose, 1959). Different resource management processes can yield heterogeneous outcomes for organizations holding similar resources and facing similar environmental contingences (Zott, 2003; Lippman and Rumelt, 2003).

A small but growing number of researchers have attempted to elucidate the role of managers in managing resource-related processes within RBV and dynamic capability logic. For example, Sirmon et al. (2007) defined resource management as the comprehensive process of *structuring*, *bundling* and *leveraging* the firm's resources with the purpose of creating value for customers and competitive advantages for the firm. In parallel with research on the development of resource management, Helfat et al. (2007) developed a related logic that focused on asset orchestration emerging from the dynamic capability literature. Dynamic capabilities are an extension of RBV, and highlights explicitly the role of managers when they "purposefully create, extend or modify [the firm's] resource base" to create value to achieve sustainable advantages (Amit and Schoemaker, 1993; Eisenhardt and Martin, 2000; Teece et al. 1997; Winter, 2003).

Although a growing number of researchers attempted to look inside the "black box", additional theory development is required to add richness to our current understanding of how to manage resources to create value in dynamic environments (Hefat et al., 2007; Sirmon et al., 2011). This is particularly the case for big data. What is particularly interesting about data is that it often increases in value the more it is used. Moreover, it is self-regenerative, in that the identification of a new piece of information immediately creates both the demand for, and conditions for production of, subsequent pieces (Huber, 1984; Porat, 1976; Stiglitz, 1975). The unique characteristics of big data differ significantly from those of traditional physical assets that are generally appropriable ("either I have it or you have it"), scarce, decrease return to use and are non-renewable (Porat, 1976; Stiglitz, 1975; Glazer, 1991). Kraaijenbrink et al. (2010) proposed that differentiating between resources and emphasizing their different characteristics could prove fruitful to offer a more robust explanation of RBV. Existing understanding of resource management literature has not distinguished the differences between rivalrous and non-rivalrous resources, nor is it safe to assume that existing theory could be extended to account for the value creation from big data.

Knowledge -based view

The knowledge-based view (KBV) in strategy has largely extended resource-based reasoning by suggesting that knowledge is the primary resource underlying new value creation, heterogeneity, and competitive advantage (Barney, 1991; Grant, 1996; Kogut & Zander, 1996). Proponents of this view perceive knowledge as the principal strategic resource, and argue that firms supersede markets in their ability to create and harness this resource (Eisenhardt and Santos, 2001). The KBV proposes that the primary reason for the existence of the firm is its superior ability to integrate multiple knowledge streams, for the application of existing knowledge to tasks, as well as for the creation of new knowledge (Grant, 1996). Grant (1994: 375) further states that “At the heart of this theory is the idea that the primary role of the firm, and the essence of organizational capability, is the integration of knowledge”.

The knowledge based view highlights the role of the individual as the prime driver in the creation of organizational knowledge (Nonaka, 1994), and conceptualizes the existence of a firm as an institution that integrates knowledge that resides within and across individuals (Grant, 1996). While some scholars have argued the primacy of the individual (Grant, 1996; Simon, 1991), most have focused on a collective locus of knowledge (Eisenhardt and Martin, 2000; Winter, 2003; Zollo & Winter, 2002). The role of data scientists is often accentuated as the key source of knowledge that has direct impact in contributing to the firm’s value creation opportunities from big data (Davenport and Patil, 2012). A data scientist is a specialized individual who possesses tacit knowledge to understand and analyse big data (Davenport and Patil, 2012). An interesting question follows: Is value creation from big data about the attributes and abilities of the individual data scientist involved, or is value creation from big data about a collective process that is independent of individuals such as data scientists? Further, as knowledge does not always flow easily within the organization, another question arises as to what the process is of transferring individual knowledge to the whole organization in the context of big data? Research has yet to shed light on such questions. Felin and Hesterly (2007: 213) pointed to the need for empirical investigation of the individual drivers of knowledge-based value: “Thus opening up the proverbial black box of the firm by explicating the underlying a priori capabilities and knowledge of the individuals involved provides a natural starting point and micro foundation for explaining the creation of new value”.

Organization structure

The postulate of a trade-off between efficiency and flexibility has gained increasing attention from management and strategy scholars (Rindova and Kotha, 2001; Rothaermel et al., 2006; Davis, Eisenhardt and Bingham, 2009). Organizations with too little structure have limited guidance to generate appropriate behaviors efficiently (Weick, 1993; Okhuysen and Eisenhardt, 2002; Baker and Nelson, 2005), while organizations with too much structure are over-constrained and lack flexibility (Siggelkow, 2001; Martin and Eisenhardt, 2010). This dilemma has created a “paradox of administration” (Thompson, 1967:15) because organizations operating in a dynamic business environment require both efficiency and flexibility. In order to manage this dilemma, scholars have pointed out that a moderate structure balance between these two extreme states will lead to high performance (Bingham et al., 2007; Weick, 1976). Research shows that high performing organizations resolve this tension by using a moderate amount of structure to improvise a variety of high-performing solutions (Brown and Eisenhardt, 1997, 1998). For example, Brown and Eisenhardt (1997) found that high-tech firms with a moderate number of simple rules (i.e., semi-structure) are more flexible and efficient (quickly creating high quality, innovative products while responding to market shifts) than firms with more or fewer rules.

This stream of enquiry is particularly relevant in the context of big data. The magnitude of data volume and the speed at which it should be analysed and acted upon requires organization to be flexible and agile to cope with a flow of opportunities that emerge from the data insights. Extant literature has been built around the central argument of how an organization balances flexibility and efficiency, to the neglect of key factors that are relevant to big data. For example, with limited managerial attention, how can a firm manage flexibility and efficiency when it seeks to capture heterogeneous opportunities in the big data context that, as noted, is high volume, high velocity and high variety? With a group of specialists, such as data scientists, who are able to generate data insights, how should the firm be structured in order to transmit knowledge across its boundaries in order to maximize the value from big data? A superior knowledge base can be associated with greater flexibility and faster reaction to environment changes (Grant, 1996), how then should a firm be structured to be able to execute the data insights in a timely manner in order to appropriate value? An understanding of how organizational structure shapes flexibility and enables fast reaction in a big data context is crucial to providing additional insights to the existing literature.

Network collaboration

Departing from the traditional RBV, this stream of research contains efforts to differentiate firms in terms of their ability to build unique resources from external alliances (Dyer and Singh, 1998; Lavie, 2006). It complements the RBV by arguing that critical resources may span firm boundaries and therefore a firm needs to build relational rent from a network of alliance relationships in order to gain “complementarity”, “co-specialization”, and “synergy” resources that ultimately leads to superior firm performance (Lavie, 2006; Lippman and Rumelt, 2003; Mahoney and Pandian, 1992; Peteraf, 1993). By building such alliances, firms that are part of an inter-connected network can gain access to resources without paying their full acquisition costs (Dyer and Singh, 1998; Lavie, 2006). Such resource layering through building relation-specific assets with a firm's close alliance partners can be safeguarded by a contractual or long-term relationship agreement (Dyer and Singh, 1998). Open networks are characterised by alliances that are multi-firm based, with partners operating on a "modular" basis, moreover, they are dynamic as the scope of the sharing network evolves over time as new members join (Bharadwaj, et al., 2013; Han et al., 2012). This contrasts with firm collaboration in a closed alliance context which accentuates a centralized approach and a firm's bargaining power to ensure its ability to generate superior economic rent (e.g., Gulati and Gargiulo, 1999; Lavie, 2006), a model which no longer appears sufficient to provide a robust explanation for superior firm performance in the new information age.

Inspired by Chesborough (2003), open innovation highlights that as the boundary between a firm and its environment becomes more permeable, external ideas are important information resources for the firm's competitive advantage. Many scholars proposed an Open Collaborative Ecosystem (OCEs) (Baldwin and Von Hippel, 2011; Curley and Formica, 2013) and Open Strategy (Tavakoli et al., 2015) that underscores integrated collaboration, with co-created shared value and innovative network. This was manifested by companies such as Facebook, Alibaba and Apple, that started to investigate the significant potential beyond conventional boundaries in order to tap into unique ideas from various constituencies within the innovative ecosystem through open platform strategy. Such an open process enriches a company's own knowledge base through the synergetic effect from external resources, which leads to a great increase of a company's innovativeness (Laursen and Salter, 2006). As a result, firms are able to accelerate the rate of innovation and consequently create a more compelling competitive position (Chesborough and Appleyard, 2007). Although research in this domain has focused strongly on concepts such as open innovation and open ecosystem, the empirical understanding of how companies transform such external resources to create value arguably remains in a black box (Bharadwaj, et al., 2013; Goes, 2014; Munford, 2014; Sirmon, Hitt and Ireland, 2007).

Taken together, these brief summaries of the dominant themes in theory and research clearly demonstrate their value in understanding the process of value creation. Studies from information technology (IT) research streams also illuminate the magnitude and impact of data-related analytical problems in business organizations. A growing number of IT researchers have articulated the significant impact of big data and how big data is going to transform the business landscape (Goes, 2014; Rai, 2016; Chen, Chiang and Storey, 2012). However, despite the great effort devoted to emphasise the significance of big data, many firms are still struggling to capture value opportunities associated with big data. Recently, scholars have attempted to bridge the disciplines between information system strategy and strategy-as-practice and called for a joint agenda to take full advantage of such theoretical synergies (Lazer, Pentland and Adamic, 2009; Giles, 2012; Whittington, 2014). As yet, there does not appear to be a holistic explanation of "how" firms transform big data to create value, as well as why variations exist across firms in relation to their capabilities to manage big data. This understanding is important, as recent researchers and practitioners suggest that data is playing an increasingly significant role in driving the sustainable value creation of the firm (George et al. 2014; Koutroumpis and Leiponen, 2013; McAfee and Brynjolfsson, 2012; Mckinsey Global Institute, 2011).

It follows from these observations that the best way to understand the process of managing big data to create value is to undertake an inductive analysis that, while building on existing literature, allows new analytical insights on how big data drives the value creation of the firm. The purpose of this study is twofold. First, we look inside the "black box" and explore how firms transform big data to create value. Second, as some firms are more able to create value from big data than others, we investigate why firms differ in their ability to create value from big data. In the following section, we elucidate our study's sample, data collection, interview protocol, and data analyses.

RESEARCH METHOD

We adopt a "naturalistic inquiry", which applies inductive logic to obtain insight (Lincoln and Guba, 1985). Inductive studies are particularly helpful in developing and extending new theoretical insights and when the research question is process oriented, such as ours. Multiple case studies were adopted in order to create the opportunity to triangulate information collected and to augment external validity, help guard against observer bias and allow for replication logic (Eisenhardt, 1989; Miles and Huberman, 1994; Yin, 2003). This approach allows us to extend existing theory and accordingly develop theoretical explanations for the observed phenomena (Locke, 2001).

Research setting

Internet platform companies (IPCs) are the setting of our research, and one where the core product is data. First, IPCs such as, Alibaba, are the epitome of data-based organizations because they are established on the internet and have been collecting data since their inception (Doan et al. 2011). Compared to traditional companies that largely depend on physical assets to drive internal and supply-side efficiency, IPCs primarily depend on their ability to generate information to enable/facilitate the interaction between different groups of users in order to create value (Parker and Van Alstyne, 2005). Thus, the value of data is particularly salient in the context of IPCs. Second, rather than protecting data that has been accumulated and collected by the firm, many IPCs open up their platforms and share their data with third parties. This phenomenon provides an interesting lens to understand how firms transform such open data to create value. The owners of open platforms normally expose their application programming interfaces (APIs) to allow third parties free entry into the supply of the platform, including its technology, database and interaction with the focal platform's users (West, 2003). This differs fundamentally from open source software where individual programmers and users develop a range of software applications and distribute them to the public for free.

