Journal of the Association for Information Systems

Editorial ISSN: 1536-9323

Big Data Research in Information Systems: Toward an Inclusive Research Agenda

Ahmed Abbasi
McIntire School of Commerce, University of Virginia
abbasi@comm.virginia.edu

Suprateek Sarker
McIntire School of Commerce, University of Virginia
School of Business, Aalto University
suprateek.sarker@comm.virginia.edu

Roger H. L. Chiang
Carl H. Lindner College of Business, University of Cincinnati
roger.chiang@uc.edu

Abstract:

Big data has received considerable attention from the information systems (IS) discipline over the past few years, with several recent commentaries, editorials, and special issue introductions on the topic appearing in leading IS outlets. These papers present varying perspectives on promising big data research topics and highlight some of the challenges that big data poses. In this editorial, we synthesize and contribute further to this discourse. We offer a first step toward an inclusive big data research agenda for IS by focusing on the interplay between big data's characteristics, the information value chain encompassing people-process-technology, and the three dominant IS research traditions (behavioral, design, and economics of IS). We view big data as a disruption to the value chain that has widespread impacts, which include but are not limited to changing the way academics conduct scholarly work. Importantly, we critically discuss the opportunities and challenges for behavioral, design science, and economics of IS research and the emerging implications for theory and methodology arising due to big data's disruptive effects.

Keywords: Big Data, Behavioral, Business Analytics, Business Intelligence, Design Science, Economics of IS, Information Value Chain, Research Directions.

1 Introduction

As business processes become major differentiators for organizations in many industries, organizations are increasingly using analytics to "wring every last drop" of value from those processes (Davenport, 2006). Consequently, companies now view their data as a primary business asset (Redman, 2008). In organizational settings, the information technology (IT) function is tasked with managing and integrating data as an "enabler" of data-driven business processes and decision making (Chandler, Hostmann, Rayner, & Herschel, 2011; Lycett, 2013). Big data's rise has further amplified the importance of IT in this role (Horan, 2011), resulting in important implications for IT managers and scholars within and beyond the information systems (IS) discipline.

Recent editorials and special issues in top IS journals have discussed big data analytics from different vantage points. For example, Chen, Chiang, and Storey (2012) highlight the various domains and information sources associated with big data. They also touch on their vision for evolving analytics toward big data -- from the traditional structured relational database-driven paradigm to "business intelligence and analytics 2.0", which leverages Web and unstructured content, and to "business intelligence and analytics 3.0", which, in addition, encompasses mobile and sensor-based data. Goes (2014) presents valuable taxonomies for big data infrastructure and big data analytics. Agarwal and Dhar (2014) discuss challenges and opportunities pertaining to big data in information systems research and note that IS researchers are well positioned to take advantage of opportunities in this area. Sharma, Mithas, and Kankanhalli (2014) offer several research questions that IS researchers need to pursue. Adding to this conversation is the call for papers for *MIS Quarterly*'s upcoming special issue that emphasizes the need for IS researchers to examine and exploit big data's disruptive nature (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2014).

All of these commentaries, editorials, and/or special issue introductions highlight important facets of big data research in IS. In this editorial, we synthesize takeaways from prior expositions and expand on them by focusing on the interplay between big data's unique characteristics, these characteristics' implications for the information value chain, and potential areas of inquiry for the three major IS research traditions. We present a framework (Figure 1) that highlights this interplay and helps one to generate (and potentially refine) a set of meaningful research questions on this topic for the IS discipline. There has been a fair amount of IS research on the information value chain, which is the cycle of converting data to information to knowledge to decisions to actions and, thereby, generate additional data. Major bodies of IS research pertaining to the information value chain have examined the derivation and management of knowledge and decision making and actions; however, the effects of big data on the value chain remain relatively unexplored. Scholars and others often define big data by four key characteristics: volume, velocity, variety, and veracity; however, big data is not simply a matter of injecting additional scale, variation, speed, or noise to research data sets. As Chris Anderson of *Wired* famously said (2008, emphasis added), "Because in the era of big data, more isn't just more. More is *different*". In other words, these characteristics have the propensity to disrupt the traditional information value chain and result in a new "big data information value chain". Such disruption not only presents opportunities for novel research from within or across the different IS research traditions (e.g., positivist, interpretive, and critical behavioral research, design research, and economics of IS-based research) but also raises epistemological questions and challenges for some of the IS research traditions.

This editorial proceeds as follows. First, in Sections 2, 3, and 4, we provide an overview of the traditional information value chain and related IS research, big data's disruptive characteristics, and the resulting big data information value chain. With these sections, we highlight the profound impact of big data on people, processes, technologies, and, consequently, on organizations, industries, and virtually every facet of the world we live in[1]. Second, in Sections 5 to 11, we discuss possible research directions and implications for the three traditions of IS research alluded to earlier. In a way, we highlight the fact that we have much to learn about the big data phenomena and scholars in each tradition can and need to play a role in the research endeavor. We emphasize that that, by highlighting the three traditions and certain questions in our editorial, we do not wish to exclude other forms or areas of inquiry. In fact, we believe that we will see new traditions, genres, and areas of inquiry spring out of the IS community's engagement with phenomena related to big data.

[1] Those already familiar with the notion of "big data" and its impacts on the value chain may prefer to skim through this discussion or perhaps proceed directly to the discussion on research directions (starting in Section 5).

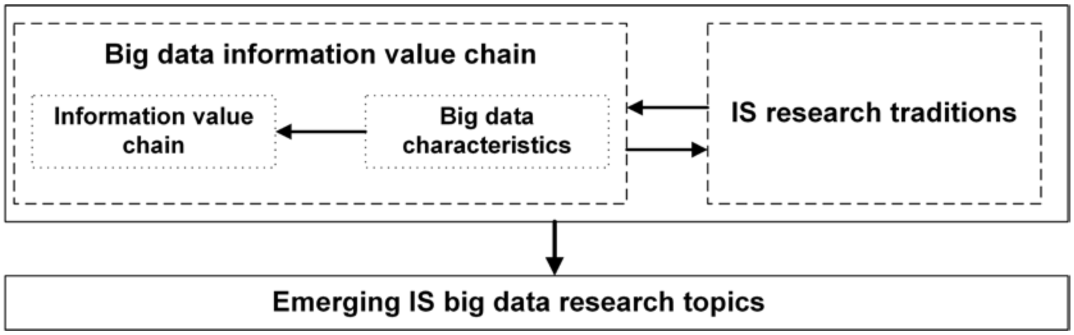

Figure 1. Overview of the Big Data Information Value Chain Perspective[2]

2 The Information Value Chain

The information value chain is the cyclical set of activities necessary to convert data into information and, subsequently, to transform information into knowledge (Fayyad, Piatetsky-Shapiro, & Smyth, 1996a, 1996b; Han, Kamber, & Pei, 2006), which individuals use to make decisions and take action. The decisions and actions then result in outcomes such as business value and additional data (Sharma et al., 2014). Each stage of the value chain encompasses people, processes, and technologies (Chandler et al., 2011). Information value chains operate in a given context; for instance, at the enterprise level, in a centralized IT unit, or for a specific functional or business unit. Figure 2 illustrates the traditional information value chain prior to the era of big data. The list of people, processes, and technologies shown in Figure 2 are illustrative, not exhaustive. As a further caveat, one could place certain processes and technologies differently depending on how one interprets data, information, and knowledge and how one delineates between decisions and actions. Here, we present the examples to illustrate the interplay between people, processes, and technologies across the stages of the traditional information value chain (i.e., a pre-big data era baseline). In Section 3, we contrast this traditional value chain with the new information value chain resulting from big data's disruption.

One can broadly categorize the set of activities in the information value chain into two groups: knowledge derivation and decision making (Chandler et al., 2011; Goes, 2014; Sharma et al., 2014). In the traditional information value chain, structured data is predominantly stored on premises in organizations' data centers that use relational database management systems (RDBMS). Many organizations integrate various structured data sources into data warehouses and/or data marts using extract, transform, and load (ETL) technologies. Database administrators, database managers, and ETL developers typically perform these tasks. Programmers or data analysts then process and analyze the data stored in these systems using structured query language (SQL) and use the resulting data as input for report generators, business intelligence (BI) tools, and/or analytical models incorporated in predictive technologies. In the value chain's knowledge stage, the people tasked with deriving knowledge from the data intersect with the decision makers; in other words, in the knowledge stage, the enablers and producers interact with the consumers of information (Chandler et al., 2011). In this stage, technologies such as knowledge management systems, corporate wikis, reporting tools, BI dashboards, and expert systems preserve or present existing knowledge or facilitate the creation of new knowledge through processes such as forecasting or reporting. Technologies such as decision support systems (DSS), recommender systems, and collaboration tools support managers and analysts' decision -making processes. The people, process, and technologies at various stages are also influenced by contextual factors such as the organizational/department/unit culture and IT governance.

Traditionally, IS research on the information value chain has also focused on the two aforementioned areas: 1) deriving knowledge and 2) decision making (Sharma et al. 2014). Scholarship in the area of deriving knowledge encompasses areas such as large bodies of work on knowledge management, structured data mining, and database design (e.g., Alavi & Leidner 2001; Chiang, Barron, & Storey, 1994; Storey, Chiang, Dey, Goldstein, & Sudaresan, 1997). The IS body of literature on decision making

[2] The arrow from the big data information value chain to the IS research traditions indicates the epistemological/paradigmatic considerations and challenges big data may bring to the way we conduct research and the types of criteria we might privilege in examining the big data phenomenon. The arrow between big data's characteristics and the information value signifies the disruptive impact of the four Vs on the traditional value chain.

is rich and extensive. Major areas of emphasis include research on designing decision support systems (DSS) and behavioral research on the effectiveness of IT artifacts or other decision aids for supporting decision making (Nunamaker, Chen, & Purdin, 1991; Wixom & Watson, 2001; Shim et al., 2002; Arnott & Pervan, 2008).

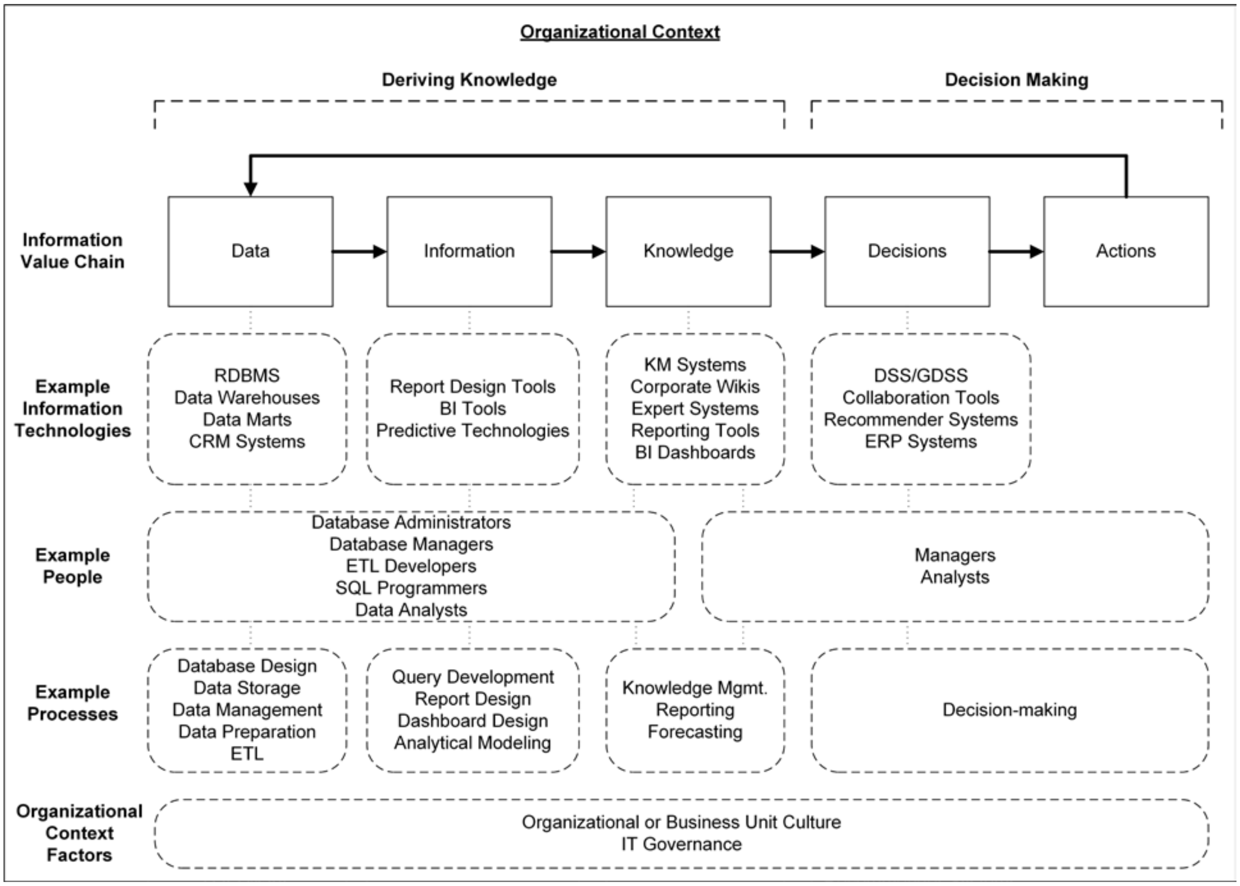

Figure 2. The Traditional Information Value Chain and Examples of the Accompanying People, Processes, and Technologies

In this era, academic scholars and practitioners have tended to use relatively scarce, largely static, and deliberately sampled and collected data (Kitchin, 2014a, 2014b). Having described the traditional information value chain, we now move on to describing the key characteristics of big data and elaborating on how these characteristics might disrupt the information value chain.

3 Enter Big Data: The Four Vs

One can separate big data and “regular-sized” data based on the presence of a set of characteristics commonly referred to as the four Vs: volume, variety, velocity, and veracity (Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012; Goes, 2014).

3.1 Volume

The U.S. Library of Congress, which archives both digital and offline content, has collected hundreds of terabytes of data (Manyika et al., 2011). Interestingly, the average company in 15 of 17 industry sectors in the United States has more data stored than the Library of Congress (Manyika et al., 2011), which underscores the fact that big data is pervasive across industries including finance, manufacturing, retail, health, security, technology, and sports. For a detailed discussion of various applications domains for big data, see Chen et al. (2012). Furthermore, in the vocabulary of big data, petabytes and exabytes have now replaced terabytes. For instance, large retailers each collect tens of exabytes of transactional data every year (McAfee & Brynjolfsson, 2012). To put these volumes into perspective using the classic grains of sand analogy, if a megabyte is a tablespoon of sand, a terabyte is a sandbox two-feet wide and one-inch deep, a petabyte is a mile-long beach, and an exabyte is a beach extending from Maine to North Carolina.

3.2 Variety

Organizations are now dealing with structured, semi-structured, and unstructured data from in and outside the enterprise (Schroeck et al., 2012). The variety includes traditional transactional data, user-generated text, images, and videos, social network data, sensor-based data, Web and mobile clickstreams, and spatial-temporal data (Chen et al., 2012; McAfee & Brynjolfsson, 2012). Effectively leveraging the variety of available data presents both opportunities and challenges.

3.3 Velocity

The speed of data creation is a hallmark of big data. For instance, Wal-Mart collects over 2.5 petabytes of customer transaction data every hour (McAfee & Brynjolfsson, 2012). With respect to unstructured data, over one billion new tweets occur every three days, and five billion search queries occur daily (Abbasi & Adjeroh, 2014). Such information has important implications for "real-time" predictive analytics in various application areas, ranging from finance to health (Bollen, Mao, & Zeng, 2011; Broniatowski, Paul, & Dredze, 2014). Simply put, analyzing "data in motion" presents new challenges because the desired patterns and insights are moving targets, which is not the case for static data.

3.4 Veracity

The credibility and reliability of different data sources vary. For instance, social media is plagued with spam, and Web spam accounts for over 20 percent of all content on the World Wide Web (Abbasi & Adjeroh, 2014). Similarly, clickstreams from website and mobile traffic are highly susceptible to noise (Kaushik, 2011). Furthermore, deriving deep semantic knowledge from text remains challenging in many situations despite significant advances in natural language processing.

4 The Big Data Information Value Chain

In his seminal book *The Innovator's Dilemma*, Christensen (1997) introduced and elaborated on the idea of disruption as relevant here. In organizational settings, disruptive phenomena significantly alter value chains. Indeed, big data and its four "V" characteristics have had a profound impact on the people, processes, and technologies related to the information value chain. As Mayer-Schönberger & Cukier (2013, p. 19) note, "The era of big data changes how we live and interact with the world". Simply put, big data means big disruption (Newman, 2014). Figure 3 illustrates this disruption in three ways. First, the new value chain involves a different set of people, processes, and technologies. While IT is known to exist in a constantly changing landscape, we can clearly see the accompanying changes to the people and processes attributable to big data as a disruptor. Second, there is greater amalgamation of technologies into "platforms" and processes into "pipelines" in the value chain's knowledge-derivation phase. Third, we see greater reliance on data scientists and analysts across all stages of the value chain to support self-service and real-time decision making.

Big data's four Vs clearly change how one stores and manages data. In terms of technical considerations, data's volume, velocity, and variety in organizations have caused IT departments to consider distributed storage architectures capable of handling large quantities of unstructured data. In terms of the technology, NoSQL ("not only SQL") systems, such as those leveraging Hadoop (Dean & Ghemawat, 2008) and/or Spark, have emerged as being better suited for the larger volumes and variety of unstructured data, while organizations commonly use in-memory databases to exploit velocity in real-time applications (Heudecker, 2013). In addition, firms are increasingly interested in *collecting* social media and sensor-based data (Chen et al., 2012) to supplement the internal data sources they have traditionally relied on, which contributes to the data's variety. However, using such data sources comes with an array of data-quality and credibility concerns, which require appropriate data-management, data-preparation, and knowledge-management activities. Data's volume and velocity have also caused IT departments to shift physical on-premises data centers to cloud-based infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and database-as-a-service (DBaaS) offerings better suited to meet organizations' elastic computing and storage needs (Buytendijk, 2014). The interplay between traditional schema-based structured data storage and management, new schema-less big data technologies, and organizations' increasing reliance on cloud computing can create complex big data architectures. In some respects, the key data-management and storage questions that practitioners pose have shifted to “what other internal/external data sources can we leverage” and “what kind of enterprise data infrastructure do we need to support our growing needs?”.

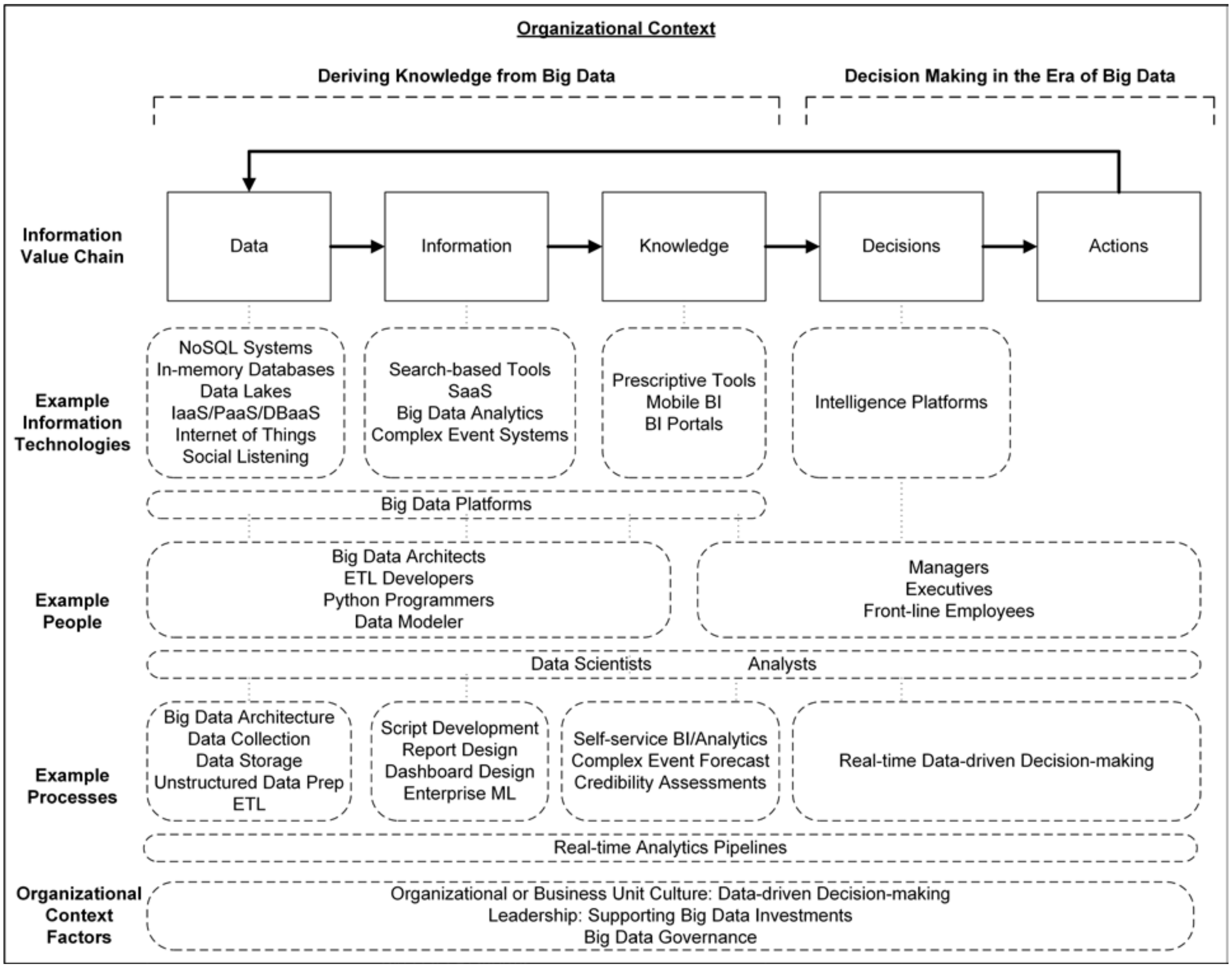

Figure 3. The Big Data Information Value Chain and Examples of Related People, Processes, and Technologies

The shift towards schema-less data storage and management coupled with organizations’ increasing desire to leverage big data as a source of competitive advantage has brought about the rise of data scientists (Davenport & Patil, 2012; McAffe & Brynjolfsson, 2012). In the words of Davenport and Patil (2012, p. 73), “data scientists are the people who understand how to fish out answers to important business questions from today’s tsunami of unstructured information”. Data scientists and scripting-oriented programmers now perform (or at least complement those who do perform) many of the activities pertaining to deriving knowledge from internally and externally collected data sources that database managers and SQL programmers traditionally performed. Data scientists also work closely with analysts and management in the decision making phase (Davenport & Patil, 2012). Furthermore, data lakes, which are essentially data warehouses or data marts specifically intended to serve as “sand boxes” for data scientists to experiment in, are becoming increasingly pervasive (Buytendijk, 2014). While ETL developers still play an important role, data-integration tasks increasingly entail fusing various noisy structured and unstructured data sources.

Consistent with the trend toward IaaS/PaaS/DBaaS, many technologies pertaining to the value chain’s information stage are also running in the cloud and accessed via software-as-as-service (SaaS) (Buyetdijk, 2014). Big data analytics allows data scientists and modelers to use enterprise machine learning—distributed scalable online algorithms running atop Hadoop platforms that can digest the volume and variety of information at unprecedented speeds. To offer an example, big data analytics running on Hadoop allowed Sears to push personalized promotions to customers more accurately and faster, which reduced the lead-time to one week compared to eight weeks in their previous data warehouse-based implementation (McAfee

& Brynjolfsson, 2012). Complex event-processing systems can analyze real-time sensor-based spatial-temporal data (Heudecker, 2013). For instance, the U.S. grocery chain Kroger has used overhead infra-red sensors to count customers and anticipate the number of currently needed checkout lanes and the number needed in 30 minutes, which has resulted in the company's reducing average customer wait times from four minutes to 26 seconds (Coolidge, 2013).

Big data's velocity and the trend toward data driven decision making have created an exciting paradigm shift in how organizations create and leverage knowledge for decision making. The biggest shift is organization's consuming analytics in real time with the rise of "self-service" BI/analytics (Chandler et al., 2011). This shift is attributable to the fact that "Big data, and the fast pace and complexity of today's marketplace, require that leaders make decisions faster than ever before" (Kiron et al., 2011, p. 5). Self-service BI/analytics allows various employees in an organization, including managers and executives, to independently generate custom reports, run basic analytical queries, and access key performance indicators across various devices without relying on IT or decision analyst support. These factors help organizations avoid time-consuming hand-offs and make decisions in an agile manner (Chandler et al., 2011).

The organizational context undoubtedly has impacts on the information value chain. Firms transformed by big data are often ones where a "strong top-line mandate to use analytics supports a culture open to new ideas" (Kiron et al., 2011, p. 5). Similarly, business units or departments with a longer-standing tradition of data-driven decision making, such as finance and operations, tend to leverage big data more relative to departments such as human resources (Kiron et al., 2011). Furthermore, big data governance practices have important implications for the availability, quality, maintenance, and security of the variety of novel structured and unstructured data sources that are part of the big data information value chain.

5 Big Data: Implications for IS Research

The big data information value chain has several implications for IS research. The first set of issues relate to deriving knowledge from big data. In this area, prior editorials have examined the effects of "big research dataset" usage in academic research. These effects primarily pertain to sensemaking, which is the process of deriving knowledge based on information extracted from big data (Lycett, 2013). Below, we discuss the epistemological/paradigmatic issues, theoretical implications, and methodological challenges pertaining to "big research datasets". However, we must emphasize that research implications of deriving knowledge from big data extend beyond the challenges of using "big research datasets".

The disruption to the traditional information value chain attributable to big data affords a plethora of research opportunities for the three IS research traditions. For instance, as organizations collect and store more customer data than ever before, privacy, security, and ethical considerations come to the forefront. With the proliferation of NoSQL and Hadoop systems, the advent of data lakes, and the popularity of in-memory databases, we need big data modeling formalisms and integration artifacts. Increased emphasis on complex event forecasting, credibility assessment, and social media analytics presents opportunities for novel prediction/description artifacts. Social listening platforms and the Internet of things allow novel forms of analysis pertaining to user-generated content and sensor-based data rich in knowledge, opinions, emotions, location, and geographic information. More broadly, with organizations treating data as a primary asset, assessing and, in some cases, quantifying the value of volume and variety relative to the costs of veracity becomes of paramount importance to evaluate the effectiveness of big data investments. We discuss some of these opportunities in Sections 6-8.