Following a purposeful sampling method (Eisenhardt, 1989; Miles and Huberman, 1994), we selected five firms in the IPC industry ranging from an e-commerce platform, a web search engine to a web instant messaging provider. All firms are publicly held. As shown in Table 1, five companies, that were previously successful with a proprietary platform strategy and have adopted 'open platform strategies' (West, 2003) in the last eight years, were eventually selected for this study. Among these five firms, we selected three firms that had been successful in creating value from big data and two that had not. We defined successful value creation as our informants did, in terms of positive characteristics (e.g., 'made a huge difference to our job') and negative ones (e.g., 'too much hype, don't see the point'). The firms were selected as the unit of analysis.

[Insert Table 1 about here]

Data collection

We obtained information from four sources: 1) semi-structured interviews with 42 informants from the sample firms, including CEOs, directors and senior managers who were usually one level subordinate to the CEO, and with senior data engineers in significant roles in their companies; a total of 34 additional interviews with the IPCs' partner firms and third party developers and repeated semi-structured interviews with the same participants through on-site visits and conversations via telephone, Skype and WeChat; 2) reports and strategic memos produced by the companies for the period between February 2008 and March 2013; 3) extensive archives including newspapers, internet sources and corporate materials published between March 2000 and July 2014; 4) informal follow-ups with emails and phone calls for clarification. This diverse range of sources was designed to improve the likelihood of gaining a complete and accurate picture (Yin, 2003) as well as providing textual accounts of debates and discussions and strengthening confidence in the findings (Jick, 1979).

The semi-structured interviews were conducted in Chinese, ranged from 60 to 150 minutes, recorded (unless disallowed by the interviewees) and transcribed verbatim within a week of the interviews by the first author and professional transcribing and translating service provider. The interviews were structured into four sections. First, we asked questions about the background and business strategy of an informant's firm. Second, using an open-ended format we asked the informant to describe the process by which firms transform big data to create value and key challenges they have encountered during such value creation process. We then asked informants to describe the role of data experts such as data scientists in contributing to the value creation process. This was followed by questions about how managers' action/abilities influence the value creation process. More details of the interview questions can be found in the Appendix.

We adopted a "court room" style of interviewing, pushing for concrete illustrations to increase the data trustworthiness (Eisenhardt and Graebner, 2007). Additional interview questions were added to the interview protocol in order to probe emergent themes or to take advantage of special opportunities that may have arisen in a given situation (Eisenhardt, 1989). We assured anonymity to informants to encourage candour. The steps taken are likely to have mitigated informant and other biases and provided detailed and accurate accounts.

Data analysis

We began by conducting within-case descriptions, where the case studies were built based on data and key constructs (Eisenhardt, 1989; Yin, 2003). Within case evidence was acquired by taking notes and writing narratives. For this purpose, we focused on analyzing the interview data as well as integrating and triangulating facts from various data sources. Triangulation of archival and interview data enables richer and more reliable description of each case (Jick, 1979) and improves construct validity (Yin, 2003). This process led to the discovery of different modes of value creation as well as the mechanisms that enable/inhibit such value creation processes.

Next, cross case analysis was conducted, looking for similar constructs and themes in the cases (Eisenhardt and Graebner, 2007). To preserve the integrity of replication logic across cases (Eisenhardt, 1989; Yin, 2003), we began this cross-case analysis after most data had been collected. As research progressed, we sought to verify the emerging dynamics by using supplemental data sources, especially non-interview data (Jick, 1979). From the emerging constructs and themes, we identified the tentative relationships between constructs and attempted to discern correlational tendencies between the value creation modes and the mechanism of enablers and inhibiters. We then used replication logic to further refine these initial relationships by visiting frequently each case in order to compare and contrast the specific constructs, relationships and logics. To ensure the validity of the analysis and theory building, we presented our analysis in an academic workshop in order to stimulate and induce alternative explanations. The latter were accepted or rejected based on their consistency with the data and/or theoretical logic. We continued reading broadly in an effort to gain insight into the data (Glaser and Strauss, 1967). As the theoretical frame clarified, we compared it with the extant literature to highlight similarities and differences, strengthen the internal validity of findings, sharpen construct definitions and raise the generalizability of the emergent theory. We also presented the inductive model to informants inviting their feedback and comments. These interactions were conducted through face-to-face meetings, telephone discussion, and email dialogue.

RESULTS: AN INDUCTIVE THEORY OF VALUE CREATION FROM BIG DATA

Through iteration of the data described in the previous section, a theoretical framework emerged. Our inductively derived theoretical framework of value creation from big data is shown in Figure 1.

[Insert Figure 1 here]

The framework shown in Figure 1 helps to explain how firms manage big data to create value and why firms differ in their abilities to perform better than others. To trace the chain of causation, we have identified two modes of value creation processes from big data. Creating value from internal data is transaction-driven, where the firm mainly focuses on data analysis to generate superior economic rent that can be exclusively enjoyed by the firm. Creating value from the firm's open data network is relation-driven, where the firm mainly focuses on data collaboration to generate superior economic rent that can be collectively enjoyed by the firm and its partners. Looking inside the "black box", we found that data is location, context and relevance sensitive; when it is combined with different ideas and datasets, it enables firms to see new, better and different patterns that are impossible to see in isolation.

The findings revealed that some firms are better than others in transforming big data to create value. In as much as there are gains from an internal data set, that is, when firms are able to *democratize, contextualize*, *experiment* with data and *execute* data insights in a timely manner to create value, these firms generally perform better than others in creating value from the firm's internal data network. Gains from data collaboration with external partners tend to reinforce gains from creating value from an internal dataset, which generate greater value creation opportunities when coupled with additional capabilities and datasets. In order to become the leader in an inter-connected data network, a firm's central position is largely influenced by the potential of both its data relevancy to contribute to partners' value creation processes, and of its technology relevancy to provide a data infrastructure that can facilitate data flows and data exchange within the firm's extended data network. Therefore, the differences in gains from data mining and gains from data collaboration were caused by heterogeneity between firms' capabilities to harness data internally and heterogeneity between firms' capabilities to manage relationships externally.

In the following section, we discuss each value creation mode in further detail. We then shift the focus of the analysis from the question of how firms transform big data in order to create value to examine why firms differ in their abilities to create value from big data.

FINDINGS

Two distinct patterns of value creation modes from big data emerged from the empirical evidence: creating value from the firm's internal data set and creating value from the firm's external data network (see Figure 1). Our analysis indicates that the modes of value creation from big data exhibit common features (the firm's capability to harness the data internally and its capability to manage relationships with external partners in an open data network) with variations in explaining how some firms are able to generate value better than others from the same value creation process. In this section, we first provide a brief picture of each mode, and then present a second layer of findings explaining why some firms differ in their capabilities to create value from big data better than others. We intersperse the narratives with significant quotes intended to illustrate our interpretation, as recommended by Langley and Abdallah (2011), and we display additional selected quotes in Table 2, to illustrate and document the robustness of our claims.

[Insert Table 2 about here]

Creating value from the firm's internal dataset: capability to harness data internally

This pattern of value creation refers to the process of exploiting and exploring a large volume of data from the firm's internal database in order to identify consistent patterns and systematic correlations among different variables in a particular representational form, and transforming these insights into actions that leads to identification of new opportunities for creating value that increase the customer's willingness to use/pay. The selected IPCs have accumulated a large volume of data over time which allows sophisticated analysis and pattern matching. There is a recurring theme among all companies that exploiting data enables a firm to conduct "demand driven" product/service development. A senior director from Alpha commented:

Every time customers use our platform, they left a trail of bread crumbs about who they are, what they like, etc. There are many types of data; administrative data, transactional data, behaviour data and real time data, etc. For example, we rely on customer behaviour data to understand how they react to where we place the click button, the size and font of words displayed on the website, the colour, shape. (Alpha, 04a)

In a similar vein, Tangu divided user activity data from one particular product into three categories: early product utilization stage, peak product utilization stage and end of product utilization stage. This chain of evidence enables Tangu to gain a holistic understanding about the entire user experience journey, which provided valuable information guiding future product improvement and innovation.

We also noted that firms use big data to provide value added and personalized service to different groups of consumers. Alpha introduced a series of services, such as, online store design, account management, and transaction data analysis (number of daily orders, the variation of daily online traffic, the metrics of visitors, sales, etc.) to assist the sellers on their platform to improve transaction efficiency. Based on big data, Alpha was able to build its own credit scoring model that provided comprehensive analysis about client behavior, transaction history, credit rating and review content. This knowledge led to a range of financial services including small financial loans to its customers. Similar examples can be found among other IPCs. For example, Bray developed a set of big data predictive programs for monitoring diseases like hepatitis, Chinese New Year travel, earthquakes, FIFA World Cup victories and movie box office success. One of the senior data engineers from Bray noted that:

Now we have the luxury to have the real-time data which can really make a difference to our lives. We can use such data to have a fairly accurate understanding about, for example, how crowded it is now at Forbidden City. Data is time sensitive too. (Bray, 009a)

As the arbiters of value, the customer plays the key role in determining the value of the product/service offered by the firm (Priem, 2007). The data generated from the demand-side can guide and help a firm to understand the consumer's needs, which leads to the identification of new opportunities for creating value that increase the customer's willingness to use/pay. Firms equipped with the insights generated from big data are more able to respond to unexpected opportunities. The outcome of data exploiting can enhance a firm's isolating mechanism to allow it to decrease threats of imitation, thus increasing the maintainability of an advantage based on the insights generated from big data. This discussion leads to the following proposition:

Proposition 1: In a big data context, exploiting data facilitates preferential access to a greater variety of demand-side opportunities, thus increasing the firm's potential value creation.

How do some firms create more value from internal data than others? Prior research placed emphasis on the role of "data scientists" who have the specialized skills and knowledge to analyse big data (e.g., Davenport and Patil, 2012) in contributing to the firm's value creation opportunities from big data. This perspective implies that companies that have more data scientists have better chances of creating value from big data.

The findings from our research indicate a different view. Firms that hired many data scientists do not always generate better value creation opportunities. Rather, it was the process of data management where managers are able to *democratize*, *contextualize*, *experiment,* and *execute* data insights in a timely manner to create value. In the following sub- sections, we provide a detailed account of our observations to support our emergent theme (Eisenhardt and Graebner, 2007). We display additional selected quotes in Table 3 to illustrate and document the robustness of our claims.

[Insert Table 3 about here]

Capability to democratize data

We define data democratization as the firm's capability to integrate data across the firm and enable a wider range of employees to access and understand data where it is needed at any given time. One challenge dealing with big data, often highlighted during our interviews, is the sheer volume of data itself, as described by many informants as "*we can easily drown in the big data sea*". We found significant variations in the scale where internal employees can benefit from exacting insights from big data. The quotes shown in Table 3 indicate that data democratization is associated with better value creation.