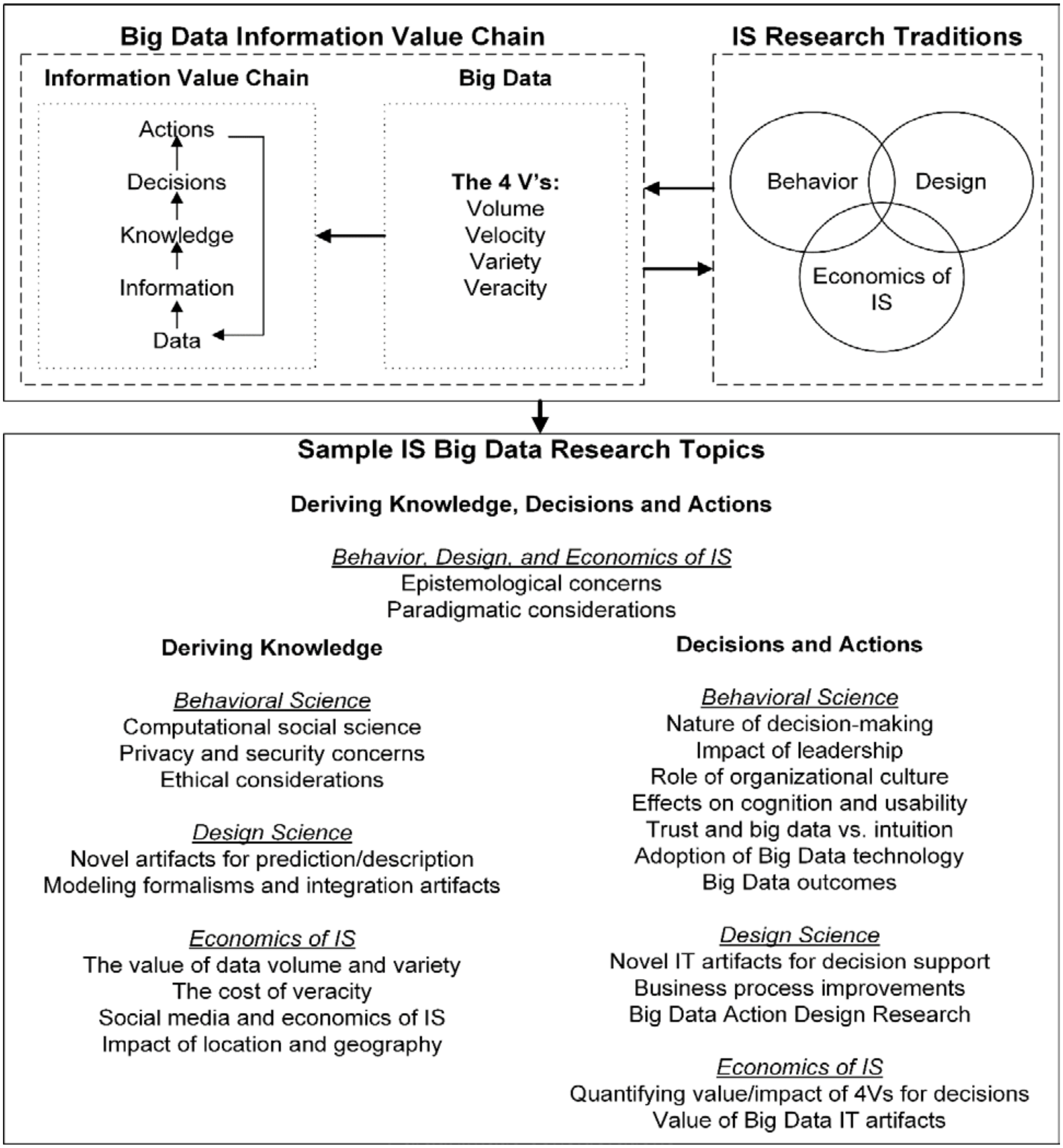

Figure 4. Toward a Big Data Research Agenda for IS

The second set of implications pertains to big data's impact on decisions and actions. This area of inquiry, which few prior IS papers have focused on, can potentially view big data and its characteristics as impacting IT artifact-related perceptions and behaviors in behavioral studies. We also believe that the four Vs of big data potentially change the very nature of IT artifacts, much like communication and collaboration technologies altered decision support systems and knowledge management, which gives rise to a new class of big data IT artifacts. IS research needs to not only contribute to the design but also examine the feasibility and effectiveness of such IT artifacts for different stakeholders. For instance, with the proliferation of real-time data-driven decision making and self-service analytics, executives, managers, and front-line employees are increasingly beginning to use big data to support timely decision making, which raises questions about the tension between data and intuition, implications for the nature of decision making, and appropriate ways to quantify the value/impact of the 4Vs on decisions. The above-mentioned issues also prompt researchers to think about the effects of cognition and usability and of the broader organizational context that includes but is not limited to organizational culture and leadership, big data investment, and adoption outcomes. Real-time analytics pipelines built on big data platforms are augmenting or automating various business processes. A question that arises is: under what circumstances can (or should) one employ such alternatives for improving business processes? Big data also presents opportunities for designing and implementing novel decision-support artifacts and necessitates research on assessing the value of big data IT artifacts in different organizational contexts. In the ensuing sections, we elaborate on these and other opportunities.

The key takeaway is that big data is not only different but also highly disruptive to the academic research process and to practice related to data, which requires that we re-assess our research assumptions, methodologies, and substantive questions. Furthermore, due to big data's impact on people-process-technology, these implications extend to behavioral science, design science, and the economics of IS. In their paper interestingly entitled "Why IT Fumbles Analytics", Marchand and Pepper (2013) discuss the usability and cognitive challenges associated with big data analytics. They argue that scholars and practitioners have focused too much on big data's technical facets and not enough on the people and their institutional and social environments, which has created a lack of socio-technical harmony that IS implementation initiatives often need to succeed. They note that "big data and other analytics projects require people versed in the cognitive and behavioral sciences, who understand how people perceive problems, use information, and analyze data in developing solutions, ideas, and knowledge" (p. 109), which appears to be well aligned with the core strengths of the IS discipline. Keeping in mind the aforementioned sets of implications from the perspective of the information value chain, Figure 4 outlines promising areas of research for big data research in the IS discipline. In Sections 6-8, we discuss possible research areas for the three IS traditions. We reiterate that the goal here is to maintain a broad and inclusive perspective and to be illustrative rather than exhaustive in our coverage.

6 A Big Data Research Agenda for Behavioral IS Research

6.1 Epistemological Concerns of Big Data and Behavioral IS Research

The four Vs, particularly volume and variety, present challenges not only regarding the changing nature of big data phenomena as they exist in practice but also regarding how, and with what assumptions, research is conducted to investigate such phenomena. Some commentators have brought to question the value of the traditional scientific model of research (which often involves constructing hypotheses based on guesses and then deductively testing the hypotheses using carefully sampled data (e.g., Gregor & Klein, 2014)) in a data-abundant environment associated with big data (Anderson, 2008; Kitchin, 2014b). Will "machine-generated correlations" on big data be enough to specify "inherently meaningful and truthful" patterns and relationships (Kitchin, 2014b, p. 135)? Can these correlations render front-end theorizing (before empirical analysis), which is at the heart of the scientific research model, meaningless (Kitchin, 2014b)? And, even if these patterns do not lead to understanding, will we be able to accurately predict behaviors, which is all that some important stakeholders may care about (Shmueli & Koppius 2010; Agarwal & Dhar, 2014)? Is an inductive "mode of science" in development, wherein algorithms "spot[s] patterns and generates theories" (Steadman, 2013, in Kitchin, 2014b, p. 131)? Are we nearing the "end of theory" (Anderson, 2008) given that "data tells the truth" under the assumption that studies have access to exhaustive data (that is, n = all), whereas "theory is merely spin" (Kitchin, 2014b, p. 135)?

6.2 Behavioral IS Research on Deriving Knowledge from Big Data

Many scholars, including Agarwal and Dhar (2014), do not perceive a fundamental transformation in the philosophy underlying research. Rather, they see potential for using a "guided knowledge-discovery" process that leverages big data in conjunction with traditional data-collection methods to generate insights and preliminary hypotheses worthy of further examination through a hybrid process of induction, deduction, and abduction. One perspective is that big data triggers a "paradigm shift towards computational social science" research (Chang, Kauffman, & Kwon, 2014, p. 67). In this perspective, scholars have argued that new "unobtrusive" big data information sources facilitate realism and generality with appropriate levels of control. Examples include using social media and Web clickstreams to enhance customer survey data to better understand the "voice of the customer" (Kaushik, 2011, p. 9), using large social networks to understand the dynamics of influence (Aral & Walker, 2012), and including mobile sensor-based data for enhanced spatial-temporal behavior analysis. Yet, other researchers note that such data is not "neutral, objective, and pre-analytic in nature" but reflects a certain underlying world-view, philosophy, or theory-in-use that has informed the design of measurement instruments (e.g. Kitchin, 2014b, p. 2). Researchers and data scientists will need to reflect on this perspective on big data when making truth claims.

Another major area in the broad realm of computational social science is workforce analytics (Davenport, Harris, & Shapiro, 2010). Today, many organizations use standard statistical techniques to combine employee perceptions derived from surveys with objective data from economic reports and/or measured through technology usage logs and sensors to make connections between employee satisfaction levels and sales, productivity, retention rates, and shrinkage levels (Coco, Jamison, & Black, 2011). For example, Best Buy found that a 0.1 percent increase in employee engagement resulted in a $100,000 annual increase in revenue per store (Davenport et al., 2010). Since workforce analytics examines the interplay between employee perceptions, technology usage, and business value-related outcomes, behavioral IS researchers should be well suited to lead research studies in this important area.

While the preliminary results appear promising and present a great opportunity for the IS community to further develop and apply suitable data analytic techniques on large data sets, we need to critically reflect on several issues. First, we need to examine the assumptions underlying the data (e.g., how well the data represents the population and whose interests may be excluded or overrepresented) since the condition of "n = all" does not hold in most cases (Kitchin, 2014b). Second, we need to be aware of the assumptions underlying the analytic techniques (i.e., how value-free the techniques and algorithms are, and how compatible the assumptions underlying the techniques are with the nature of data). Third, and more importantly, we need to ensure that applying big data analytics actually leads to desirable economic and humanistic outcomes for relevant stakeholders. Clearly, both qualitative and quantitative researchers have an important role to play in rethinking and refining how big data is collected, prepared, analyzed, and presented and in investigating the actual processes and consequences of using big data analytics. For example, qualitative researchers, who one might not typically think of big data researchers, can seek to contribute to this arena by examining fundamental questions. These may include: "Are decision making processes at various levels of the organization being transformed due to big data, and, if so, how?" or "What kind of fit is needed (say, between the architecture/algorithms and the organizational structure/culture) for big data initiatives to be effective in organizations, and how can one cultivate such a fit?".

When using big data sources to derive insights, privacy considerations become paramount. In fact, Barocas and Nissenbaum (2014, p. 33) contend that "big data extinguishes what little hope remains for the notice and choice regime", which provocatively points to the futility of organizations' sharing their policies regarding collected data and the "opt-out options" they may provide. Analytics concerns how data from various sources is assembled and mined (Fayyad et al., 1996). However, what data actually gets mined, and using what techniques, is often emergent and not necessarily defined upfront, which makes individuals particularly vulnerable to privacy invasions. Interestingly, a poll conducted by KD Nuggets showed that "data mining" has remained the prevalent term in academia, while "analytics" has become more widely adopted in industry over the past decade (KDNuggets, 2011). It appears that "data mining" has a more invasive connotation, whereas industry has carefully marketed analytics as being progressive and associated with intelligence and business value. While analytics may have made data mining more socially acceptable, the underlying privacy concerns, which big data's characteristics such as volume and variety amplify, remain. We know about many high-profile examples of privacy infringement in both industry and academic contexts, including one where the father of a teenage daughter discovered that she was pregnant owing to Target's highly accurate marketing efforts (Hill, 2012). Similarly, Facebook's intentionally manipulating users' moods has raised many critical questions (Kramer, Guillory, & Hancock, 2014; Agarwal & Dhar, 2014). Often times, organizations tout "informed consent" and upfront notice as an answer to critics; yet, as Barocas and Nissenbaum (2014, p. 32) note, "upfront notice is not possible because new classes of goods and services reside in future and unanticipated uses". Identifying acceptable levels of intrusion, and finding the right balance (and the principles of balancing) between insights obtained due to access to big data and the infringement such access results in is an important area of inquiry for IS scholars.

Similarly, it is virtually unavoidable that big data will continue to see large breaches. With the rise of self-service analytics and with access to sensitive data more pervasive than ever in and across organizations, security behavior challenges regarding compliance, insider threats, and so on will grow as important areas of research.

Big data analytics involves deriving knowledge and gaining insights. The pursuit of knowledge has always had a close relationship with ethics. One primary component of ethics already mentioned is privacy considerations. Beyond privacy, firms and researchers leveraging big data may consider deontological ethical criteria of consistency, equality, accountability, integrity, and an all-around conscientiousness (e.g., Chatterjee, Sarker, & Fuller, 2009a, 2009b). These issues underlie much of the concerns related to the "transparency paradox[3]" and the "tyranny of the minority[4]".

6.3 Behavioral IS Research on Implications of Big Data for Decisions and Actions

As industry moves towards self-service analytics using big data, several research questions arise, such as the impact of sources and collection methods on data credibility, the impact of access to knowledge on employee satisfaction and knowledge transfer, and organizational norms for knowledge access and transfer (Chatterjee & Sarker, 2013). Further, as big data introduces novel IT artifacts that support large-scale, self-service, real-time analyses and decision making from vastly integrated enterprise-wide analytics, behaviors and perceptions remain critical to the process of effectively converting knowledge to appropriate decisions and actions. These emerging data sources, decision making processes, and IT artifacts present an opportunity to revisit questions related to constructs, such as trust, leadership, knowledge transfer, and decision making.