The Alpha case illustrates the linkage between data democratization and value creation. Alpha data director argued "*make big data serve everyone (in the company), not just top management and executives*". In 2009, Alpha introduced data tools to enable a wider scale of data application. For example, the data team introduced a data product called "Wu Liang Shen Zhen", a data visualization tool that monitors the "*heartbeat*" of hundreds of thousands of customers. "*It is like an electrocardiogram that hospitals use to monitor a patients' pulse, if there is any abnormal sign, it draws our attention to it straightway*" (Alpha, 017a). This tool was a result of collaboration between the data team and the customer service team. The tool enabled the customer service team to closely monitor the transaction activity of millions of customers based on historical and real time data, in order to detect any abnormal activities. As the data director described:

"*We want everybody to like data and benefit from data. It is not easy as people often think only data experts can navigate and understand these sets of numbers. So, our goal is to turn big data into something that is tangible and easy to understand and use. Such productizing big data requires our team to work with other teams to understand their needs and customers' needs to come up with something that can benefit them on a daily basis." (Alpha, 014a)*

Information gathered from other informants, published industry sources and archival data confirmed that this company was well known for its data democratization capabilities. Data products such as "Seismograph" that monitors which functions customers complain about the most, including gaining access to the original customer complaint, and "Time Machine" that traces the evolutionary journey of customers, were designed from big data to enable a wide use of data applications. The informants from the same company often expressed a high level of confidence in the data team, referring to them as the *"magic maker".* One informant explained:

"*We do not need to have a PhD in physics or advanced IT skills to understand the data, everything is so straightforward to understand, it makes our job so much easier. They really did a fantastic job here*". (Alpha, 002a)

The Tangu case also indicates the link between the democratization of data and value creation. For example, in Tangu each department has its own data team. The purpose of the data team is to "*assist everybody in the department to be able to use data*" and to "*coordinate with other departments in terms of data integration and data sharing*". Informants often noted that "*without such collaboration between data experts and non-data experts, it would be impossible to build the bridge between the pure correlations (generated from big data) and real business value of such correlations*" (Tangu, 008a). The following observation provided by a senior data engineer at Tangu displays the relationship between the data democratization and conversion process.

"Big data may sound fancy but big also means slow, cumbersome, and complex. We used to set up the data team as a separate unit, which meant only a handful of people were able to crunch the data. The problem was then that the value of the data has not been used in an efficient manner as only a small group of people were working on it. We wanted to empower every single member of our team with the insights generated from big data." (Tangu, 003a)

In contrast, there was little mention from firms that generate limited value from big data concerning the enabling of large scale data applications. What was commented on suggested that employees' access to data and their understanding of data was not particularly germane to the value creation process. For example, one deputy executive was asked to describe the process of how data was used at the general managerial and operational levels. He answered that the "*business intelligence team analyse the results on a daily basis and present it to the senior management team, and the management team can select what is relevant and forward it to the lower level" (Cheu, 001a).* Such information distribution was confirmed by other informants from the same firm. However, one informant commented, "*Sometimes we don't really know the purpose of the document, to us, it's more like tick boxing exercises*" (Cheu, 010a)

The evidence on the linkage between the firm's capability to democratize data for wider use of application and value creation is consistent with the converse found at Shrong. For example, informants described that while other departments put a cap on how many people they could hire, the data department was given a green light. "*The problem is that they hired so many people*", one informant said, "*so far I didn't see much value from it*". At Shrong, informants often emphasized the technical role of data scientists but there was lack of interest in encouraging a wider scale of data use. In such firms, the informants indicated that big data was typically a luxury product for executives and data experts.

Overall, our findings concerning data democratization capability calls into question the implicit assumption that the analytical skill of individual data scientists is sufficient to contribute to a firm's value creation opportunities. Our evidence indicates that the process of data scientists transmitting relevant knowledge across internal firm boundaries to enable wider and collective data application is positively associated with better value creation. Leaders in these firms tend to view data as a public good that can be used and accessed by all employees. This discussion suggests the following proposition:

Proposition 2: In a big data context, the greater the firm's capability to democratize data to enable a wide range of data applications, the greater the likelihood the firm will increase its potential value creation.

Capability to contextualize data

This capability refers to the firm's ability to assign meaning as a way of interpreting the data within which an action is executed. Firms collect a significant amount and different types of data, including data on customer behaviour, market demand, shifting preferences and changing customer needs. The firm's capability to identify the contextual clues to gain a holistic view of customers will be positively associated with better value creation.

A key insight is that those firms that created more value from big data than others tended to actively break down the data barriers that existed within the firm. Such data integration enables firms to gain a "360 degree context" for customers. For example, Alpha have been working on integrating internal data since 2004, as documented in an in-house memorandum. The effort was made to integrate data from multiple systems, inviting collaboration among different functional units. As one informant commented:

"We have been working on a unified user cookie and log format for about 10 years. We wanted to build a data warehouse so people get a 360-degree picture of our customers. It is all about joining the dots and understanding data within a context" (Alpha, 021a).

The evidence we collected from other informants from the same company and from the secondary sources confirmed this view. It was emphasized that the real value of big data is not just the volume of the data itself, rather, it was the connection between "individual customers" and "their everyday life". As pointed out by a senior data expert, "*It is not just about what decision he/she made, it's about what scenario that drives him/her to make that decision*" (Alpha, 005a). Similarly, this view is also shared by Tangu and Bray. Creating a closed data loop, collecting and analysing customer data from multiple channels and magnifying the context about individual needs was constantly highlighted during our interviews. As explained by one informant,

"There are many different types of data, demographic data, historical data, situational data and locational data. By drawing all of these data together, you can have a full context of every customer interaction. Incomplete data often gives you a partial picture about customers. It's like the story of blind man and an elephant." (Tangu, 013a)

In contrast, we noted that many frustrations were expressed regarding information hoarding among business units within firms that created less value from big data. For instance, several informants from Cheu pointed out that data was stored in diverse lines of business and this prevented the firm from gaining a holistic view of individual customers. This is illustrated in the following observation:

"*Incomplete data can only tell you part of the story and people here seem to have the mentality to hold the information. It is not a win or lose situation here, it's not even about competition because when we mix our data together, we both benefit from it. The mentality needs to be changed and infrastructure needs to build to support this internal data collaboration otherwise everybody is doing their own thing with their incomplete set of data, nobody is doing a good job in this case.*" (Cheu, 011a)

In addition to failing to integrate data internally, many informants also noted that the outcome of data analysis from big data was often "*too abstract*" and "*you don't get much out of it from* *their data report*". A similar view was shared by other informants who described the information presented in the data report as "*too technical*" and "*not sure where the business value lies*". One informant pointed out *"delivering data and correlation without the context, it's rather pointless" (Cheu, 006a).* Our evidence also revealed that companies that are less active in building infrastructure to integrate internal datasets are also less active in engaging cross- department communication. This is reflected in the following:

"Not everybody can do (data analysis) that's for sure so we leave it for the experts. They dig out what is relevant first then we decide how to use our resources to make it happen or make changes." (Cheu, 010a)

From firms that have created more value from big data, we noticed that such cross-department collaboration was a common practice. Data teams frequently communicated with other departments to co-create value from big data. As one informant lamented:

"Albert Einstein once said 'If I had only one hour to save the world, I would spend fifty-five minutes defining the problem, and only five minutes finding the solution'. It is very time consuming and pointless to approach data with no question in mind. You have to put data in a context. So we spent a lot of time talking to our colleagues from different departments, what problems do they have? What data is relevant for this problem? Can it be solved by data? If data is available, what data do you need? With a particular context in mind, it is much easier to find a solution from the data."(Alpha, 005a)

As a result of this patterned collaboration, firms are more likely to understand data in different contexts, therefore leading to a better understanding of individual customers. This, in turn, made firms perform better and achieve prominence in creating value from big data. This suggests that:

Proposition 3: In a big data context, the greater the firm's capability to encourage cross-department collaboration to contextualize data, the greater the likelihood the firm will increase its potential value creation.

Capability to experiment with data

Data experimentation is the firm's capability to promote "trial and error", cultivate an inquisitive attitude towards data, encourage continuous experimenting with the data and monitor the changes. Across the sample companies, we found notable variations in the firms' actions in engaging in activities associated with data experimentation.

A positive example is Tangu, where a senior data analyst commented:

"People sometimes get very passive and just let computers take over. You need to have the balance between the 'yin' and 'yang'. For me, data analytics is the 'yin' and you need the 'yang' which entails coming up with new things to test the data, to test how the market reacts and fluctuates. There is a lot of hype about the role of data scientists and what they can do. Often they neglect the right side of your brain, the creative side of things." (Tangu, 020a)

Information from other informants from the same firm also accentuated the importance of data experimentation in driving the firm's value creation. A product development manager described the process of testing data in order to try out different ideas:

"*We sometimes try out different designs and features on our products just to see how the data react. With the internet, we can control the scale in which we wish to experiment. It is a process where you try it out, you gather feedback, and you improve, then try it out again on a larger scale, then feedback- and then improve." (Tangu, 002a)*

The findings suggest that a "trial and error" organizational culture, when coupled with a greater level of data accessibility, tends to have a better chance to transform value from big data within the firm's internal network. This combination is reflected in the following observation from a senior marketing director at Alpha:

"At the beginning, people felt quite intimidated by the data. They view data as it is, a scary set of numbers and statistics that they don't know where and how to start with. We want them to be curious about the data, to ask what things they could try to find out to improve customer experiences or make their jobs easier. The data support center can then begin to interrogate

the data with a clear focus. This is extremely important given the amount of data we have. We also encourage them to try out new ideas in incremental steps to see how the (real-time) data reacts. We had one idea coming from one of our interns whose idea was to produce personalized video clips from the historical data to capture our customers' journey from the time they first opened up a shop on our website. The results were very successful." (Alpha, 001a)

In contrast, informants from other firms put a lot of emphasis on the robustness of the data itself, where the experimenting element was mentioned rarely. For example, a Shrong executive expressed a high level of confidence in the firm's data analysis skills. Informants from the same company described the top management's attitude to big data as "*let the numbers do the talking"* or "*they stick too much to the number itself*". Similar examples can be found at Cheu, where the informants were asked to describe the role of big data in assisting their work activities. What was mentioned suggested that the insights generated from big data played a crucial role in influencing their jobs. "*It is a hard-core science*," as one informant described "*you don't want to mess with it*".

Our findings indicate that firms showing a tendency toward cultivating a learning and experimental culture also tended to have better conversion rates from the data. The findings suggest that a "trial and error" approach, when coupled with a greater level of data accessibility, tends to have a better chance to transform value from big data within the firm's internal network. This chain of evidence is consistent with the recent emphasis on managerial capability with the stress on flexibility, creativity and timing rather than efficiency and control (Amit and Schoemaker, 1993; Teece, 2007; Teece et al. 1997). This discussion suggests:

Proposition 4: In a big data context, the greater the firm's capability to cultivate a learning and experimental culture towards big data, the greater the likelihood the firm will increase its potential value creation.

Capability to execute data insight

This capability refers to the firm's ability to transform data insights into actions that lead to identification of new opportunities that increase the customer's willingness to use/pay and thus for creating value. Our analysis found a clear pattern of variation in how firms execute big data insights. Several informants from Alpha commented that the real value of big data depended heavily on the speed of the firm's execution ability. As one informant emphasized:

"*We had many discussions around how are we going to react quickly to these data and how we can capture the opportunities emerging from data there and then. The solution is not easy, it requires structural change, process change and also mind-set change. These are the hidden challenges associated with the management of big data. It is a different management where you build a process to enable people to make sense of the data and take actions to respond to the data in a timely manner. You need to catch that moment"*. (Alpha, 011a)

Archival data, such as minutes of meetings and internal newsletters, confirmed this view. Different from the traditional hierarchal decision making process, operational employees from Alpha have the flexibility and freedom to act based on the data insights. "*It is embedded in our daily routine, act more like a microscope, help us to see things better, clearer and quicker so we can make timely and right decisions.*" (Alpha, 009a) Individual team members were able to monitor the market trends and customer activities through a data tool based on historical and real-time data every day. Employees observing any abnormality emerging from the data were able to react to the situation in real time or could discuss it with the team leader, after which action was able to be implemented within 24 hours. "*It's not just about strategy anymore*", claimed one informant, "*it's the ability of execution*".

Similarly, the Tangu case also illustrates the linkage between capability to execute data insights and value creation. For example, one informant from Tangu described the execution ability to big data as "*creating* a *harmonious relationship with big data, like a dancing partner, data move, you move".* She further explained:

“*If you are constrained by the authority and hierarchy, then you lose that movement, agility and flow. Things change fast, and you have to catch that moment. It is all about speed. Time in this age is more important than money”. (Tangu, 001a)*

Tangu created a decentralized data ownership and empowered employees with data insights to make timely decisions. In contrast, firms that created less value from big data usually adopted a centralized data ownership, where only data experts and top executives could access the data. The decision-making process generally originated from the top, then cascaded down to the operational employees. For example, a data team would produce a data report, present it to the top management, this report would be forwarded to middle management, and then finally filtered to the operational employees. One informant from Shrong commented: “*it takes layer after layer of digestion, once it is your turn to digest and do something, it always quite late. So sometimes I don't really understand the purpose of the report. What do they want us to do? Read it?*” As a result of this kind of decision-making process, firms often missed opportunities to grasp the value emerging from real-time data, which hindered the firm’s ability to create value from big data.

Our findings support the view that decentralized power enhances the pace of decision making. This is consistent with Brown and Eisenhardt (1997, 2009) who suggest that firms with simple or few rules that combine improvisation tend to be more flexible to market shifts. This discussion leads to the following proposition:

Proposition 5: In a big data context, the greater the firm’s capability to build a decentralized decision-making process to enable agile execution of big data, the greater the likelihood the firm will increase its potential value creation.

Creating value from the firm’s open data network

This pattern of value creation refers to the process of attracting and integrating separate external datasets and resources, and managing this open data network to create value. Such collaborations can result in significantly higher value being co-created in comparison to value harnessed through the firm's internal big data. The alliance partners contribute complementary and supplementary data/capabilities, so together they are a source of value that individual partners could not build in isolation. All five IPCs emphasised that their platforms' internal resources and capacity can only be stretched to a certain extent to meet the diversified and heterogeneous market demand, particularly in the mobile internet age. Therefore, new infrastructure capabilities, as well as new resources centred on improving customer experience, need to be created and built in order to remain competitive in a fast-changing market.

Several informants highlighted the network externality effect of opening their resources to third-party developers, resulting in a distinctive resource within their network innovation systems. For example, a senior director from Tangu commented:

"We tried to do everything ourselves in the past. After we opened up our platform, we noticed that the most popular games were actually coming from outsiders. It is not about the data itself, it is about how to connect your data to with the 'a-ha' moment where you can provide better products for your customers based on the data. Data does not generate innovation, people do." (Tangu, 005a)

In addition to exposing the public APIs that allow third parties free entry to the supply of the platform, including its database, firms also coalesced datasets from different external sources. For example, in 2013, Shrong embarked on a collaborative partnership with Alpha, with the intention of combining two datasets in order to bring unique and valuable services to both platform users. A senior executive from Shrong commented:

"We both sit on a huge amount of data. They have huge data about e-commerce and we have huge data on customer behavior on the mobile internet. When we drew our data together, we

noticed that there were so many things that could be done or improved. That is the beauty of data, when you mix and match them, new and different patterns occur.' (Shrong, 005a)

Another example is the collaboration between Tangu and a local search engine, who agreed to integrate their respective datasets in order to jointly develop and cross-promote their products and services. Similarly, Bray pulled data together from various airline companies and the Civil Aviation Authority of China. This action not only improved airline operation efficiency but also provided a better experience for customers. A Chinese electrical appliance manufacturer entered into a collaborative relationship with Alpha with the aim of building internet products which allow the management of domestic appliances by remote control and enables communication between devices. A senior director from this electrical appliance manufacturer noted:

"Our data has different strengths and sheds different lights on our products and customers. We accumulated years of experience and data in delivering electronic products, and they (Alpha) also sat on a huge amount of data regarding market trends, customer preferences and different metrics. Such online and offline integration is the perfect marriage for product innovation (Alpha-partner, 2a).

A similar example is Cheu's collaboration with offline travel agents and travel relevant partners, such as Taxi Company, to further discover new opportunities from the collective dataset where all parties' resources can be fully maximized. When asked about whether they would be worried that such offline travel companies would "steal" Cheu's online business, a senior manager commented:

"We all have different strengths, resources and ideas, it is not like they don't have access to important data. Data is cheap to get and store these days and it is coming from so many sources including physical products. It is all about mixing and matching, every time you put different data together, you will see different results, different trends and different patterns. There are so many opportunities in the market. We want to grow the whole market together, instead of competing head to head in a small market space." (Cheu, 015a)

Similarly, Alpha and a local logistic company engaged in a “joint data crunching” project in order to improve delivery efficiency. A senior operational director from a logistic company commented:

“We have so many parcels to deliver and our drivers are under a lot of pressure. We captured and tracked thousands of package movements from telematic sensors and they (Alpha) tracked thousands of transaction activities from their websites. The data pot enabled us to redesign our logistic system. Together, we can predict how many packages will be generated for which province during different periods of the week.” (Alpha-partner, 6a)

The evidence resonates well with the role of external actors in contributing to the firm’s value creation in conjunction with the firm’s internal R&D by stimulating innovation through a combination of the efforts of a diverse pool of complementary firms, leading to superior firm performance (Chesbrough, 2003; Chesbrough and Appleyard, 2007). This open process enriches the firm’s own database through the synergetic effect from external knowledge sourcing, which leads to a great increase in innovativeness (Laursen and Salter, 2006). As a result, the firm is able to create a more compelling competitive position (Chesbrough and Appleyard, 2007).

It was evident that IPCs and their collaborative partners were able through collaboration to combine their datasets into “synergistic bundles” to generate economic rent (Madhok and Tallman, 1998). Our evidence demonstrates that the value of data increases when it is combined with different sets of data, a process which fundamentally shifts the nature of the relationship between the firms and their network partners from competitive to collaborative. Value can be realized in the form of new products and business model innovation, which ultimately expand the market opportunities for firms and their partners to create and capture value. This leads to the following proposition:

Proposition 6: In a big data context, data layering with inter-industry and intra-industry partners' data, allows preferential access to a greater variety of opportunities that generate additional sources of value for customers, thus increasing a firm's potential value creation.

A key observation is that instead of passively opening their platforms and sharing their data, all the study's IPCs acted in the role of supporter/enabler by providing tailored financial, technical and analytical data support to their partners. The evidence of this support is found consistently across all cases. Rather than focusing on acquiring more bargaining power with their partners, the five IPCs all adopted a conciliatory approach by providing free support first and only extracting a small commission fee when the developers started to obtain profit from their own operations. A senior director from Bray stated that:

"We are not like the United States, we are very short of talent and our infrastructure conditions are unevenly distributed across different provinces. There are so many people that would like to take advantage of the resources we have, but they simply do not have the data, knowledge or money to do so. You can see most of us (IPCs) acting as a gardener trying to fertilize the ground to make the flowers blossom. You plant the seeds, you fertilize them, and they will grow and then you can make profit from fruits or vegetables from the plant. It is like making a profit from your resources' interest. They tell us about their ideas and we help them to crunch the data. The stronger they get, the more profitable we are." (Bray, 002a)

Acting as an incubator, Bray launched Chengdu High-Tech Zone's 'Mobile Internet Venture Building', aiming to provide a range of services including free office space to start-up developers based in China's Western region. As another informant from Bray commented, "*When people have good ideas, we give ideas enough oxygen to make sure they grow*" (Bray, 010a). The concept "*the stronger they are, the stronger we will become*" was repeatedly emphasized across all the interviews. This idea was communicated clearly across the platform and its complementors. Bray further launched the "Light App" that is designed to increase the visibility of less popular apps, and its indexation feature through search engines boosts the light app's promotion and monetization capabilities. It bypasses the app store and makes finding services and information easy for customers.

Evidence revealed that all IPCs worked closely with their indirect business partners in order to overcome the deficient infrastructure/market conditions in China. Alpha offered tailored training programmes generated from big data to start an "Alpha villages" project that helped farmers and entrepreneurs from small and isolated villages to start their own businesses on Alpha's platform. The first village to take up e-commerce on a large scale was Dongfeng village in Jiangsu Province where more than 1,000 households started selling furniture goods online. A director of one of the trading agencies that Alpha worked with noted:

"They are offering training and banking services to help local farmers start their own business on its platform. They trained us so we can train more farmers from different villages. It is all free of charge." (Alpha-partner, 1a)

Another consistent theme emerging from the data is that all the leading IPCs are building their own ecosystems by providing technical and data analysis support for small-medium IPCs (SM IPCs). For example, companies such as Alpha, Tangu and Bray help SM IPCs to build up their online traffic, provide data-driven technological support to help SM IPCs to design and implement more advanced web monitoring systems, and offer free training sessions to help partners maximize the data potential. Rather than focusing on the short-term firm–level financial returns, all five case firms simultaneously devote more attention to cultivating an ecosystem that enable others to grow. We noted that in an emerging economy such as China, where the support market is either absent or weak, often referred to as an institutional void, the selected firms perceive it as a new opportunity space for further ecosystem building. For the firms in this study, the complexity of institutional situations in China, such as market deficiency, shortage of talent, underdeveloped infrastructure and a wide wealth gap between urban centres and rural villages, acted not as constraints, but as a stimulus for IPC value creation activities. One informant from Alpha commented:

"*E-commerce in the United States is a dessert. In China, it is a main course. Because the infrastructure of commerce in China was so bad that we had to develop and nurture our own system to make it work, so the boundary of our firm is getting bigger and bigger. And the infrastructure of e-commerce in the United States was so good, that e-commerce is a supplement" (Alpha, 010a)*

This statement suggests that the under-developed institutional environment of China provided a fertile ground for Chinese IPCs to use big data to nurture and strengthen their direct and indirect value creation providers within their ecosystem, thus acting as a catalyst to further strengthen and enlarge the scope of a firm's extended network.

The dominant view of investing in an ecosystem has two extremes, one is to make limited or no ecosystem investment in an institutional void context because no immediate or direct economic returns will be generated (e.g., Khanna and Palepu, 2000). The other view is to build a perfect ecosystem where firms can capture most of the economic value. Contrary to extant theories, our data suggest an additional alternative, in an emerging economy where the market support mechanisms are weak, firms can proactively engage in seeding an ecosystem by developing and cultivating new business partners while not trying to dominant the space and capture all the value. In our cases, the market was not viewed purely as the main place to appropriate value, instead, it was perceived as an important mechanism to stimulate building a broad ecosystem to foster social and economic development. This view is consistent with scholars who have championed the role of social entrepreneurs (e.g., Venkataraman, 1997; Peredo and Chrisman, 2006; Mair and Marti, 2009), whose primary objective is to create social value while creating economic value is a necessary condition to ensure financial viability (Mair and Marti, 2006). Our evidence revealed that by opening the firm's data, thus attracting and enabling resource constrained partners to participate more actively in its ecosystem, can lead to many sustainable business opportunities, consequently providing considerably more sustainable value creation. This discussion leads to our next proposition:

Proposition 7: In emerging countries where the support markets are absent or weak, the firm's capability to invest in seeding a big data ecosystem, while not trying to dominant the space and capture all the value, allows sustainable access to a greater variety of capabilities and resources, thus increasing a firm's potential value creation.

How do some firms create more value from a data ecosystem than others? The dominant view of network alliances is that building relation-specific assets with the firm's closed alliance partners can be safeguarded by a contractual or long-term relationship agreement (Dyer and Singh, 1998). In particular, according to this perspective, the firm's bargaining power and centralized approach in a closed alliance network is crucial to ensure the firm's ability to generate superior economic rent (e.g., Krause, et al., 2007; Lavie, 2006).

Contrary to extant theories, we noted that while firms engage in an open data network where alliances are multi-firm based, partners operate on a "modular" basis. The scope of the sharing network is dynamic as it evolves over time as new members join (Bharadwaj, et al., 2013; Han et al., 2012). Consequently, the traditional approach is insufficient to offer a robust explanation in the open data network context. Our findings indicated that firms perform better in managing their position in open networks if they are able to leverage their large customer numbers, databases and critical services provided by the open data platform infrastructure in order to contribute to their external partner's value creation.