Along these lines, Davenport and Harris (2007) analyze numerous successful organizations to derive key elements of the "anatomy of an analytical competitor". Scholars have used these traits to categorize organizations' capability maturity levels from aspirational to transformed (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Two of the important differentiating elements they identify are people and culture. With respect to these two elements, Davenport and Harris wonder: how does analytical leadership emerge? What are the characteristics of analytical executives? What role do different c-level executives play? What if executive-level commitment is lacking? What traits of big data analysts make them effective? How do organizations transform from intuition-based decision making to a data-driven decision making paradigm? Edward Deming is often falsely attributed the famous quote "In God we trust, all others bring data". As we mention earlier, it is often (naively) believed that data and algorithms result in what is necessarily just and true. However, the importance of intuition and judgment, especially in uncertain environments cannot be underestimated (Davenport, 2006; Yetgin, Jensen, & Shaft, 2015). Given that there is a natural tension between data and intuition (Davenport & Harris, 2007), what is the ideal balance between data and intuition? What kind of theory can inform this balancing act? What role does trust play? How do the four Vs impact user perceptions and intentions to use big data IT artifacts? More broadly, we need to understand if and how we should revise existing decision making models (e.g., Hodgkinson & Starbuck, 2008) to reflect how decisions are actually made in organizations using big data and big data analytics. Also, given people and culture can potentially act as impediments to the adoption of big data analytics in organizational settings, what theories and models are appropriate for avoiding implementation failures due to human and cultural issues? From an HCI standpoint, as dashboard-based visualizations become the norm for "managerial cockpits", we need to investigate what the implications of the four Vs are on how users handle cognitive loads resulting from big data. Finally, we need to conduct balanced assessments of outcomes of the implementation/adoption and use of big data (i.e., big data infrastructure and big data analytics). We need to conduct critical, intensive assessments of the *actual* impact of big data investment and use and understand if and how one can attain instrumental benefits (such as performance and profitability) and humanistic benefits (such as empowerment and freedom). In other words, big data offers the opportunity to re-examine some of the same phenomena pertaining to decisions and actions that behavioral IS researchers have examined over the years for different waves of technological innovations.

6.4 Summary of Sample Big Data Research Opportunities for Behavioral IS

As Table 1 shows, behavioral researchers have numerous opportunities to contribute to scholarship on big data. Some potential areas include epistemological reflections and methodological development, contributions to computational social science, and reformulations of existing theories and development of new theories of human decision making/human behaviors for the new data abundant environments using traditional or newer computational approaches. In particular, privacy, security, and ethics of big data have significant implications

[3] This paradox is related to the observation that "datafication" (Lycett, 2013) is assumed to make processes transparent, yet the ways in which data is collected and mined remains unknown/inaccessible to the public (Richards & King, 2013).

[4] This is related to the idea that companies tend to make inferences and generalizations based on data from the minority who agree to share data and, thereby, silence (i.e., make irrelevant) the views of those who refuse to give consent (Barocas & Nissenbaum, 2014, p. 32).

and, hence, deserve special attention. Developing in-depth consultable case studies that capture the intricacies of applying big data approaches in complex social environments would also be of value.

Note that the areas listed below in the table are not an exhaustive list of possible research avenues. Rather, they signify the breadth of possibilities and directions that have opened up from this interest in, and trend toward harnessing, big data.

Table 1. Big Data and Behavioral Research: Sample Research Opportunities

Value chain stage(s)	Possible research topics	Possible areas of inquiry
Deriving knowledge, decisions, and actions	Epistemological concerns	• What implications does big data have for the traditional deductive scientific research model? • Is an alternative big data-driven “inductive mode of science” in development or even feasible in the IS discipline? If so, how can we ensure the validity of such knowledge? If not, how can one use the inductive mode in conjunction with the traditional deductive model? • What is the role of prediction versus explanation in big data research? • How must one adapt traditional research methodologies/methods (qualitative and quantitative) to investigate phenomena of interest in big data environments?
Deriving knowledge	Computational social science	• How can longitudinal big data channels, such as social media, mobile location, and Web clickstreams, enhance our understanding and explanation of user behaviors? • How can sentiments and affects appearing in user-generated content, such as online word-of-mouth, inform our understanding of user behaviors and intentions? • What can large online social networks reveal about patterns of influence and/or information propagation? • What new insights can work logs and other unstructured sources reveal about relationships between employee actions and employee productivity, satisfaction, and/or customer-oriented outcomes? • What are the threats to validity of knowledge computationally derived, and what are the ways to mitigate these threats? • What is the nature of theory or theorizing that is consistent with this form of research (i.e., computational social science)?
	Privacy and security concerns	• What are the principles by which one can manage invasiveness/infringement of privacy in business enterprises, in academia, and in wider society? • How can big data contribute to a better understanding (and resolution) of the privacy paradox in which, on one hand, users desire personalization, innovative technologies, and novel communication channels, and, on the other hand, seek privacy and anonymity? • Since informed consent is often seen as ineffective, what other controls/policies need to be put in place? How can we assess and ensure their effectiveness?
	Other ethical considerations	• What implications does big data have for consistency, equality, accountability, integrity and an all-around conscientiousness? • What is the impact of sources and collection methods on data credibility? • How might access to information and knowledge attained from big data affect employee satisfaction and performance? • What should the organizational norms be for access and transfer of knowledge attained through big data analytics? • What would be the key elements of an ethical code for organizations and analysts/data scientists using big data and big data analytics?
Decisions and actions	Nature of decision making	• How do organizations/individuals/groups actually make decisions in the big data environment? • To what extent do traditional decision making models hold in the new environment?

Table 1. Big Data and Behavioral Research: Sample Research Opportunities

	Leadership	• How does analytical leadership emerge? • What are the characteristics of analytical executives? • What role does c-level leadership play in firms' abilities to leverage and "compete on big data analytics?"
	Organizational culture and governance	• How do organizations transform from an intuition-based decision making culture to a data-driven decision making culture? • What is the role of IT departments in supporting big data analytics? • What big data technology architectures and configurations (or analytics techniques) fit with different types of organizational cultures (and governance)?
	Cognition and usability	• What are the capabilities and constraints of big data information sources in supporting cognition and decision making? • As dashboard-based visualizations become the norm for managers, what are the implications of the four Vs on users' cognitive load and decision making performance?
	Trust and big data versus intuition	• What is the ideal balance between data and intuition? • What does trust in big data mean? What role does trust (e.g., trust in big data, people, and processes) play in adopting and effectively using big data?
	Adoption and adaptation of big data techniques and technologies	• How do the four Vs impact user perceptions and intentions to use big data IT artifacts? • What are the key personality traits of good analysts and how do these traits impact data and technology usage and performance?
	Big data outcomes	• How do we assess big data initiatives when considering the perspectives of relevant stakeholders and the potential instrumental and humanistic outcomes?

7 A Big Data Research Agenda for Design Science Research

7.1 Paradigmatic Considerations for Design Science Research on Big Data

We begin by discussing broader paradigmatic considerations of big data on design science; namely, 1) the effects of focusing on IT artifacts that *emphasize information more than systems/technology* and 2) implications of big data analytics artifacts on kernel design theories. Design is a product and a process (Walls, Widmeyer, & El Sawy, 1992; Hevner, March, Park, & Ram, 2004). The design product is a construct, model, method, and/or instantiation. The design process involves an iterative cycle of "test and learn" or "build and learn, evaluate and learn" (Simon, 1996; Nunamaker, 1992). In big data analytics, the process of deriving knowledge and insights from data is, in some ways, analogous to the design process. The most prevalent process for guiding analytics is the cross-industry standard process for data mining (CRISP-DM) (see Figure 6). CRISP-DM involves iteratively performing several phases 1) problem/business understanding, 2) data understanding, 3) data preparation, 4) modeling, 5) evaluation, and 6) deployment (Chapman et al., 2000). The first phase emphasizes the importance of tackling significant problems/opportunities and identifying data mining goals/success criteria. The second phase involves an inventory of available data sources, including assessing data quality. The third phase involves selecting appropriate sources and specific variables, and cleaning and reformatting the data. The fourth phase includes using predictive, descriptive, or prescriptive analytics method to analyze the data (Chandler et al., 2011). The fifth phase involves evaluating models and their connection to problem/business outcomes. Finally, the sixth phase involves analyzing the artifact in the field outside the lab/production environment.

CRISP-DM is to data mining what the system development lifecycle (SDLC) was to traditional information systems development, with CRISP-DM similarly appearing in the introduction section of various data mining and business analytics textbooks. The commonalities between CRISP-DM and the traditional system design process (Simon, 1996; Nunamaker, 1992) presents great potential for design science research geared towards producing novel IT artifacts capable of deriving knowledge and insights from big data. For instance, following the CRISP-DM design process, new constructs, models, methods, and instantiations could enhance BI dashboards or predictive, descriptive, prescriptive, and/or diagnostic model-based technologies (Shmueli & Koppius, 2011; Chandler et al., 2011). Recently, we have already begun to see research in this vein with new predictive IT artifacts developed using the design science tradition that espouses informal connections to elements of CRISP-DM (Abbasi, Zhang, Zimbra, Chen, & Nunamaker, 2010).

Comparing and contrasting CRISP-DM with SDLC is also interesting for another reason. One can consider big data artifacts as a shift in the design science tradition toward more significantly emphasizing artifacts that support information relative to systems or technology. For instance, only one of the stages depicted in CRISP-DM pertains to actually deploying the artifact, whereas at least four stages relate to processing, modeling, and evaluating data/information. CRISP-DM barely emphasizes the key stages of SDLC: users' system-related requirements, post-deployment system implementation, and system usage/evaluation. While the implications of this shift are not necessarily epistemological in nature, it does introduce paradigmatic considerations for design science. For instance, what is the appropriate balance between information and systems/technology in design research geared toward big data? Furthermore, how does research on the design of information-centric big data IT artifacts relate design science (in IS) to information and computer science? Our discipline has a clear need to have conversations on this issue.

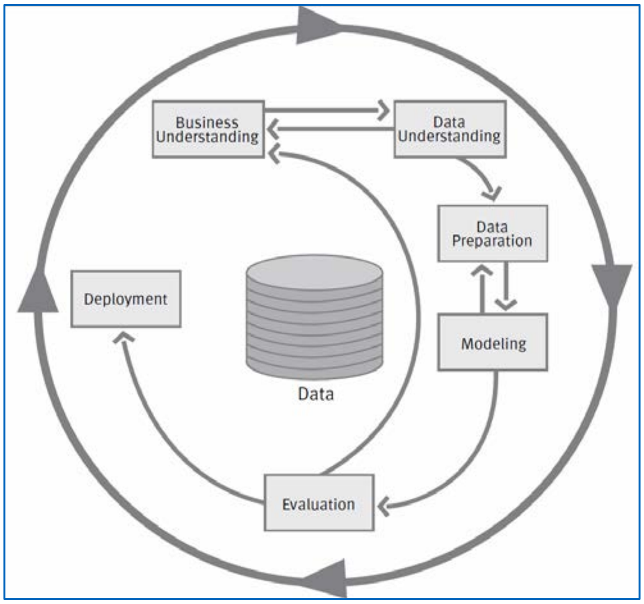

Figure 6. The CRISP-DM Analytics Process Embodies Similar Intuitions to the Iterative Design Process Advocated in the IS Design Science Tradition

For predictive analytics, big data has also had an impact on the kernel theories providing design guidelines. Classical theories of statistics such as the information theory and Bayes theorem embody simple yet powerful knowledge that have guided the design and development of some of the commonly used predictive and descriptive analytics methods traditionally employed in industry and academia. However, with the growth of big data in recent years, the four Vs introduce various challenges for data mining methods grounded in these traditional statistical theories. The volume of data presents computational constraints. The sparsity and structural nuances of new varieties of data sources, such as text and multimedia, poses representational richness issues. Some have also found traditional statistical methods to be susceptible to veracity concerns.

Data-mining methods grounded in the statistical learning theory overcome many of these limitations (Vapnik, 1998) and attain unprecedented results for tasks such as image recognition, text mining, phishing detection, and fraud classification. Consequently, despite appearing in the past 15 to 20 years, Vapnik's research on the statistical learning theory has already garnered over 150,000 citations according to Google Scholar, with more than half of those citations appearing in the past five years. One can largely attribute the success of statistical learning theory-based methods to their having a strong theoretical foundation and robust analytical power that is well suited for big data. However, other big data analytics methods, such as deep learning (LeCun, Bengio, & Hinton, 2015), do not have significant theoretical underpinnings. Indeed, critics have argued that they provide power without the underlying statistical theory (Gomes, 2014). As design research continues to explore the dichotomy between prediction and explanation (Shmueli & Koppius, 2010), the precise role of kernel theories in big data IT artifacts will remain a central question. A key question is: should IS research adopt an "ends-justify-the-means" perspective in which predictive power trumps methodological transparency and explanatory potential? We discuss the broader “prediction versus explanation” issue in greater depth in Section 11.

7.2 Design Science Research on Deriving Knowledge from Big Data

Setting aside the aforementioned paradigmatic considerations of designing big data IT artifacts, design science has much to offer in the burgeoning realm of predictive analytics, including novel constructs, models, methods, and instantiations leveraging big data. Scholars in the IS community have also applied predictive analytics at both “micro” and “macro” levels of granularity (Brown, Abbasi, & Lau, 2015). Chang et al. (2014) refer to this range of data granularities as the micro-meso-macro data spectrum. One of the most exciting opportunities pertains to predicting/analyzing micro-level outcomes (Agarwal & Dhar, 2014). Specifically, Brown et al. (2015, p. 6) note:

> *Micro-level predictive analytics involves making inferences about future or unknown outcomes pertaining to individual firms, people, or instances. Micro-level prediction contains greater intra-entity information, including perceptual constructs, [transactions/logs on] individual behavior, and spatial-temporal indicators.*

We have already begun to see some exciting, cutting-edge examples of micro-level prediction. For example, Wang and Ram (2015) predicted individuals’ sequential purchase patterns using spatial, temporal, and social relationship features on a test bed encompassing three million transactions initiated by thirteen thousand customers from nearly three hundred locations. This and other related studies highlight the “art of the possible” and the “art of the valuable” for IS research on novel predictive design artifacts. Future IS research could leverage objective (e.g., observed transactions and logs) and perceptual (e.g., survey, sentiment, voice transcript, and interview) data in conjunction with various intermediate decisions and actions to predict individuals’ behaviors with applications in marketing, e-commerce, security, health, and finance (Abbasi, Lau, & Brown, 2015).