Capability to build data relevancy

This capability refers to the extent of the firm's data and data analysis skills that are applicable to contribute to the external partner's value creation. Such a mechanism was constantly brought up during our interviews. For example, one informant from Bray commented:

"*We are competing (with other platforms) to attract all the (external) useful resources, (including) app developers, entrepreneurs, other platforms and other companies. We try to make ourselves as attractive as possible for potential alliances based on our data and skills. You need to make yourself indispensable to others; you need to make yourself relevant." (Bray, 006a)*

We noted that firms that have a central network position tend to possess a substantial amount of data and have better data analysis capability in supporting external partners' value creation. Many informants highlighted the network externality effect of opening up their resources to third-party developers, resulting in a distinctive resource within their network innovation systems. In addition to attracting developers, firms are also competing to build a data network. Findings indicate that new patterns and correlations in terms of location, relevancy and context, appear when one set of data is integrated with other datasets. This is consistent with the self-generating nature of information assets (Glazer, 1991). Coalescing datasets from different sources can therefore further stimulate the heterogeneous generation of value from the data. When asked about the key mechanism that drives partners to choose one platform over others, many informants commented that the accumulated value of the data platform and its analytical skills and service were the main factors. One of the developers commented:

"The top ones (platforms) all have huge customer numbers, we are talking about at least millions (of customers), and thinking about the value of these data, it is huge. Each platform has its unique strength, so it up to you to see where your strength and idea fits. How their data and service is going to help us to come up with new to better service our customers." (Tangu-partner, 2a)

In contrast, where a firm is struggling to build a central network position within its open data network, this is often associated with the firm's limited ability to create value from its existing database, as represented in the following description:

"*If we can't make sense of our data, how can we help others to see the real value of our data, how can we help others to make sense of the collective data pot. If we can't master this, people would only show very limited interested in partnering with us*". (Bray, 001a)

By building an environment enabling external actors to create value though accessing, integrating and creating opportunities from the focal firm's resources, such firms become hubs. They have a decisive influence in connecting external actors and developing "new rules of the game" that serve the inter-connected data network. This leads to our next proposition:

Proposition 8: In a big data context, the greater the firm's capability to contribute to external partners' value creation through its internal data and data analytical skill, the more likely it will increase its potential value creation in its extended data network.

Capability to orchestrate an open data platform

Our findings also highlighted that the firm's ability to build an infrastructure that enables data to flow within its data network is positively associated with its ability to create value from the firm's open data network. This capability refers to the functionality, usability and compatibility of technological infrastructure to facilitate communication and value creation within a firm's extended data network. This is also exemplified in the following from Cheu:

"In the data age, data will be connected and shared. While you connect all the data together, the storage, processing and sharing mechanisms need to be established to make sure it is easy to use, compatible with others' [technology]. Your system needs to be able to put everything together seamlessly, the levels of virtualization of hardware, software and storage. Whoever has the ability to build this [infrastructure] will have an influential role to play in this network. This is a huge challenge, but also represents huge opportunities. You can see the battle has already started." (Cheu, 001a)

A similar example can be found at Tangu, which invested heavily in building an open data platform. In addition to the technology related platform infrastructure, Tangu also paid close attention to building a data security system. Based on archival data and internal communications, Alpha invested around $1 billion on its cloud service and opened data centres in Singapore, the Middle East and the US. The underlying technology is an open data processing services platform, which our informant claimed can process petabytes of data in six hours (a petabyte is 10^{15} bytes of data). The platform's application goes beyond e-commerce. For example, the genomic research institute BGI used it in 2013 to sequence genes more quickly. ODPS has also been used to track weather patterns and pharmaceutical drugs sold in China. While such data generate great value, managing such a cloud platform is a challenge, as described by the data expert:

"*Managing that volume of data comes with big responsibility, how you can ensure data security, this includes access-authorization and data auditing, how can you monitor the data exchange, what the governance structure should be. You connect this amount of data and you have so many data users that are active on your platform, how can you manage this process, how can you monitor the data quality, how can you measure it?*". (Alpha, 020a)

This view was consistently emphasized in all of our cases. Our analysis indicated that firms which have better capabilities to build a technological infrastructure that enables data flow and data sharing within its extended data network, tended to be associated with greater levels of ability to build a leader position in an open network. This leads to our next proposition:

Proposition 9: In a big data context, the greater the firm's capability to manage an open data network platform the more likely it will increase its potential value creation in its extended data network.

CONTRIBUTION

By providing a new theoretical framework grounded in qualitative evidence, this research provides an important contribution to our knowledge of big data management. Organizational research has long recognized the role of resources as the key ingredient that drives the value creation of the firm (Barney, 1986; 1991; 2001). Although scholars have argued the need to understand the black box of “how” firms transform resources to create value, investigations of the specific process by which firms create value from resources have been insufficient (Barney and Arikan, 2001; Sirmon et al. 2007; Sirmon et al. 2011; Kraaijenbrink et al. 2010; Priem and Butler, 2001a, b). We therefore adopted a process-oriented approach in order to understanding how firms transform big data to create value. Our findings suggest that it is not the data itself, but rather what firms do that leads to value creation.

The main outcome and contribution of this work is an inductive process model that not only shows the process of value creation from big data, but also sheds insights into the mechanisms that can be used to explain how some firms are better at extracting value from big data than others. Data itself does not automatically generate value for customers. Several scholars have highlighted that it is not resources themselves, but the firm’s capability to manage the resource-related process that makes a difference to the development of value creation (Adner and Helfat, 2003; Bharadwaj, 2000; Helfat, et al., 2007; Sirmon et al., 2011). Our analysis showed that firms differ in their abilities to extract value from big data both internally within the firm and externally across the extended-data network. In developing our theoretical framework, several new insights emerged, in particular the notion of big data management, where managers can d*emocratize*, *contextualize*, *experiment* with data and *execute* data insights in a timely manner to create value, as well as the firm’s network orchestration capabilities that generate value creation from the firm’s extended data network.

Our analysis revealed four types of internal firm ability for harnessing big data to create value: data democratization, data contextualization, data experimentation and data execution. The facility to implement each of these capabilities critically impacts the firm’s ability to create value from big data. In regard to debates of the knowledge-based view as to whether it is the individual or the collective that is the source of new value, we propose that individual knowledge, such as that possessed by data scientists, is insufficient to maximize value creation from big data. Rather, value creation is closely associated with a collective process that transmits relevant knowledge across internal boundaries of the firm. This resonates with those scholars that emphasise that the collective locus of knowledge is the key to drive value creation (e.g., Eisenhardt & Martin, 2000; Kogut & Zander, 1995; Winter, 2003; Zollo & Winter, 2002). This view informed an understanding of the four knowledge management capabilities within the black box of the firm.

In addition to focusing on creating value directly from the firm's internal data, firms also pursued a collaborative strategy to create value indirectly from their data network, where a diversified and dynamic knowledge base becomes a heterogeneous resource network that is rare, scarce and difficult for competitors to imitate. The basis of competition is no longer confined to the firm-level but is extended to the data network level. Authority is no longer fully controlled by the firm, but is distributed among the different dataset owners, as each party is dependent on the others in order to generate new patterns, trends and ideas from the collective dataset. This is consistent with the "co-opetition" phenomenon where firms engage in simultaneous cooperation and competition with each other (Brandenburger and Nalebuff, 1996; Gnyawali and Madhavan, 2008). The nature of competition has thus moved away from a focus on the "I win, you lose" mind-set towards concerns with being excluded from participating in adding value to the given stock of available data. This finding concurs with prior research showing that Open Collaborative Ecosystems (OCEs) (Baldwin and Von Hippel, 2011; Curley and Formica, 2013) underscore integrated collaboration, co-create shared value and cultivate innovative networks, which lead to the proliferation of value creation opportunities.

Our findings expand on observations of capability at the firm level of analysis, by suggesting that through an open data network, the firm is no longer a self-contained entity. Through its extended data network, the firm's ability to build a leader position will depend on its capability to establish, maintain and reinforce symbiotic relationships with its partners. Our findings provide new insights about the importance and need for identifying managerial capabilities that result in significant outcomes for value creation from big data. This furthers our understanding of the resource orchestration literature by applying theory to the “big data” context.

Our findings also lend support to the assertion that firms which combine improvisation with low-to-moderately structured rules to execute a variety of opportunities are more flexible and able to respond to market shifts (Brown and Eisenhardt, 1997; Bingham, Eisenhardt, and Furr, 2007; Davis, Eisenhardt and Bingham, 2009). Our findings revealed that firms with a tendency to build a decentralized data structure, encourage cross-department collaboration, and promote “trial and error” and an inquisitive attitude towards data, are more likely to create value from big data. The focus of these companies was on building a flexible and decentralized organizational structure that allows improvised and agile responses to the data. Leaders in these firms also tend to view the data as a public good that might be accessed and understood by all employees.

Our findings also shed light on the unique institutional conditions in China. The accelerated rates of internet and smartphone penetration, cheap labour costs, uneven subnational infrastructure conditions and skills shortages in China, served as a stimulus that generated many value creation opportunities for companies. Prior research presented two extremes to invest in an ecosystem: Some scholars proposed making no or limited investment in an institutional void situation as no immediate or direct financial returns would be generated, others suggested building a complete ecosystem where firms can capture most of the financial value. In contrast, our data provided an alternative: By proactively investing in seeding an ecosystem through developing and cultivating new business partners while not trying to appropriate all the value allows a firm to have sustainable access to a greater variety of capabilities and resources. Our analysis revealed rich examples of how firms in China use the value generated from big data to support the conditions of institutional voids, which further broadened the firm's extended network. Acting as a "hub" for the firm's extended data network in a guanxi-oriented society, accentuating ties and reciprocal relationships can further strengthen a firm's ecosystem network. This research makes a unique contribution to contextualization by highlighting the distinctive institutional conditions in China and how they shape the firm's value creation activities in a big data context.

There is an important caveat from this study for practicing managers. Our data do not place data scientists at the centre of the value creation process. Our view is that it is the manager's capability to democratize, contextualize, experiment with data in a collective process, and build an organization structure to enable an agile response to execute data that makes an important difference in the value creation process. The findings also highlighted that in addition to harnessing data within the firm's existing data set, managers could seek broader potential value creation opportunities with external partners. By being the enabler in the inter-connected data network, a firm can have more influential power in managing such an expanded network, which leads to much more sustainable value creation opportunities. Our findings suggest that action patterns that involve different value creation processes, particularly managerial capabilities in which organizational competences can develop, create value better than others. This could offer informative concepts and relationships that managers can use to make deeper and richer assessments of the ways in which they manage data to create value.

LIMITATIONS AND FUTURE RESEARCH

In common with other research, this study has several limitations. First, our qualitative approach where a relatively small number of cases are analysed from a specific industry

(Perry, 1998) provides a limited basis for generalisation (Chetty, 1996). Moreover, we only study IPCs that adopted an open platform strategy after they were previously successful with proprietary strategies, therefore, our findings ought to be interpreted with caution as they may not be applicable to IPCs that initially unsuccessfully pursued proprietary strategies or to IPCs that have intended to build an open platform strategy from inception. In addition, this investigation focused on China, where market imperfections and scarcity of resources are particularly pressing. Whether our findings are replicable in other cultural settings or whether they are unique to the Chinese context is an empirical question. Future research can expand or test (e.g. using quantitative methods) our model on other industries from different countries.