There is no doubt that user-generated content sources such as social media have enhanced our understanding of various micro and macro-level phenomena in recent years (Chen et al., 2012), which presents great opportunity for developing social media analytics artifacts for collecting, monitoring, analyzing, summarizing, and visualizing social media data (Zeng et al., 2011). However, a major challenge remains in ensuring high veracity of such data sources. As Zeng et al. (2011, p. 14) note: “issues such as semantic inconsistency, conflicting evidence, lack of structure, inaccuracies, and difficulty in integrating different kinds of signals abound in social media”. We need IT design artifacts capable of identifying, quantifying, accounting for, and alleviating veracity concerns in information sources such as social media by assessing key information quality dimensions such as usefulness, relevance, and credibility. Examples of preliminary research in this vein include work pertaining to online spam and deception detection (Zhang et al., 2014). Such artifacts can be potentially beneficial in various application areas including marketing, finance, public policy, and health (Zeng et al., 2011; Abbasi & Adjeroh, 2014).

Big data presents numerous opportunities for new design-oriented work pertaining to the earlier stages of the value chain; namely, data and information. There is a long-standing tradition of design science work on modeling formalisms and ontologies (Wand & Weber, 2002). New forms of user-generated content present opportunities to enrich existing ontologies and develop new ones and to introduce new conceptual models and grammars. A related area involves extending classification principles from conceptual modeling to modeling of information categories in big data, which forms the basis for many forms of predictive and descriptive analytics (Parsons & Wand, 2013). Embley and Liddle (2013) expect conceptual modeling to address big data challenges by adopting the perspective that the design activity related to big data is fundamentally about structuring information. Conceptual modeling research should make big data’s volume searchable, harness the variety uniformly, mitigate the velocity with automation, and check the veracity with application constraints. The IS research community needs to direct some attention to the tasks of investigating, addressing, and/or exploiting big data’s four Vs jointly and effectively.

Similarly, database design and data integration have been a major area of focus for prior works (Storey et al., 1997), and there are opportunities for research on the integration of a variety of structured and unstructured data sources available in organizational settings. As we highlight earlier when discussing the big data information value chain, the four Vs have increased the complexity of managing, storing, and integrating data. The challenge remains in investigating and establishing an acceptable architecture to integrate, manage, and implement both structured and unstructured data under one unified platform. The traditional relational data model and the corresponding relational database management systems (RDBMSs) cannot meet the heterogeneity (variety) challenge of big data. Many consider NoSQL as the potential data management solution for big data, but many architectural alternatives exist that range from Hadoop/Spark to Hadoop and RDBMS in parallel to Hadoop (for unstructured data) inputting into RDBMS (Heudecker, 2013; Buytendijk, 2014). We need research to examine the feasibility, fit, and business value of such alternatives and to provide guidelines for big data architectures based on organizational and industry-level contexts.

For design research pertaining to knowledge derivation and representation, big data presents both advantages (i.e., volume and variety) and disadvantages (i.e., velocity and veracity). With the large volume and a variety of data sources, big data can certainly enhance ontology learning by automatically deriving domain knowledge by mining unstructured and semi-structured data (Buitelaar, Cimiano, & Magini, 2005). For example, user-generated content contributed freely in social media presents opportunities to enrich existing ontologies and develop new ones and to introduce new conceptual models and grammars. As we allude to earlier in this section, some IS scholars have recently suggested that domain ontologies may play a critical role in the conceptual model for managing and implementing big data (Embley & Liddle, 2013).

7.3 Design Science Research on Supporting Decisions and Actions from Big Data

Scholars have long used design science to guide the design and development of various decision support systems, including group support systems, recommender systems, personalization, contextualization, and collaboration technologies (Nunamaker et al., 1991; Adomavicius & Tuzhilin, 2005; Arnott & Pervan, 2012). They have also used it to develop BI-related DSS artifacts (Chung, Chen, & Nunamaker, 2005). More recently, Lau et al. (2012) developed ABIMA, a big data business intelligence DSS for mergers and acquisitions. ABIMA integrates large volumes of financial metrics derived from structured databases with unstructured sources, such as financial news articles, search engine results, and documents crawled from the Web. The system, practitioners specializing in mergers and acquisitions evaluated, is the type of research that epitomizes how one can use design science for research on big data DSS. Other types of big data IT artifacts that support the decision making process could be ones designed to support real-time decision making that possibly incorporate user feedback-based or system-generated credibility assessments of underlying information sources (Jensen, Lowry, Burgoon, & Nunamaker, 2010; Jensen, Averbeck, Zhang, & Wright, 2013).

Given that big data analytics significantly emphasizes enhancing business processes (Davenport, 2006), business process improvement driven by big data constitutes an important research area (Baesens et al., 2014). Potential avenues include developing automated artifacts for discovering and optimizing processes and employing analytics-driven methods in various internal-facing and external-facing business processes that range from operations and human resources to customer relationship management (Davenport & Harris, 2007). As an extreme example of automation, the use of big data analytics to replace human involvement from certain business processes has already begun to take shape (Davenport & Kirby, 2015); in Section 4, we mention real-time analytics pipelines that are replacing traditional business processes. However, in many contexts, big data analytics provides complementary “augmentation” to human-driven processes (Davenport & Kirby, 2015). Augmentation and automation signify a departure from the traditional human-centered computing paradigm toward autonomous computing albeit with varying degrees of separation depending on the level of human involvement. Nevertheless, this shift necessitates reconsidering guidelines for the design product and design process associated with such artifacts (e.g., requirements gathering in contexts where there are no users). When designing such artifacts, what role might theories and principles from, for example, the cognitive computing and artificial intelligence literature play?

Given the dynamic and nascent technological and organizational environments related to big data, it would be interesting to see if (and how) one could productively use the action design research (ADR) method to develop robust big data IT artifacts (Sein et al., 2011). For instance, action design research incorporates provisions for varying levels of end user involvement that could provide the necessary flexibility for designing big data artifacts in contexts ranging from traditional user-centered decision support to business process augmentation/automation.

7.4 Summary of Design Science Research on Big Data

Big data presents several *wicked* problems: how should IS researchers balance a big data-oriented design science research agenda with those being pursued in reference disciplines such as engineering, statistics, and computer science? Goes (2014, p. iv) touches on this challenge by noting that at least five different departments at his institution were, in some way, explicitly related to big data research. Goes cautions us by noting that, without guidelines for shaping the IS big data research agenda, "each unit can contribute to the big data paradigm, but at present the approach resembles that well-known cartoon of making sense of an elephant by grabbing isolated parts of the animal". In our assessment, we can say that, relative to other disciplines, IS design science researchers are uniquely positioned to provide the appropriate mix of rigor along with humanistic and instrumental relevance. Further, our research often seeks to offer generalizable design principles and guidelines abstracted from the development of contextualized big data IT artifacts that can potentially help address other important problems. The sample research opportunities in Table 2 reflect this view. We believe the design science perspective on big data analytics represents an important future area of emphasis for IS research.

Table 2. Big Data and Design Science Research: Sample Research Opportunities

Value chain stage(s)	Possible research topics	Possible areas of inquiry
Deriving knowledge, decisions, and actions	Paradigmatic considerations	• What is the appropriate balance between information and systems/technology in design research geared toward big data in the IS discipline? • What implications does big data IT artifacts' potential shift in focus from systems to information have for the design process? • How might the characteristics of big data affect the nature of kernel design theories that are potentially useful? • What is the IS "signature" for big data design research (i.e., what is the scope/nature of big data IS design artifacts relative to reference disciplines such as computer science, marketing, engineering, and statistics)?
Deriving knowledge	Novel artifacts for prediction or description	• How can one leverage the volume and variety of big data to develop novel artifacts for predicting/describing macro versus individual/micro-level phenomena or events? • How can design science research build novel artifacts for deriving knowledge from big data sources, such as user-generated content, to advance research in other disciplines, including marketing, finance, and health? • How can design guidelines of big data analytics artifacts better compensate for the veracity of input data? What novel veracity-assessment artifacts can we develop to shed light on information relevance, usefulness, and credibility?
	Modeling formalisms and integration artifacts	• Can new forms of user-generated content enrich existing ontologies, enable the development of new ones, and introduce new conceptual models and grammars? • What is the potential for extending classification principles from conceptual modeling to modeling of information categories in big data? • Can conceptual modeling address some of the challenges of big data by making the volume searchable, harnessing the variety uniformly, mitigating the velocity with automation, and/or checking veracity with application constraints? • How can design science inform the state-of-the-art integration, management, and implementation of organizational big data initiatives in light of the four V challenges? • What design theories do we need to guide big data architectures based on organizational and industry-level contexts?
Decisions and actions	Novel IT artifacts for decision support	• How can IS contribute guidelines for design artifacts that support real-time decision making from big data? • What is the role of credibility assessment as a design guideline in big data decision support systems?
	Business process improvements and automation	• What is the potential for automated process discovery and optimization? • How can big data analytics improve business processes? • Are existing design theories sufficient for real-time big data analytics environments in which run-time and autonomy considerations create nuanced design requirements? What are the alternative theories that may be valuable?
	Big data action design research	• Given the emerging nature of big data, how can we use approaches such as action design research (ADR) to guide the development and harnessing of big data IT artifacts in organizational settings?

8 A Big Data Research Agenda for the Economics of IS

8.1 Epistemological Concerns for Big Data and the Economics of IS

The economics of big data has important implications for information systems. Just as scholars once used the "economics of information" to describe the value of information asymmetries in marketplaces (Stigler 1961), now, with firms competing on analytics, access to information that can enable enhanced analytical capabilities and insights facilitating differentiation has ushered a new era of "knowledge is power". Quantifying this power is critical.

Beyond some of the issues discussed in Section 6.1, the epistemological implications for the economics of IS community primarily relate to research methodology. These include the deflated p-value problem (Lin, Lucas, & Shmueli, 2013), increasing emphasis on prediction versus explanation (Shmueli & Koppius, 2010), and construct validity when using unstructured and log-based data sets; however, given that many of these issues are applicable to multiple IS research traditions, we discuss them in greater detail in Section 11.

8.2 Economics of IS Research on Deriving Knowledge from Big Data

The value of information has been a longstanding area of inquiry in the economics of IS tradition (Banker & Kauffman, 2004). In the context of big data, assessing information's value is more critical than ever. One research direction analyzes the relative value contributions of the four Vs (e.g., value of data volume and variety) for deriving knowledge from big data. As organizations treat data as an asset (and, in many Web 2.0 business models, as the primary asset), quantifying its value has become a major discussion topic both from a broad business value perspective (which includes the implications for third party data brokers and data markets) and from a more traditional accounting perspective. This emphasis on data as an asset has spurred *infonomics*: the theory, study, and discipline of assigning economic significance to information. For instance, a recent McKinsey report states that public data sources pertaining to education, energy, healthcare, transportation, and consumer finance (collectively dubbed "open data") have the potential to create USD$3 trillion in annual business and/or societal value (Manyika et al., 2013).

At a more micro level, individual organizations are routinely interested in identifying the most useful public/private data sources and quantifying the precise value of these sources. For example, Bardhan, Oh, Zheng, and Kirksey (2015) used demographic, clinical, and administrative data from 67 hospitals in northern Texas gathered over a four-year period to build models capable of predicting and describing congestive heart failure patient readmissions. Their model demonstrates the importance of health IT-related variables, which prior or baseline models have not considered impactful but that could help hospitals save millions of dollars by avoiding costly readmission-related penalties. Many other recent studies also suggest that using more data instances and variables can improve predictive capabilities (Junque de Fortuny et al., 2013). These findings raise questions regarding complexity, model management, and cost-benefit tradeoffs. Researchers wonder: if bigger is better, how much is too much? As Junque de Fortuny et al. (2013, p. 219) ask, "is it worth undertaking the investment to collect, curate, and model from larger data sets?"

Prior to big data's rise, firms were beginning to derive the last iota of predictive power from a set of data they had access to often by building predictive models that were increasingly complex. Netflix is one example. In 2006, Netflix offered a US$1 million prize to any team that could improve their existing movie recommendation models by 10 percent. The competition concluded on July 26, 2009. Over the final two years, the performance of the best predictive models improved by less than 2 percent, whereas the complexity of the solutions increased dramatically. The final winning model from a team appropriately named BellKor's Pragmatic Chaos blended results from hundreds of underlying base models.

Similarly, scholars commonly describe loan risk-assessment models at financial services firms developed in the past two decades, which largely leverage the same structured data sources, as complex, variegated black-box arrangements of models on top of models delicately combined (Derman, 2011). Former Goldman Sach's Lead Quantitative Analyst Emmanuel Derman highlights the pitfalls of complexity and poor model management over time in his book *Models Behaving Badly*. He describes the role played by complex models in the 2008 financial crisis (Derman, 2011); the narrative reminds readers of the classic saying "complexity is death". Big data creates an opportunity to not only enhance these models' analytical capabilities but also reduce the inherent risks associated with using them. However, not all information sources are created equal. A key challenge facing organizations is finding a way to quantify the value of information that considers both insightfulness and risks.

Research on pricing for data sets/data sources in the booming data broker markets can help firms make more-informed decisions in the data marketplace. Here, all of the standard services management questions apply, including those that Rai and Sambamurthy (2006) articulate. What are the best strategies for bundling of data? Should firms pursue flat or usage-based monthly service rates versus one-time sales? How do alternative service rate plans impact data usage? Big data's impact on data-based marketing and pricing, omni-channel marketing (Song, Sahoo, Srinivasan, & Chrysanthos, 2014), and attribution are other potential areas of inquiry. The value of information also raises questions about intellectual property rights, especially in the context of user-generated content.

The cost of veracity can potentially offset the value of data volume and variety. The existing body of knowledge on data warehousing and business intelligence, which one can consider a close predecessor to big data, has emphasized the importance of data quality as an antecedent for the success of data warehousing initiatives (Wixom & Watson, 2001). Hence, quantifying the adverse effects of incomplete or inaccurate information is essential yet challenging for mitigating risk in the era of big data analytics. Such research could connect data quality to the effectiveness of business outcomes.

The analysis of social media has garnered considerable attention from the economics of IS community in recent years with many outstanding avenues of inquiry. As organizations move towards the "social-ecosystem" encompassing the use of social media for various employees and customer-oriented activities, we can ask how firms can leverage social media for internal communication and collaboration, external engagement, and listening/ideation (Zeng, Chen, & Lusch, 2010). What is the value of insights, engagement, and internal communication usage through social technologies? The interplay between social media channels and marketing effectiveness is another important area receiving considerable attention in the economics of IS community (Song et al., 2014). In that vein, questions include: what are the key factors impacting social media marketing effectiveness? How does peer influence impact social media marketing? What is the role of social media in viral marketing? A recent related stream of work examined the usefulness of location and geographic information for the analysis of choice, price, competition, and mobile marketing.

8.3 Economics of IS Research on Implications of Big Data for Decisions and Actions

The economics of big data analytics extend further down the information value chain beyond knowledge acquisition to decisions, actions, and their ensuing consequences. In his book entitled *The Value of Business Analytics*, acclaimed business analytics guru Evan Stubbs talks about the challenges business analysts and data scientists face when attempting to quantify the value of an analytics project or portfolio of initiatives (Stubbs, 2011). For example, when should firms invest in big data, and what are the potential returns? As one answer to this question, Tambe (2014) found that firms with significant existing data sets who invested in Hadoop were associated with 3 percent faster productivity growth. A related question is: what implications does big data's ushering in the increased usage of cloud-based SaaS and DBaaS have for platform economics and strategy? Research on quantifying the business value of big data analytics, including the implications of the four Vs, could inform the existing body of knowledge. Potential research directions include work measuring the return on investment for big data technologies, the impact of data variety and veracity on quality of decision making, and the economics of real-time decisions using big data.

8.4 Summary of Economics of IS Research on Big Data

In some ways, the economics of IS tradition is ideally suited to tackle problems pertaining to deriving knowledge from big data. In particular, beyond certain methodological adjustments (see Section 11), the empirically driven, inductive reasoning-oriented applied econometrics approach is well aligned to addressing these types of questions. The economics of IS also has an essential role to play in examining the economic value of big data insights and big data analytics-driven decision making. Table 3 summarizes some of the important research opportunities.

Table 3. Big Data and Economics of IS: Sample Research Opportunities

Value chain stage(s)	Possible research topics	Possible areas of inquiry
Deriving knowledge, decisions, and actions	Epistemological and/or methodological concerns	• How must traditional research methods be adapted to investigate big data environments? • What is the role of prediction versus explanation in big data research? • How can we ensure the validity of constructs derived from noisy unstructured and log-based data sources?
Deriving knowledge	Value of data, volume, and variety	• What is the value of various data sources and channels in terms of quality of insights, enabling new capabilities, and quantifiable business value? • Regarding the impact of data volume on insights and business value, how much is enough and how much is too much? • What is the value of data volume and variety from a risk management perspective? • As data becomes an asset, what role can third party data brokers and data markets play? What are the pricing and market structure implications? • As firms monetize user-generated content, what are the implications for intellectual property? • How can the volume and variety of data inform data-based marketing and pricing? • What are the benefits and challenges of data variety for omni-channel marketing analysis and attribution?
	Cost of veracity	• How can we quantify the business impact of low veracity data? • Which types of data quality issues are the most impactful?
	Social media and economics of IS	• What is the role of social media in enterprises for internal communication, customer engagement, and listening/ideation? • Regarding the economics of social media, what is the value of insights, engagement, and internal communication usage through social technologies? • What are the key factors that impact social media marketing effectiveness? • How does peer influence impact social media marketing? • What is the role of social media and incentive schemes in viral marketing?
	Impact of location and geography	• How can location and geographic information impact research on choice, pricing, and competition? • What are the implications of geo-targeting in mobile marketing?
Decisions and actions	Quantifying value and impact of four Vs on decision making	• What is the impact of data variety and veracity on the quality of decision making? • What are the key factors influencing business value in the context of real-time decision making using big data?
	Value of big data IT artifacts	• How do we measure the return on investment for big data technologies? • When should firms invest in big data, and what are the potential returns? • As big data ushers in increased usage of cloud-based SaaS and DBaaS, what are the implications for platform economics and strategy?

9 Cross-tradition Research on Big Data

There are many opportunities for research at the cross-sections of behavior, design, and economics of IS. In some respects, big data's scale and complexity afford and encourage cross-tradition research projects. The work on human-computer systems design has traditionally been at the intersection of design science and behavior, such as cognitive psychology and decision science (Banker & Kaufmann, 2004). In this vein, one obvious direction is to design and develop big data IT artifacts that researchers may subsequently evaluate in terms of their positive impact on behavior (Zahedi, Abbasi, & Chen, 2015). Another connection between the design and behavioral traditions relates to research on IS strategy. Big data analytics ultimately focuses on improving business processes (Davenport & Harris, 2007) to attain a strategic competitive advantage, a perspective that is highly congruent with the resource-based view of the firm. Designing enterprise-wide big data analytics in a manner that maximizes the potential for competitive advantage in different types of industries and for different organizational cultures and governance archetypes is a potentially valuable direction. Related to this area, knowing how to align alternative big data architectures (such as ones based on Hadoop, traditional RDBMS, or hybrid models) with business strategy and a firm's data environment and understanding important criteria and success factors in the decision making process remain important issues. Researchers can use established approaches in IS such as case studies, laboratory experiments, and survey studies in investigating such topics.

Design science and the economics of IS also appear to have potential in the context of big data research. One example is cost-sensitive classification or regression in which one quantifies the values of true and false positives/negatives and incorporates them into the design of predictive artifacts (Bansal, Sinha, & Zhao, 2008; Zhao, Sinha, & Bansal, 2011). Here, existing work has mostly focused on the predictive artifact and less on the methodology for deriving cost matrix values. The value of information in big data IT artifacts represents another important area at the cross-section of design and economics of IS. Recently, many studies have designed novel IT artifacts focused on mining big data sources as decision-support aids (e.g., Lau, Liao, Wong, & Chiu, 2012). It remains unclear what the business value of such artifacts truly is with respect to key business performance indicators. Another already potent research area at the cross-section of design and economics of IS pertains to optimization (Banker & Kaufmann, 2004). Here, big data's variety presents opportunities. For instance, online product reviews could possibly enrich product design optimization (e.g., Balakrishnan & Jacob, 1996). Similarly, consumer sentiments and demand forecasts based on user-generated content could enhance pricing optimization; however, in both of these examples, the tension between the value and veracity of social media and other user-generated data sources could present an interesting dynamic, worthy of an in-depth inquiry.

Big data presents numerous opportunities at the intersection of economics of IS and behavior as well. For instance, behavioral economics is a well-established area (Tversky & Kahneman, 1974). In the context of big data, research could examine managers' propensities to combine data of varying quality/credibility with, for instance, intuition in real-time versus non-real-time settings. Such work would borrow from theories in economics, cognitive psychology, decision making, and risk management (Banker & Kaufmann, 2004).

10 Big Data: Implications for Theory

Many scholars have reflected on big data's possible implications on theory. One point of view, albeit extreme, is that big data renders the role of theory—sometimes seen as fictional and value-laden—unnecessary and obsolete, and replaces it with patterns derived directly from data that reflect nothing but the truth (e.g., Kitchin, 2014b). On the other hand, many scholars argue that, in the absence of theory, data lacks "order, sense and meaning" and that "theories without data are empty; data without theories are blind" (Harrington, 2005, p. 5, cited in Sarker et al., 2013, p. xiiii). While this debate is likely to continue without immediate resolution, we do not foresee theories disappearing or diminishing in importance because of research using big data. To the contrary, we foresee that some of the theories will become more robust because "researchers now have a medium for theory development through massive experimentation in the social, health, urban, and other sciences" (Agarwal & Dhar, 2014, p. 444). This is in part due to easier data collection and enhanced control and precision, realism, and generality associated with big data (Chang et al., 2014). At a broader level, we believe that scholars have the opportunity to reflect on the changing nature of the theorizing process and on the characteristics of theories developed in a data-abundant environment.

From an information value chain perspective, some recent big data IS studies and editorials have touched on the role of theory when leveraging big data sources for discovering knowledge. As previously mentioned, we believe that, in addition to big data "information sources", big data's characteristics embodied in the four Vs afford important opportunities for research, that can both borrow from novel theories and contribute to existing theories, related to deriving knowledge as well as decisions and actions. Below, we outline a few examples.

The impact of data variety and velocity on problem-solving accuracy and time constitutes an important research topic. The cognitive fit theory (CFT) provides an excellent and robust theoretical lens for examining this topic. Earlier work on CFT examined the importance of congruence between problem task and problem representation on users' mental representation and overall problem-solving performance (Vessey, 1990). While initial studies focused on tables versus graphs, subsequent work extended CFT to other specialized representations such as maps (Dennis & Carte, 1998), and considered the impact of users' prior domain knowledge and the effect of subtasks (Shaft & Vessey, 2006). All of these findings have important implications when examining the impact of big data characteristics for problem solving in general, and specifically in the context of dashboards depicting a variety of information in real-time. On the other hand, big data characteristics, such as variety and velocity, can also potentially offer theoretical extensions. Data visualization dashboards often incorporate multiple tabs with coordinated views depicting real-time data (Andrienko & Andrienko, 2003). The effects of problem solving in such multi-representation, multi-subtask, real-time situations remain unclear, and this offers great potential to contribute to theory. CFT, with suitable adaptations, can also inform the design/construction of novel user interface artifacts (Vance, Lowry, & Egget, 2015) for presenting big data.

Organizations using big data routinely ask managers and analysts to monitor and present key findings using reporting tools that integrate traditional structured data sources with novel social listening, web clickstream, sensor-based, and open data. Practitioner studies have suggested that analysts often do not perceive such tools to be useful, with obvious implications for the business value of such artifacts (Kaushik, 2011). Adoption models represent an excellent theoretical lens for examining the impact of perceived usefulness and ease of use on behavioral intention to use such reporting tools and actual use (Davis, 1986, 1989). The effects of mandatory versus voluntary reporting and trust are also important considerations (Brown, Massey, Montoya-Weiss, & Burkman, 2002; Gefen, Karahanna, & Straub, 2003). In turn, big data's four Vs could provide important insights that can inform the extensive body of knowledge pertaining to technology adoption (Venkatesh, Morris, Davis, & Davis, 2003; Venkatesh, Thong, & Xu, forthcoming). For instance, the variety of data could have potentially contrasting effects on users' perceptions of usefulness and ease of use, which users' levels of experience may moderate. Further, as data veracity becomes increasingly relevant to big data IT artifacts, the implications for trust (in both the data and the artifact) and eventually for behavioral intention to use also present interesting issues to investigate.

Chaos theory studies the behavior of dynamic systems that are highly sensitive to initial conditions, where small changes in initial conditions can yield widely diverging outcomes (Gleick, 1987; Sprott, 2003; Werndl 2009). Prior studies have already discussed the potential of big data for theory development via computational social science or massive experimentation (Agarwal & Dhar, 2014; Chang et al., 2014). Chaos theory could be beneficial in macro-level computational social science research, since it "appears to provide a means for understanding and examining many of the uncertainties, nonlinearities, and unpredictable aspects of social systems behavior" (Kiel & Elliott, 1996). Furthermore, as IS design science research explores novel predictive artifacts utilizing big data, chaos becomes an important consideration. The inclusion of big data should make predictive artifacts more accurate, stable, and valuable. However, for complex event forecasting in situations where chaos is present, prediction can be problematic since model assumptions, which are typically based on probabilities of various patterns (i.e., connections between observed initial conditions and eventual observed outcomes), may not hold true (Sprott 2003). Consequently, seemingly small errors in initial condition prediction probabilities can result in large errors between longer-term forecasts and actual outcomes (Werndl, 2009). We are already beginning to see Chaos theory concepts incorporated in problems such as weather forecasting and traffic prediction, where, in turn, the results are informing our understanding of chaos in these application areas.

Somewhat related to chaos is black swan theory, which focuses on highly improbable or surprising, high-impact events that are often incorrectly rationalized in hindsight (Taleb, 2005, 2007). The key idea is that probability-centered analysis and thinking that diminishes the importance of outliers or the unobserved is problematic for appropriately managing the risks associated with black swan events. Taleb's views on the limitations of statistics and his seemingly negative portrayal of statisticians has raised several (possibly valid) rebuttals from the statistics community (e.g., Westfall & Hilbe, 2007). Nevertheless, his central tenet appears to have merit. As big data further creates the shift towards data-driven decision making, risk management pitfalls such as attempting to predict extreme events, overreliance on the past, psychological biases against less likely outcomes, and overemphasis on standard deviations are likely to be exacerbated (Taleb, Goldstein, & Spitznagel, 2009). Black swan events also have important implications for the design of process automation relying on big data. We have already seen automated loan risk assessment and algorithmic trading engines fail miserably due to such events (Taleb, 2007; Derman, 2011). Black Swan Theory could shed light on studies examining risk considerations in data-driven decision making where traditional decision theories may be inadequate. Similarly, it can inform the design of process automation in big data environments such that the unknown unknowns are given proper consideration.

In summary, big data has potentially important implications for theory. From an information value chain perspective, big data sources and associated IT artifacts have distinct implications for both knowledge acquisition and for decisions and actions (and related outcomes), which include system usage, performance, and satisfaction. The key nuances of big data artifacts stem from the four V characteristics. On one hand, these characteristics may simply inform well established IS theories by playing the role of antecedent constructs or of moderator/mediator variables. On the other hand, these characteristic can introduce complexity and risk in IT artifacts increasingly relying on big data, thereby opening up exciting new possibilities for utilization of theories that have seen relatively limited usage in IS.