Due to the novelty of big data, this study prompts several paths for future research. Further research could address questions such as: How is our model of large scale sharing across all kinds of databases compatible with the business strategies of organizations that are involved in this business? What are the necessary conditions that must be in place to build such shared ecosystems, and how will they share the risk or losses from inter-firm arrangements? How should firms manage data privacy issues? To what extent should companies open their platforms and how do firms benefit from these interfaces? What are the capabilities required to stimulate ecosystem collaborations?

References

Adner R and Helfat CE (2003) Corporate Effects and Dynamic Managerial Capabilities, *Strategic Management Journal* 24 (10):1011-1025.

Amit R and Schoemaker P (1993) Strategic Assets and Organizational Rent, *Strategic, Management Journal* 14 (1):33-46.

Baker T and Nelson RE (2005) Creating something from nothing: Resource construction through entrepreneurial bricolage. *Administrative Science Quarterly*, 50: 329-366.

Baldwin CY and Von Hippel E (2011) Modelling a Paradigm Shift: From Producer Innovation to User and Open Collaborative Innovation, *Organization Science* 22(6): 1399-1417.

Barney JB (1986) Strategic Factor markets: Expectations, luck, and business strategy, *Management Science* 32(10): 1231–1241.

Barney JB (1991) Firm Resources and Sustained Competitive Advantage, *Journal of Management* 17(1): 99-120.

Barney JB. (2001) Resource-Based Theories of Competitive Advantage: a Ten-Year Retrospective on the Resource-Based View, *Journal of Management* 27(6): 643-650.

Barney JB and Arikan AM (2001) The Resource-Based View: Origins and Implications. In Hitt MA, Freeman RE and Harrison JS (eds) *Handbook of Strategic Management* Oxford: Blackwell, pp. 124-188.

Bharadwaj AS, El Sawy MA., Pavlou PA and Venkatraman N (2013) Digital business strategy: toward a next generation of insights, *MIS Quarterly* 37 (2): 471-482.

Bingham CB, Eisenhardt KM and Furr NR (2007) What makes a process a capability? Heuristics, strategy, and effective capture of opportunities' *Strategic Entrepreneurship Journal* 1: 27-47.

Brandenburger AM and Nalebuff BJ (1996) *Co-Opetition*. New York: Doubleday.

Brown SL and Eisenhardt KM (1997) The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly* 42: 1-34.

Brown SL and Eisenhardt KM (1998) *Competing on the Edge: Strategy as Structured Chaos*. Boston: Harvard Business School Press.

Chen H, Chiang RHL, and Storey VC (2012). Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, 36: 1165-1188.

Chesbrough H (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press.

Chesbrough HW and Appleyard MM (2007). Open innovation and strategy, *California Management Review* 50(1), pp. 57-76.

Chetty SK (1996) The Case Study Method for Research in Small and Medium Sized Firms, *International Small Business Journal* (15): 73–85.

Curley M and Formica P (2013) *The Experimental Nature of New Venture Creation, Capitalizing on Open Innovation 2.0*. London; Springer.

Davenport TH and Patil D J (2012) Data scientist: the sexiest job of the 21st century. *Harvard Business Review* 90: 70-76.

Doan A, Ramakrishnan R, and Halevy AY (2011) Crowdsourcing Systems on the World-Wide Web, *Communications of the ACM* 54(4): 86-96.

Dyer JH. and Singh H (1998) The relational view: cooperative strategy and sources of inter organizational competitive advantage. *Academy of Management Review* 23(4): 660–679.

Eisenhardt KM (1989). Building theories from case study research, *Academy of Management Review* 14: 532-550.

Eisenhardt KM and Martin JA (2000) Dynamic Capabilities: What are they? *Strategic Management Journal* 21(10/11): 1105-1121.

Eisenhardt KM and Santos F (2001). Knowledge-based view: a new theory of strategy in: Pettigrew AM, Thomas H and Whittington R (eds), *Handbook of Strategy and Management*, London: SAGE: 139-164.

Eisenhardt KM and Graebner ME (2007) Theory Building from Cases: Opportunities and Challenges, *Academy of Management Journal* 50(1): 25–32.

Felin T and Hesterly WS (2007) The knowledge-based view, nested heterogeneity, and new value creation: philosophical considerations on the locus of knowledge. *Academy of Management Review*, 32: 195–218.

Financial Times (2016) Yuval Noah Harari on big data, Google and the end of free will. Available at https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5dd5a28c (accessed in Oct 2016).

Gandomi A and Haider M (2015) Beyond the hype: big data concepts, methods and analytics. *International Journal of Information Management* 35(2):137-144

George G, Haas MR and Pentland A (2014) Big Data and Management, *Academy of Management Journal* 57(2): 321-326.

Giles J (2012) computational social science: making the links. *Nature* 488(7412):448-450

Whittington R (2014) Information system strategy and strategy as practice: a joint agenda. *Journal of Strategic Information Systems* 23(1): 87-91.

Glaser BG and Strauss AL (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, NY: Aldine.

Glazer R (1991) Marketing in an Information-Intensive Environment: Strategic Implications of Knowledge as an Asset, *Journal of Marketing* (55): 1-19.

Gnyawali DR and Madhavan R (2001) Cooperative networks and competitive dynamics: a structural embeddedness perspective. *Academy of Management Review* 26: 431-445.

Goes PB (2014) Big Data and IS Research. *MIS Quarterly* 38(3): 3-6.

Grant RM (1991) The resource-based theory of competitive advantage: Implications for strategy formulation. *California Management Review*, 33(3): 114–135.

Han KW, Oh KS, Im RM, Chang H, Oh and Pinsonneault, A (2012). Value co-creation and wealth spill over in open innovation alliances, *MIS Quarterly*, 36 (1):1-26.

Helfat CE, Finkelstein S, Mitchell W, Peteraf M, Singh H, Teece D and Winter SG (2007) *Dynamic Capabilities: Understanding Strategic Change in Organizations*. Malden, MA: Blackwell.

Hoopes DG, Madsen TL and Walker G (2003) Why is There a Resource-Based View? Toward a Theory of Competitive Heterogeneity, *Strategic Management Journal* 24 (10): 889–902.

Huber GP (1984) The Nature and Design of Post- Industrial Organizations, *Management Science* 30: 928-951.

Jick TD (1979) Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly* 24: 602-611.

Khanna T and Palepu K (2000) The future of business groups in emerging markets: Long-run evidence from Chile. *Academy of Management Journal 43*(3): 268-285.

Kogut B and Zander U (1996) What firms do? Coordination, identify and learning. *Organization Science* 7(5): 502-518.

Kraaijenbrink J, Spencer JC and Groen AJ (2010) The Resource-Based View: A Review and Assessment of its Critiques. *Journal of Management* 36: 349-372.

Krause D, Handfield, RB and Beverly BT (2007) The relationships between supplier development, commitment, social capital accumulation and performance improvement. *Journal of Operations Management* 25: 528 – 545.

Langley A and Abdallah C (2011) Templates and turns in qualitative studies of strategy and management. *Research methodology in strategy and management* 6: 201-235.

Laursen K and Salter A (2006) Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. *Strategic Management Journal* 27: 131-150.

LaValle S, Lesser E, Shockley R, Hopkins MS and Kruschwitz N (2011) Big Data, Analytics and the Path from Insights to Value. *MIT Sloan Management Review* 52: 20-31.

Lavie D (2006) The competitive advantage of interconnected firms: an extension of the resource-based view. *Academy of Management Review* 31(3): 638–658.

Lazer D, Pentland A and Adamic L (2009) Computational social science. *Science* 323 (5915): 721-723.

Lepak DP, Smith KG and Taylor MS (2007) Value creation and value capture: A multi-level perspective. *Academy of Management Review*, 32 (1):180-194.

Liebowitz J (2013) Big data and Business Analytics. CRC press.

Lippman SA and Rumelt RP (2003) A bargaining perspective on resource advantage. *Strategic Management Journal* 24: 1069-1086.

Lincoln YS and Guba EC (1985) *Naturalistic Inquiry*. Beverly Hills, CA: Sage.

Locke K (2001) *Grounded Theory in Management Research.* Thousand Oaks: CA: Sage.

Mahoney JT (1995) The Management of Resources and the Resource of Management. *Journal of Business Research* 33: 91–101.

Mahoney JT and Pandian JR (1992) The Resource-Based View Within the Conversation of Strategic Management. *Strategic Management Journal,* 13: 363–380.

Mair J and Marti, I (2006) Social entrepreneurship research: a resource of explanation, prediction and delight. *Journal of World Business*, 41:36-44.

Mair J and Marti I (2009) Entrepreneurship in and around institutional voids: A case study from Bangladesh. *Journal of Business Venturing*, 24(5): 419-435

Martin JA and Eisenhardt, KM (2010) Rewiring: Cross-business-unit collaborations and performance in multi-business organizations. *Academy of Management Journal* 53(2):265-301.

McAfee A and Brynjolfsson E (2012) Big Data: The Management Revolution, *Harvard Business Review* 90(10): 60–68.

McKinsey Global Institute. (2011) Big Data: The Next Frontier for Innovation, Competition, and Productivity. June 2011. Lexington, KY: McKinsey & Company

Miles MB and Huberman AM (1994) *Qualitative Data Analysis: An Expanded Sourcebook.* United States of America: Sage.

Munford M (2014) Rule changes and Big Data Revolutionise Caterham F1 chances. *The Telegraph*, Technology Section, 23 February.

Nonaka I (1994) A Dynamic Theory of Organizational Knowledge Creation. *Organization Science,* 5(1): 14-37.

Okhuysen GA and Eisenhardt KM (2002) Integrating knowledge in groups: How formal interventions enable flexibility. *Organization Science*, 13: 370-386.

Parker G and Van Alstyne M (2005) Two-sided network effects: a theory of information product design. *Management Science* 51: 1494-1504.

Penrose ET (1959) *The Theory of the Growth of the Firm*. New York: John Wiley.

Peteraf MA (1993) The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal* 14: 179–191.

Peredo AM and Chrisman JJ (2006) Toward a Theory of Community-Based Enterprise, *Academy of Management Review* 31(2): 309–328.

Perry C (1998) Processes of a Case Study Methodology for Postgraduate Research in Marketing. *European Journal of Marketing* 32: 785-802.

Porat MU (1974) Defining an Information Sector in the U.S. Economy. *Institution for Communication Research*, Stanford University.

Priem RL (2007) A consumer perspective on value creation. *Academy of Management Review* 32: 219–235.

Priem RL and Butler JE(2001a) Is the Resource-Based View a Useful Perspective for Strategic Management Research? *Academy of Management Review* 26: 22–40.

Priem RL and Butler JE (2001b) Tautology in the Resource-Based View and the Implications of Externally Determined Resource Value: Further Comments. A*cademy of Management Review* 26: 57-66.

Rai A (2016) Editor's comments: Synergies Between Big Data and Theory. *MIS Quarterly*. 40(2): piii-xi.

Rindova V and Kotha S (2001) Continuous morphing: Competing through dynamic capabilities, form, and function. *Academy of Management Journal*, 44: 1263-1280.

Rothaermel FT, Hitt M and Jobe L (2006) Balancing vertical integration and strategic outsourcing: Effects on product portfolios, new product success, and firm performance. *Strategic Management Journal*, 27: 1033- 1056.

Rumelt RP (1984) Towards a Strategic Theory of the Firm," In Lamb B (eds), *Competitive Strategic Management,* Englewood Cliffs, NJ: Prentice-Hall, pp.556-570.

Schonberger VM and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. John Murray General Publishing Division.

Simon HA (1991) Bounded rationality and organisational learning. *Organization Science*, 2:125–134.

Sirmon DG, Hitt MA and Ireland RD (2007) Managing Firm Resources in Dynamic Environments to Create Value: Looking Inside the Black Box', *Academy of Management Review* 32: 273-292.