Our final comment regarding theory and big data is while one cannot understate the role of "theory" in big data research, we do need to acknowledge that theory has different forms in different traditions of research, and, thus, as research community, we need to be open to different types of abstractions offered as theoretical contributions.

11 Big Data: Implications for Methodology

The characteristics of big data test beds have important implications for the norms of analyses. One significant implication that has recently garnered attention is the "deflated p-value" problem. In their *Academy of Management Journal* editorial, George et al. (2014) suggest that the statistical methods and metrics used to examine big data sets may need to incorporate alternative techniques from statistics, computer science, applied mathematics, and econometrics. They state (p. 323): "The typical statistical approach of relying on p values to establish the significance of a finding is unlikely to be effective because the immense volume of data means that almost everything is significant".

In addition to statistical significance and co-efficient signs, one may also need to consider effect sizes and variance when testing hypotheses on big data sets (Lin et al., 2013; George et al., 2014). As Chatfield (1995, p. 70) notes: "The question is not whether differences are 'significant' (they nearly always are in large samples), but whether they are interesting. Forget statistical significance, what is the practical significance of the results?". To quantify the extensiveness of the problem, Lin et al. (2013) examined nearly 100 IS papers published between 2004 and 2010 with research test beds exceeding 10,000 instances and concluded that nearly half failed to discuss practical significance.

As we note in Section 7 and as George et al. (2014) and others allude to, analyzing big data often requires using computer science-based methods grounded in machine learning and artificial intelligence rather than statistics. For instance, genetic algorithms are a computationally effective non-deterministic heuristic method for searching an NP-hard problem's solution space and researchers/analysts have used them for variable feature selection in many predictive analytics problems involving thousands of input variables. Similarly, deep learning methods have enabled neural networks to attain impressive classification accuracies on large data sets (LeCun et al., 2015). Deep learning methods add additional layers of processes capable of learning or representing complex patterns at the expense of further degrees of separation between the model output and the underlying model intuition. Consequently, such methods also constitute a departure from traditional statistical methods, such as ordinary least squares or simple logistic regression analyses that prior IS studies have commonly used for both explaining and predicting. Shmueli and Koppius (2010) note that there is a difference between models geared toward prediction and those geared toward explanation. In many ways, big data amplifies this dichotomy as powerful non-deterministic and/or "black box" methods gain prominence for their predictive capabilities. In many cases, such methods produce answers to the "what" without the "why".

Another implication of big data sets is construct validity/credibility concerns pertaining to variables derived from user-generated, non-survey-based data sources, often characterized by low veracity. In addition to the spam and deception traits inherent to user-generated content sources (e.g., social media) and clickstreams' data-tracking limitations, scholars operationalize many unstructured data sources as a few structured variables. We can see one prominent illustration of this point in the context of user sentiment polarity: whether the user is expressing a positive, negative, or neutral sentiment toward a given topic. Due to the volume of big data, one typically derives such constructs using software packages that rely on natural language processing methods as opposed to traditional manual coding methods. There have been thousands of studies published using social media sentiments in the recent years, including several in IS outlets. Such studies routinely make conclusions about the impact of user sentiment-related independent variables. However, scholars rarely report information on the suitability and accuracy of the underlying sentiment classification models used to operationalize the constructs, which happens despite the fact that benchmarking studies have found that many state-of-the-art sentiment analysis methods' sentiment polarity classification performances are subpar, which affects the sentiment-related analysis and conclusions drawn from it (Hassan, Abbasi, & Zeng, 2014). Moving forward, we need research-based guidelines on how to validate variables derived from natural language, clickstream, sensor, and other big data sources.

Another important issue is to consider how the penetrative, imaginative understanding of human meanings discerned by qualitative (particularly interpretive) researchers using small data can *complement* (rather than be *substituted* by) patterns derived from big data using machines/techniques/algorithms, to be able to offer a more complete picture of the phenomenon. Indeed, an emerging stream of work illustrates how findings based on qualitative "idiographic" approaches may mutually inform findings based on computational methods (Gaskin, Berente, Lyytinen, & Yoo, 2014). We are also aware that grounded theory researchers in IS (i.e., the SIG GTM community) are looking for ways in which the grounded theory methodology (GTM) principles can be effectively utilized in big data settings.

Clearly, big data is requiring us to reexamine how we analyze and validate data and interpret and discuss the findings. We need further research to assess more thoroughly the pros and cons of different methods and metrics on various types of big data sets, and to provide meaningful guidelines. In particular, due to the volume, variety, and veracity dimensions, we need to be watchful about big data's creating "false positives" in terms of statistical significance of independent variables or considerably altering the effect size. Finding ways and principles that can aid in effectively complementing and/or triangulating big data research with small data research is another issue we must take seriously. Achieving "*consilience*—that is, convergence of evidence from multiple, independent, and unrelated sources"— needs to be a matter of priority (George et al., 2014, p. 324).

12 Closing Thoughts

The arena of big data/big data analytics has captured the attention and imagination of both practitioners and academics in a variety of disciplines, not just in IS. Commentators have described the big data phenomenon as a "deluge" and as having the potential to cause long-lasting impacts on practice and academia (e.g., Anderson, 2008; Kitchin, 2014b). Indeed, thought leaders and editors of leading IS journals see much reason for optimism regarding big data's impacts on the IS discipline. For example, Goes (2014, p. viii) has encouraged the field to "embrace the changes and provide leadership in the new environment… [for which we] are uniquely positioned" and to "claim our [rightful] territory". Agarwal and Dhar (2014) view big data as an opportunity for ushering in a "golden age for IS researchers". Yet, not all scholars from IS and other disciplines unquestionably accept projections of such promise (e.g., Buhl, Röglinger, Moser, & Heidemann, 2014). Some, including Professor Michael Jordan, a "machine-learning maestro", predict the onset of "big data winter" if we continue to over-promise and overhype (Gomes, 2014) without addressing fundamental epistemological and methodological issues associated with big data (e.g., Kitchin, 2014b). Furthermore, we must address the concerns regarding the erosion of privacy and, consequently, the loss of human dignity in the face of economic imperatives (e.g., Barocas & Nissenbaum, 2014) that can lead to "digital colonization" (Buhl et al., 2014) and the "subjugation" of human interests by machines and algorithms, which socio-technical scholars have for long been wary of (e.g., Bjorn-Andersen, Earl, Holst, & Mumford, 1982). In addition, the rhetoric of making academic research relevant by helping solve immediate organizational problems through big data without adequate abstraction or without designing approaches or artifacts for addressing broad classes of problems raises questions regarding how academic research differs from practice. Also, in line with Benbasat and Zmud (2003), many scholars believe that routinely engaging in big data projects without a unique disciplinary "signature" could prove to be ominous for the IS discipline in the long run.

Hence, our position is one of cautious optimism. While we undoubtedly see potential for big data in contributing to a stronger and more relevant IS discipline—one that would have a significant social impact—we do not take the benefits for granted. To help big data research achieve its potential, we invite IS scholars to: a) critically engage with fundamental issues, such as epistemology, methodology, ethics, and the design of novel artifacts; b) rethink decision models proposed in the era of scarce data and adapt them for use in the current era of abundant data; and c) assess economic and humanistic outcomes of big data in the form of systematic, multi-paradigmatic research initiatives across the information chain value. Further, to ensure a healthy development of scholarship in this area, we see the need to carefully balance "real-world" problem-solving using big data/big data techniques with reflective inquiry and scholarly abstraction of knowledge in this area. We also encourage big data researchers from the IS discipline to participate in boundary-spanning interdisciplinary big data projects, but to balance engagement with other disciplines with conscious development and nurturing of big data approaches and objectives that are somewhat consistent with the IS discipline's socio-technical heritage.

In this editorial, we do not intend to provide definite answers or directions but to encourage inquiry, reflection, and debate on this topic by IS scholars embedded in diverse theoretical and methodological traditions. We are hopeful that the framework (Figure 4) and the accompanying tables (Tables 1, 2, and 3), while preliminary, will help energize the conversation on big data in the broader IS community and provide a roadmap for advancing scholarship in the area.

Our final comment is related to teaching, which we believe is our raison d'être. No other academic unit has the diversity of research traditions and understanding of the business, information, technology, and human issues that are essential to comprehending the various facets of the big data value chain (e.g., Agarwal & Dhar, 2014). This diversity places us in an excellent position to offer pedagogical leadership in teaching, developing curricula, and programs and to initiate industry outreach centers (Chiang, Goes, & Stohr, 2012).

In summary, we are convinced that big data is here to stay. However, we can foresee a time when big data will not be at the forefront of our conversations, as we have seen in the cases of expert systems, BPR, e-business, ERP, and groupware. Yet, few will disagree that the ideas underlying these topics have continued, and will continue, to be important knowledge areas informing research and practice in IS. We expect that big data research will do the same. For now, big data offers a stage for learning new lessons, re-learning and refining old lessons, and reflecting on assumptions that underlie our research endeavors and the complex ways in which technology, information, and humans interact to shape the world we live in.

Acknowledgments

We thank Kenny Cheng, Dirk Hovorka, Steven Johnson, Vijay Khatri, Brent Kitchens, Sridhar Nerur, and Tony Vance for their constructive comments on an earlier version of the editorial. The reactions of TSWIM 2015 participants to a keynote talk by one of the authors also helped shape several ideas, for which we are grateful. Likewise, we greatly appreciate feedback from faculty and doctoral students that participated in the 2015 Antai Graduate Summer School at Shanghai Jiaotong University.

References

Abbasi, A., & Adjeroh, D. (2014). Social media analytics for smart health. *IEEE Intelligent Systems*, *29*(2), 60-64.

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., & Nunamaker, J. F., Jr. (2010). Detecting fake websites: The contribution of statistical learning theory. *MIS Quarterly*, *34*(3), 435-461.

Abbasi, A., Lau, R. Y. K., & Brown, D. E. (2015). Predicting Behavior. *IEEE Intelligent Systems*, *30*(3), 35-43.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734-749.

Agarwal, R., & Dhar, V. (2014). Editorial: Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, *25*(3), 443-448.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, *25*(1), 107-136.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired.* Retrieved from http://www.wired.com/2008/06/pb-theory/

Andrienko, N., & Andrienko, G. (2003). Informed spatial decisions through coordinated views. *Information Visualization*, *2*(4), 270-285.

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, *337*(6092), 337-341.

Arnott, D., & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems*, *44*(3), 657-672.

Arnott, D., & Pervan, G. (2012). Design science in decision support systems research: An assessment using the Hevner, March, Park, and Ram Guidelines. *Journal of the Association for Information Systems*, *13(*11), 923-949.

Bardhan, I., Oh, C., Zheng, E., & Kirksey, K. (2015). Predictive analytics for readmission of patients with congestive heart failure: Analysis across multiple hospitals. *Information Systems Research*, *26*(1), 19-39.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2014). Transformational issues of big data and analytics in networked business. *MIS Quarterly*, *38*(2), 629-632.

Balakrishnan, P. V., & Jacob, V. S. (1996). Genetic algorithms for product design. *Management Science*, *42*(8), 1105-1117.

Bansal, G., Sinha, A. P., & Zhao, H. (2008). Tuning data mining methods for cost-sensitive regression: A study in loan charge-off forecasting. *Journal of Management Information Systems*, *25*(3), 315-336.

Banker, R. D., & Kauffman, R. J. (2004). 50th anniversary article: The evolution of research on information systems: A fiftieth-year survey of the literature in management science. *Management Science*, *50*(3), 281-298.

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around procedural privacy protections. *Communications of the ACM*, *57*(11), 31-33.

Benbasat, I., & Zmud, R. W. (2003). The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. *MIS Quarterly*, *27*(2), 183-194.

Bjørn-Andersen, N., Earl, M., Holst, O., & Mumford, E. (1982). *Information society: For richer, for poorer*, Amsterdam: North-Holland.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Broniatowski, D., Paul, M. J., & Dredze, M. (2014). National and local influenza surveillance through Twitter: An analysis of the 2012-2013 influenza epidemic. *PLoS One*, *8*, e83672.

Brown, S. A., Massey, A. P., Montoya-Weiss, M. M., & Burkman, J. R. (2002). Do I really have to? User acceptance of mandated technology. *European Journal of Information Systems*, *11*(4), 283-295.

Brown, D. E., Abbasi, A., & Lau, R. Y. K. (2015). Predictive analytics: Predictive modeling at the micro level. *IEEE Intelligent Systems*, *30*(3), 6-8.

Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big data—a fashionable topic with(out) sustainable relevance for research and practice? *Business & Information Systems Engineering*, *5*(2), 65-69.

Buitelaar, P., Cimiano, P., & Magnini, B. (Eds.). (2005). *Ontology learning from text: Methods, evaluation and applications*. Amsterdam: IOS Press.

Buytendijk, F. (2014). Hype cycle for big data, 2014. *Gartner*. Available from https://www.gartner.com/doc/2814517/hype-cycle-big-data-

Chandler, N., Hostmann, B., Rayner, N., & Herschel, G. (2011). Gartner's business analytics framework. *Gartner*.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*, 67-80.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide. *SPSS*. Retrieved from www.crisp-dm.org

Chatfield, C. (1995). *Problem solving: A statistician's guide* (2nd ed.). London: Chapman & Hall/CRC.

Chatterjee, S., & Sarker, S. (2013). Infusing ethical considerations in knowledge management scholarship: Toward a research agenda. *Journal of the Association for Information Systems*, *14*(8), 452-481.

Chatterjee, S., Sarker, S., & Fuller, M. (2009a). A deontological approach to designing ethical collaboration. *Journal of the Association for Information Systems, 10*(3), 138-169.