Sirmon DG, Hitt MA, Ireland RD and Gilbert BA (2011) Resource Orchestration to Create Competitive Advantage: Breath, Depth and Life Cycle Effects. *Journal of Management*, 37(5):1390-1412.

Stiglitz JE (1975) Information and Economic Analysis', IMSSS Technical Report 155, Stanford University.

Tavakoli A, Schlagwein D and Schoder D (2015). Open strategy: Consolidated definition and processual conceptualization. ICIS 2015 Proceeding.

Teece DJ, Pisano G and Shuen A (1997) Dynamic Capabilities and Strategic Management. *Strategic Management Journal* 18(7): 509-533.

Thompson JD (1967) Organizations in action – social science bases of administrative theory, McGraw-Hill.

Tsui AS (2007) From Homogenization to Pluralism: International Management Research in the Academy and Beyond. *Academy of Management Journal* 50(6): 1353–1364.

Venkataraman S (1997) The Distinctive Domain of Entrepreneurship Research: An Editor's Perspective, in Katz J and Brockhaus R (eds), Advances in Entrepreneurship, Firm Emergence and Growth, 3 (Greenwich, CT), pp.119–138.

Weick KE (1993) The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly* 38: 628-652.

West J (2003) How open is open enough? Melding proprietary and open source platform strategies *Research Policy* 32(7):1259-1285.

Whetten DA (2009) An examination of the interface between context and theory applied to the study of Chinese organizations. *Management and Organization Review* 5(1): 29–55.

Winter SG (2003) Understanding dynamic capabilities. *Strategic Management Journal* 24: 991-995.

Yin RK (2003) *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Yin RK (2009). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.

Yoo Y, Boland RJ, Lyytinen RK and Majchrzak A (2012) Organizing for Innovation in the Digitized World *Organization Science* 23(5): 1398-1408.

Zeng J and Glaister, KW (2016) Competitive dynamics between multinational enterprises and local internet platform companies in the virtual market in China. *British Journal of Management* 27 (3): 479-496.

Zollo M and Winter SG (2002) Deliberate learning and the evolution of dynamic capabilities. *Organization Science* 13 (3): 339-353.

Zott C (2003) Dynamic Capabilities and the Emergence of Intraindustry Differential Firm Performance: Insights from a Simulation Study. *Strategic Management Journal* 24: 97-125

Table 1: Background characteristics and data sources for cases

Company	Founding year	Ownership	Type (from inception)	Where traded	$ Sales/revenue (2013)	Year adopted open platform strategy	Number of informants	Data sources
Alpha	1999	Private	Online B2C retailer	NASDAQ	8.58 Billion	2008	11	Reports and strategic memos (27) Press articles (36) Interviews with firm's collaborative partners (8) Total interviews with informants including repeated interviews (23)
Bray	2000	Private	Online search engine	NASDAQ	5.2 Billion	2009	9	Reports and strategic memos (14) Press articles (21) Interviews with firm's collaborative partners (6) Total interviews with informants including repeated interviews (16)
Tangu	1998	Private	Online social network	SEHK	7.72 Billion	2011	9	Reports and strategic memos (25) Press articles (38) Interviews with firm's collaborative partners (10) Total interviews with informants including repeated interviews (21)

Shrong	1998	Private	Online content provider	NASDAQ	148.59 Million	2011	6	Reports and strategic memos (16) Press articles (17) Interviews with firm's collaborative partners (5) Total interviews with informants including repeated interviews (15)
Cheu	1999	Private	Online travel agent	NASDAQ	877.12 Million	2012	7	Reports and strategic memos (11) Press articles (23) Interviews with firm's collaborative partners (5) Total interviews with informants including repeated interviews (17)

Table 2 Summary of cases

Evidence					First order codes	Illustrative quotes	Value creation process
Alpha	Bray	Tangu	Shrong	Cheu	Statement about...		
A, b	A, B	A, B	A	A, B	Data driven product/service development from analyzing data	"It (big data) gave us much user experience data; their interaction with the platform itself, how they behave on our website, how they react about the new features we added on. It is like having the customer as part of our product development". "The data itself kind of made us more comfortable and confident when trying out new ideas. We could quickly come up with some ideas from the data or just some new ideas in general; we could then put it out and let the customers be the judge. They like it, we do it, they don't, we change it".	Internal data exploitation
A, B	A, B	A, b	a, b	A, b	Value added/personalized service from analyzing data	"The pattern (emerging from the big data) acted as a GPS telling us where different places are, who is congregating there and what kind of stuff they actually need. So, we are no longer led by the blind, like we were in the past". "Before (big data) you were serving a huge number of customers, it was a very vague crowd and it was very difficult to please this huge crowd. Now the technology makes it possible for us to understand the individual's (customer) behavior and preferences; you can use it to guide you to provide a personalized service".	
A, b	A, b	A, B	a	A	Business model innovations from analyzing data	"We use big data to understand client behavior and offer relevant financial services by leveraging our cloud computing service to speed up the response time to our customers and keeping our operational cost low". "One of the biggest advantages of such data is that it tells us what kind of things we missed or what kinds of spaces were untouched. These empty spaces led to many different value propositions, which enabled us to re-leverage what we have to fill those spaces."	
A, B	a, b	A, B	a, b	A, B	Diversity and heterogeneous platform from open platform	"It is like you are selling alcohol in the pub, once you make the boundary invisible, you are transformed from being a pub to a boundary-less plaza where you can sell thousands of new things". "Some have the mentality that they want to make money themselves by doing everything on their own. It is impractical and impossible to cover everything. You don't have the resources, even if you do, how thin	Attracting external resources

						can you be stretched? The thinner you stretch, the more vulnerable you are. Even if you have the resources, you don't have that diversified mind set to come up with these many new things. Sharing our database from our open platform opened up so many windows, it made possible for us to provide a "one stop" service station for our customers".	though internal data
A, B	A, b	A, b	A, b	A, b	Additional business model innovation from external partners	"The wisdom of crowds is often underestimated. Organizations often think they know better. It was an eye opener when we saw how external partners use our data to come up with different business models. We were struggling with how to make money from our platform and they (external partners) provided so many solutions. It was difficult to change that mentality because of the uncertainty and you don't know how vulnerable you will be when you open. Now after everything that has happened, we are so glad we did it at the right time" "People are overwhelmed by such massive data, if you don't use it, data is just data, it doesn't mean anything. Our ability is limited based on how much data we have. Data analysis is necessary but insufficient to maximize the whole value. Once you make it available to other people, trust me, you won't believe how many things these people can do with the data you have"	
A, B	A, b	A, B	a, B	A, b	Improved customer stickiness from sharing data	"The switching cost for customers is literately zero, given how many choices they have, how much information they can access. It is only a few clicks away. Their decisions are changing every hour, I am serious, and sometimes it is just a matter of minutes. It is very hard to get their attention nowadays. And our survival is purely dependent on their attention so we must do everything we can to get hold of that attention. What we can do is rather limited in a world like this. Through sharing our data, we are building a boundary-less human resource pool where we can collectively increase customer stickiness in every little detail. They are able to interact with a huge number of customers that we have accumulated for years and we can improve customer stickiness. Everybody wins". "They developed so many applications to address the different needs of customers; some for entertainment purposes, some for educational purposes, you name it. They (app developers) made the whole customer experience much more enjoyable and fun".	
A, B	A, B	A, B	A, B	A, B	New product development with intra industry partners	"We are collecting more data than ever before and the amount of data that has already been generated is huge. Organizations keep collecting data without having a clear idea how to use them. One thing is to share your data; another thing is to put different datasets together to generate new insights and trends. It is not like once you've used it, it loses its value. It is exactly the opposite, the more you use it in different contexts with different datasets, the more valuable it will be. We are working with Baidu to put our data together to generate much more relevant, context sensitive search results".	Attracting external resources through data conjugating

						"We are currently working with a shoe manufacturer to embed a sensor to the shoes so it can be tracked through the cloud technology. We have so many missing children every year and parents are devastated. So hopefully this new product will make a small contribution to solving this problem. It is a win-win".	
A, B	A, b	A, b	A, B	A, B	Business model innovations - with intra industry partners	"They (Shrong) are doing so well with the social networking site and their resources, their customer base and the data about their customers are very valuable assets to us. And we have millions of shoppers spending time on our website everyday buying and selling stuff. The data are collected from different perspectives: social networking and e-commerce. When you put it all together, we have made mobile shopping easier, fun and secure for our customers. Generally, customers are not willing to give you the money from their pocket, but when you make a strong case that the product/service you are providing is making their life so much better, they are more than willing to enjoy that service by paying you some money".	
A, B	A, b	A, b	A	A, b	New product and development with inter industry partners	"Everyone is talking about the 'internet of things'. The combination of offline and online by adding digital dimension or IP addresses to the products, air conditioners, cookers, your fridge, everything can be made "smart". They know the "hardware" side of things; we know the "software" side of things. When we put our strengths together, the potential is huge". "For example, we are working with the government, the public transport agency and Baidu to provide a better personalized travel experience for our customers. Stand-alone data is incomplete data; the data will become much more reliable and holistic when it's combined with different datasets".	
A, B	A, B	A, B	a, B	A, b	Business model innovations with inter industry partners	"You need to make yourself relevant to other people. In this case, you need to make your data relevant to other people. How is it going to help them? We are working with the banking services, insurance services and agriculture services by pulling our data together to provide better product/services to our customers. The tailored service for individual customers can lead to so many opportunities for new and better ways to capture value".	
A, B	A, B	A, B	A, b	A, b	Additional support for complementors	"We need thousands of entrepreneurs and we are growing our own entrepreneurs. When people have the passion, desire and eagerness to learn, we provide as much support as we can to enable them to grow. The stronger and more diversified they are, the stronger we will become". "We want to be long lasting. We always joked about wanting to build an organization that could last 102 years. It is like fishing; you could just catch as many fish as you want, but the fish will run out one day. You want to provide space and nutrition so they can last generations, which means you can make a profit over a much longer period of time. We are relying on our resources' interest to make a living".	Seeding an ecosystem through investing in institutional void condition

						"They provided free services to our farmers and free computers to the village trading center where everybody could use or learn how to use them. Some of us have never even seen one before. It was funny, nobody dared to touch it in the first couple of days and people always washed they hands before they touched it. They taught us how to set up our own stores, how to put our products on the website, and how to increase our online sales".	
A, B	A, b	A, B	a, b	A, b	Additional support for direct business partners	"The infrastructure in China is not ideal. We, as a company, do not exist in a vacuum. Everything we do is closely connected with the people and organizations around us. We generated some conclusions from the dataset and used this information to help the logistics company plan the logistics side of things. We did the same thing with the insurances company. The service that customers value should be seamlessly connected and we want to use what we have from the data to improve this connection"	
A, b	A	A, b		a	Additional support for indirect business partners	"We are working with the local training agency and local communities to set up training sessions and relevant services for our customers; some of them are particularly designed for the older customers. We understand their behavior, their preferences and concerns from the data and we know how to provide tailored support for them. In a big country like China, we have limited resources to offer training sessions to people from different places. So, we are equipping them (local training agencies and communities) with the best resources and material to enable them to do so. Some of the training sessions are free; some of the additional service comes with a reasonable fee. We are strengthening the connections within our ecosystem".	

Notes: codes for the evidence categories are as follows: "A", evidence from three + interviews with different informants from the same company; "a", evidence from less than three + interviews with different informants from the same company; "B", evidence from three + archival sources; "b", evidence from less than three + archival sources.

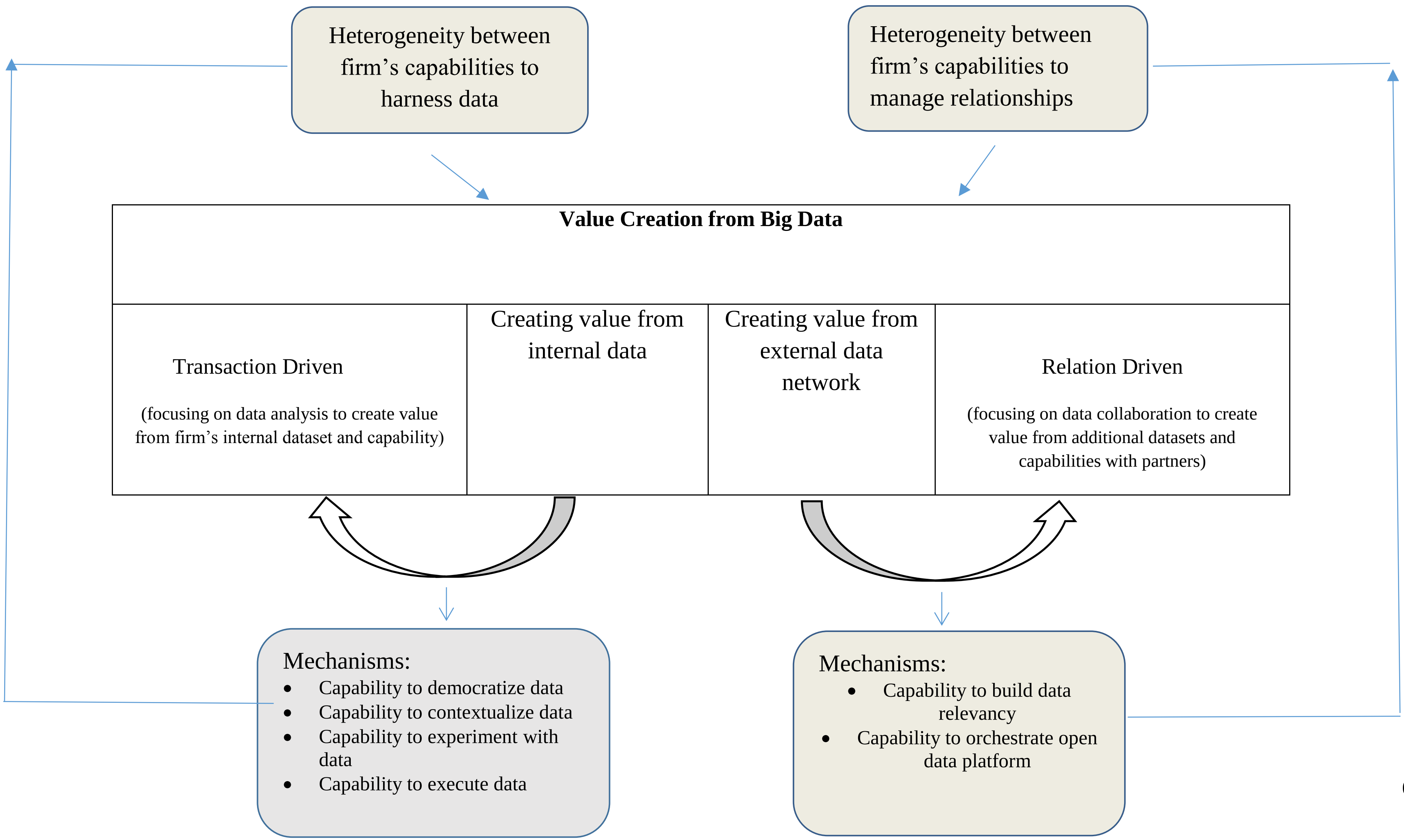
Table 3 Core capabilities and illustrative quotes

Creating value from internal dataset	
Capability to democratize data	"Most of our data analysts focused too much on the data itself, they don't have the ability of converting the knowledge they gained from data into simple tools that can serve the rest of the company, including executives, managers and front line employees. They are normally data driven rather than problem driven, this means that you need to ask the people you serve, what problem do you have? Can it be solved by data? Might data be the core of the solution? What kind of data do you think is useful?" "Our ultimate goal is to present data results into a non-data way, no complicated numbers and figures, tell a story with our data, turning data into an easy to use product or visualized tool that could turn everyone into data scientists. All of these data products are derived from data analysis. You turn the analysis concept and framework into a product, in essence, it is data generalization ability. This is the most important part of the job otherwise the request for data is huge and we will be swamped by the repetitive jobs". "It is not surprising that people get put off by the complicated numbers and graphics. It can be quite intimidating, all these statistics and numbers. The hard skill is their analytical skill, this is an essential requirement, but they also need to know how to use the data to tell a story, presenting in a way that people can get most insights from, how to sense the business context of these data, how to sense the business potential of these data, how to help other colleagues to use these data, I mean how to help other colleagues to self-serve use the data product we provided for them. Most of the organization has one, but not two and three, and it's the two and three that really set them apart from the rest." "You need a systematic approach when it comes to data management. First of all, do you want everybody or a small group of people to benefit from data? If it's the latter, then I would say that you are only touching the tip of the iceberg, I mean that you are only digging the very surface of the data value. If it's the former, you want to know the problems they are experiencing, don't know how to access the data, don't have the tools, don't have the authority, don't have the knowledge, don't know how to deal with real-time and fresher data. So, for having a data expert team, their job is to get everybody to benefit from data, solve all the problems I mentioned above".
Capability to contextualize data	"It's not just about plucking numbers, understand the correlations, this is just a very beginning of the journey, it's about understanding the context. For example, people often pay particular attention to the transaction data, who bought what, but what triggered this buying behaviour was often overlooked. They (data scientists) are very good at swimming in data, drawing correlations and visualizing data, and then making fancy, but very professional diagrams and tables for the executives. However, they often overlooked the context. Data without context is useless". "When data analysts present the data, I always ask them to discuss the business implication, how it's relevant to our customers or staff, how it might benefit our customers or staff. I don't want them to lock themselves in their own room playing with the data, I want them to talk to others from different departments to see how can they help to solve their problems through data". "You need to have that business sensitivity, it is an additional role you have to have. Without understanding the context of the data, you can never maximize the data potential. For example, rather than looking at who bought what, you need to understand under what context, or scenario he or she has the desire to purchase, what is the business potential, how can such potential be realized in the business context. This requires great commination between data experts and non-data experts".

	"Many companies when they collected data, they often found that the data is rather scattered, distributed in different data collection channels and operation staff. Incomplete data sometimes can be very misleading because it provides a partial context".
Capability to experiment with data	"People become fascinated about data itself and forgot important mechanisms to enabling to make these data "alive". The mechanism I am talking about is culture. You need to have that culture change, the culture that takes an innovative approach when it comes to data". "It is important to create a synergy between data and people. Data is important proving solid evidence, but it can only be to its best potential with the magic touch from people. That's why I say data cannot overtake all the jobs because there are certain aspects such as curiosity, creativity and imagination, things we are good at but data cannot do. That's exactly what we try to get our people to do when they approach the data, be curious, ask questions, be creative and use their imaginations, use data to create different stories, to create emotional connections with our customers". "Everything is 'we need to look at the data' or 'let the data do the talking'. We don't have the voice in the process. When it comes to data, we often have a blurred vision, it is like trying to admire a flower in foggy weather, and nobody really knows the true beauty of the flower".
Capability to execute data insight	"Data is data, unless you do something about it, otherwise it is just a set of numbers. We had many discussions around how are we going to react quickly to these data and how can we capture the opportunities emerging from data there and then. The solution is not easy, it requires structural change,process change and also mind-set change. These are the hidden challenges associated with the management of big data. It is a different management where you build a process to enable people to make sense of the data and take actions to respond to the data in a timely manner. You need to catch that moment". "Most organization collect data, actually, it's not about collecting data as it simultaneously generates itself from different sources, not many organizations can make the story come to life, the most difficult part is to act on it quickly. Because things change so fast, you need to act fast otherwise you just waste money and effort". "The mainstream of management style is layers after layers, you get data from the bottom up and then people on the top organize the resource distribution and give what people call 'strategic direction' to the people from the operational level. It was a useful method. But in the data age, from data mining and data analysis, most the decisions can be made through data, so it was completely unnecessary to report it up layers after layers, only wait a few weeks or even a few months to get a direction". "Data without action is useless. We are spoiled by the data every day and we spend way too much time on data without thinking about the things we can do there and then to respond to the insights data tell us".
Creating value from external network	
Capability to build data relevancy	"If you can't play the data well, how can you convince others to join your data party and throw their time, data and effort into it. It is not about how much data you have, volume is just a number, it is about how much you can make sense of these data and make it relevant to others. The more relevant you are, the better you are able to attract outsiders to join your network".

	“You want to put yourself in a central dot and think about how this dot is going to connect to other dots in order to make a spider’s web. The more dots rely on you, the more valuable or more central your role would be. The dot is how valuable your data is, how valuable your data is relevant to others”. “It is about what motivates others to join your data network, trust me, they (external partners) have many invitations and you want them to look at your platform, your data. It is not about how much you can make money from them, it is about how much they can benefit from coming to you. Here, the mindset needs to be changed, how can you turn yourself into a data tool to contribute to others’ success”. “People need water, electricity and air because they are heavily dependent on them. We want to turn our data into one of these essentials so we want to contribute to our partners’ life as much as possible. The more value they can get from our data and data insights, the better chance we are able to build a data ecosystem”.
Capability to orchestrate open data platform	“There are so many things to think of, how can you make an open data network, how can you manage the data security, how can you manage the coordination, how can you ensure the data compatibility? Do you have the technical infrastructure to support such an open data pot? How do you govern such an open data network?” “Managing this open data ecosystem comes with huge responsibility. You need to create rules that govern such open data interaction. You have a diversity of stakeholders with different interests, so how are you going to coordinate it and how to build policies based on collective agreement?” “The key challenge here is the management of the open data platform, the technical infrastructure is a prerequisite condition to have this platform, then it is about whether you have the ability to manage such a complicated network because you have so many data users, there are changes for the raw data, changes for the structure, correlations and logic relationships, how do you notice any abnormality in data change, how do you monitor the data quality, even the data life cycle management, it covers too much. And how can you price for all the different services in an open data platform?”

Figure 1: Theoretical framework of value creation from big data



Appendix: Interview Guide

1 Background information

- Can you please briefly tell me about yourself, e.g., your position at this company, key responsibilities, and how long have you been working here?
- Can you please provide an overview of the strategy of your company? In particular what is the strategy regarding big data?
- What is your understanding of big data?

2 General value creation process from big data

- Can you please describe how your company seeks to create value from big data?
- Can you please explain how such a process leads to commercialization/increases the customer’s willingness to use/pay?
- What are the key challenges/hurdles your company/you have encountered during such a value creation process?
- Who benefits from big data analytics? How did they benefit, and in what context?

3 The role of data experts

- What is role of data experts in your company?
- How are they involved in the value creation process?

4 Managers’ capabilities

- Can you please describe the managers’ actions or abilities that influence the value creation process?
- Have these actions you mentioned made a positive or negative impact on the value creation process/outcome?

Jing Zeng (Dr) is a Lecturer at the University of Kent at Kent Business School. Her main research interests include emerging strategies in the digital economy, business ecosystem, dynamic capabilities and innovation. She has published in journals such as *British Journal of Management* and *Management and Organization Review*.

Keith W. Glaister is Associate Dean and Professor of International Business at the University of Warwick Business School, He was previously Dean of Sheffield University Management School (2005-2013). He is Fellow of the Academy of Social Sciences and an elected Fellow of the British Academy of Management. He has published over 100 refereed journal articles and chapters, including papers in *Strategic Management Journal*, *Journal of Management Studies*, *British Journal of Management* and *Journal of World Business*. He is an editorial board member for several journals including the *British Journal of Management*.