Chatterjee, S., Sarker, S., & Fuller, M. (2009b). Ethical information systems development: A Baumanian postmodernist perspective. *Journal of the Association for Information Systems, 10*(11), 787-815.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188.

Chiang, R. H., Barron, T. M., & Storey, V. C. (1994). Reverse engineering of relational databases: Extraction of an EER model from a relational database. *Data & Knowledge Engineering*, *12*(2), 107-142.

Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems*, *3*(3), 1-13.

Chung, W., Chen, H., & Nunamaker, J. F., Jr. (2005). A visual framework for knowledge discovery on the Web: An empirical study of business intelligence exploration. *Journal of Management Information Systems*, *21*(4), 57-84.

Christensen, C. (1997). *The innovator's dilemma: When new technologies cause great firms to fail*. Boston, MA: Harvard Business Review Press.

Coco, C. T., Jamison, F., & Black, H. (2011). Connecting people investments and business outcomes at Lowe's. *People & Strategy*, *34*(2), 28-33

Coolidge, A. (2013). New technology helps Kroger speed up checkout times. *The Cincinnati Inquirer*. Retrieved from http://www.usatoday.com/story/money/business/2013/06/20/new-technology-helps-kroger-speed-up-checkout-times/2443975/

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

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

Davenport, T. H., Harris, J., & Shapiro, J. (2010). Competing on talent analytics. *Harvard Business Review*, *88*(10), 52-58.

Davenport, T. H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, *90*(10), 70-76.

Davenport, T. H., & Kirby, J. (2015). Beyond automation. *Harvard Business Review*, *94(*6*),* 59-65.

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation). Sloan School of Management, Massachusetts Institute of Technology.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113.

Dennis, A. R., & Carte, T. A. (1998). Using geographical information systems for decision making: Extending cognitive fit theory to map-based presentations. *Information Systems Research*, *9*(2), 194-203.

Derman, E. (2011). *Models. Behaving. Badly.* New York, NY: Simon and Schuster.

Embley, D. W., & Liddle, S. W. (2013). Big data—conceptual modeling to the rescue. In W. Ng, V. C. Storey, & J. C. Trujillo (Eds.), *ER conference* (vol. 8217, pp. 1-8). Springer.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-54.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27-34.

Gaskin, J., Berente, N., Lyytinen, K., & and Yoo, Y. (2014). "Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines," *MIS Quarterly*, 38(3), 849-871.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, *27*(1), 51-90.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321-326.

Goes, P. (2014). Big data and IS research. *MIS Quarterly*, *38*(3), iii-viii.

Gomes, L. (2014). Machine-learning maestro Michael Jordan on the delusions of big data and other huge engineering efforts. *IEEE Spectrum*.

Gregor, S., & Klein, G. (2014). Eight obstacles to overcome in the theory testing Genre. *Journal of the Association for Information Systems*, *15*(11), i-xix.

Han, J., Kamber, M., & Pei, J. (2006). *Data mining: Concepts and techniques*. New York: Morgan Kaufmann.

Harrington, A. (2005). *Modern social theory*. New York: Oxford University Press.

Hassan, A., Abbasi, A., & Zeng, D. (2013). Twitter sentiment analysis: A bootstrap ensemble framework. In *Proceedings of the IEEE International Conference on Social Computing* (pp. 357-364).

Heudecker, N. (2013). Hype cycle for big data, 2013. *Gartner*. Available from https://www.gartner.com/doc/2574616/hype-cycle-big-data-

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75-105.

Hill, K. (2012). How Target figured out a teen girl was pregnant before her father did. *Forbes. http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/*

Hodgkinson, G. P., & Starbuck, W. H. (2008). *The Oxford handbook of organizational decision making*. Oxford, UK: Oxford University Press.

Horan, J. A. (2011). The essential CIO. *IBM CIO C-Suite Studies*. Retrieved from http://www-935.ibm.com/services/c-suite/cio/study/

Jensen, M. L., Lowry, P. B., Burgoon, J. K., & Nunamaker, J. F., Jr. (2010). Technology dominance in complex decision making: The case of aided credibility assessment. *Journal of Management Information Systems*, *27*(1), 175-202.

Jensen, M. L., Averbeck, J. M., Zhang, Z., & Wright, K. B. (2013). Credibility of anonymous online product reviews: A language expectancy perspective. *Journal of Management Information Systems*, *30*(1), 293-323.

Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). Predictive modeling with big data: Is bigger really better? *Big Data*, *1*(4), 215-226.

Kaushik, A. (2011). *Web analytics 2.0: The art of online accountability and science of customer centricity*. Indianapolis, IND: Wiley.

KDNuggets (2011). What do you call analyzing data? *KDNuggets.* Retrieved from http://www.kdnuggets.com/polls/2011/what-do-you-call-analyzing-data.html

Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M. (2012). Analytics: The widening divide. *MIT Sloan Management Review*, 53(2), 1-21.

Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1-12.

Kitchin, R. (2014b)*. The data revolution*. London: Sage.

Kramer A. D. I., Guillory J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790.

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

Lau, R. Y. K. Liao, S. Y., Wong, K. F., & Chiu, K. W. (2012). Web 2.0 environmental scanning and adaptive decision support for business mergers and acquisitions. *MIS Quarterly*, *36*(4), 1239-1268.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444.

Lin, M., Lucas, H. C., Jr., & Shmueli, G. (2013). Research commentary-too big to fail: Large samples and the p-value problem. Information Systems Research, 24(4), 906-917.

Lycett, M. (2013). “Datafication”: Making sense of (big) data in a complex world. European Journal of Information Systems, 22(4), 381-386.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review. Retrieved from https://hbr.org/2012/10/big-data-the-management-revolution/ar

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights/Business%20Technology/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_OpenData_Full_report_Oct2013.ashx

Marchand, D. A., & Pepper, J. (2013). Why IT fumbles analytics. *Harvard Business Review*. Retrieved from https://hbr.org/2013/01/why-it-fumbles-analytics

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.

Newman, D. (2014). Big Data Means Big Disruption. *Forbes.* Retrieved from http://www.forbes.com/sites/danielnewman/2014/06/03/big-data-means-big-disruption/

Nunamaker, J. F., Jr., Chen, M., & Purdin, T. (1991). Systems development in information systems research. *Journal of Management Information Systems*, *7*(3), 89-106.

Nunamaker, J. F., Jr. (1992). Build and learn, evaluate and learn. *Informatica*, *1*(1), 1-6.

Parsons, J., & Wand, Y. (2012). Extending classification principles from information modeling to other disciplines. *Journal of the Association for Information Systems*, *14*(5), 245-273.

Rai, A., & Sambamurthy, V. (2006). Editorial notes-the growth of interest in services management: Opportunities for information systems scholars. *Information Systems Research*, *17*(4), 327-331.

Redman, T. C. (2008). *Data driven: Profiting from your most important business asset*. Boston, MA: Harvard Business Press.

Richards, N. M., & King, J. H. (2013). Three paradoxes of big data, *Stanford Law Review Online*, *66*(41), 41-46.

Sarker, S., Xiao, X., & Beaulieu, T. (2013). Qualitative studies in information systems: A critical review and some guiding principles. *MIS Quarterly*, *37*(4), iii-xviii.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of big data. *IBM Institute for Business Value*.

Sein, M., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*, *35*(1), 37-56.

Shaft, T. M., & Vessey, I. (2006). The role of cognitive fit in the relationship between software comprehension and modification. *MIS Quarterly*, *30*(1), 29-55.

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*, 433-441.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, *33*(2), 111-126.

Shmueli, G., & Koppius, O. (2011). Predictive analytics in information systems research. *MIS Quarterly*, *35*(3), 553-572.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: MIT Press.

Song, Y., Sahoo, N., Srinivasan, S., & Chrysanthos, D. (2014). Uncovering path-to-purchase segments in large consumer population. In *Proceedings of the 24th Workshop on Information Technologies and Systems*.

Steadman I. (2013). Big data and the death of the theorist. *Wired*. Retrieved from http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory

Stigler, G. (1961). The economics of information. *The Journal of Political Economy*, *69*(3), 213-225.

Storey, V. C., Chiang, R. H., Dey, D., Goldstein, R. C., & Sudaresan, S. (1997). Database design with common sense business reasoning and learning. *ACM Transactions on Database Systems*, *22*(4), 471-512.

Stubbs, E. (2011). The value of business analytics: Identifying the path to profitability. New York, NY: John Wiley & Sons.

Tambe, P. (2014). Big data investment, skills, and firm value. *Management Science*, *60*(6), 1452-1469.

Te'eni, D. (2006). Designs that fit: An overview of fit conceptualizations in HCI. In P. Zhang & D. Galletta (Eds.), *Human computer interaction and management information systems: Foundations* (pp. 205-224). London, England: M. E. Sharpe.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Vapnik, V. N. (1998). *Statistical learning theory* (Vol. 1). New York: Wiley.

Venkatesh, V, Thong, J. Y. L., & Xu, X. (Forthcoming). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems*.

Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*(2), 219-240.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research*, *3*(1), 36-59.

Wand, Y., & Weber, R. (2002). Research commentary: Information systems and conceptual modeling—a research agenda. *Information Systems Research*, *13*(4), 363-376.

Wang, Y., & Ram, S. (2015). Prediction of location-based sequential purchasing events using spatial, temporal and social patterns. *IEEE Intelligent Systems*, *30*(3), 10-17.

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, *25*(1), 17-41.

Yetgin, E., Jensen, M., & Shaft, T. (2015). Complacency and intentionality in IT use and continuance. *AIS Transactions on Human-Computer Interaction*, *7*(1), 17-42.

Zahedi, F. M., Abbasi, A., & Chen, Y. (2015). Fake-website detection tools: Identifying elements that promote individuals' use and enhance their performance. *Journal of the Association for Information Systems*, *16*(6), 448-484.

Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems*, *25*(6), 13-16.

Zhao, H., Sinha, A. P., & Bansal, G. (2011). An extended tuning method for cost-sensitive regression and forecasting. *Decision Support Systems*, *51*(3), 372-383.

Zhang, W., Lau, R., & Li, C. (2014). Adaptive big data analytics for deceptive review detection in online social media. In *Proceedings of the International Conference on Information Systems*.

About the Authors

Ahmed Abbasi is Murray Research Professor and Associate Professor of Information Technology in the McIntire School of Commerce at the University of Virginia. He is Director of the Center for Business Analytics and a member of the Predictive Analytics Lab. His research on online fraud and security, natural language processing, social media, and e-health has been funded through multiple grants from the National Science Foundation. Ahmed received the IBM Faculty Award and AWS Research Grant for his work on Big Data. He has published over fifty peer-reviewed papers in journals and conferences, including top-tier outlets such as *MIS Quarterly*, *Journal of the AIS*, *Journal of MIS*, *ACM Transactions on IS*, *IEEE Transactions on Knowledge and Data Engineering*, and *IEEE Intelligent Systems*. One of his papers was considered a top publication by the AIS. Ahmed also won best paper awards at *MIS Quarterly* and *WITS*. His work has been featured in various media outlets, including the *Wall Street Journal*, the Associated Press, and Fox News. Ahmed serves as an Associate Editor for *Information Systems Research, Decision Sciences Journal*, *ACM Transactions on MIS*, and *IEEE Intelligent Systems*, and as an editorial review board member for the *Journal of AIS*. He is a senior member of the IEEE and serves on program committees for various conferences related to computational linguistics, text analytics, and data mining. He is also a co-founder and/or advisory board member for multiple predictive analytics-related companies.

Suprateek Sarker is a Professor of Information Technology at the McIntire School of Commerce, University of Virginia, USA. He also serves as Visiting Distinguished Professor at Aalto University School of Business, Helsinki, Finland. He is primarily a qualitative researcher, and his past work has been published in leading outlets including the *MIS Quarterly, Information Systems Research, Journal of the AIS, Journal of MIS, European Journal of Information Systems, MIS Quarterly Executive, Journal of Information Technology, IEEE Transactions,* ACM *Transactions on MIS*, *Communications of the ACM*, *MIS Quarterly Executive*, and *Decision Sciences Journal*. He is currently serving as the Editor-in-Chief of the *Journal of the AIS*, a Senior Editor of *Decision Sciences Journal*, a Senior Editor (Emeritus) of *MIS Quarterly*, a member of the Board of Editors of *Journal of MIS*, and an editorial board member of *IEEE Transactions on Engineering Management*. Some of his past work has been funded by the National Science Foundation (NSF) and the Institute for the Study of Business Markets (ISBM). He is also a past recipient (with S. Sahay) of the Stafford Beer Medal awarded by the Operational Research Society, UK.

Roger H. L. Chiang is a Professor of Information Systems at Department of Operations, Business Analytics, and Information Systems, Carl H. Lindner College of Business, University of Cincinnati. Dr. Chiang's research interests are in business intelligence and analytics, data and knowledge management, and intelligent systems, particularly in database reverse engineering, database integration, data and text mining, sentiment analysis, document classification and clustering, domain knowledge discovery, and semantic information retrieval. He has over fifty refereed papers published by conferences and journals including ACM *Transactions on Database Systems*, *ACM Transactions on MIS, Communications of the ACM, The DATA BASE for Advances in Information Systems, Data & Knowledge Engineering, Decision Support Systems, Journal of American Society for Information Science and Technology, Journal of Database Administration, Journal of MIS, Marketing Science, MIS Quarterly,* and *Very Large Data Base Journal*. He has served as the Senior Editor of *The DATA BASE for Advances in Information Systems*, *Decision Sciences*, and *Journal of the AIS* and the associate editor of *Information & Management*, *Journal of the AIS*, *Journal of Database Management*, *International Journal of Intelligent Systems in Accounting*, *Finance and Management*, and *MIS Quarterly*.

Copyright © 2016 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from publications@aisnet.org.