Digital Data Streams:

Creating Value from the Real-Time Flow of Big Data

Federico Pigni
Gabriele Piccoli
Richard Watson

There is no escaping the Big Data hype. Vendors are peddling Big Data solutions; consulting firms employ Big Data specialists; Big Data conferences are aplenty. There is a rush to extract golden nuggets (of insight) from mountains (of data). By focusing merely on the mountain (of Big Data), these adventurers are overlooking the source of the revolution—namely, the many digital data streams (DDSs) that create Big Data—and the opportunity to improve real-time decision making. This article discusses the characteristics of DDSs, describes their common structure, and offers guidelines to enable firms to profit from their untapped potential. (Keywords: Value Creation, Competitive Advantage, Information Systems, Decision Making)

As organizations rely increasingly on data to determine and meet customers' needs, changes in data availability and timing influence a firm's ability to create value in the form of new products, services, or processes (Table 1). The emergence of digital data streams is creating strategic opportunities for existing firms and enabling the formation of new enterprises. The catalyst for this seismic change is the massive generation of real-time structured and unstructured data streams that organizations can leverage for decision making and operational change. Emblematic of these new enterprises is Uber, the world's largest "taxi" company, speculatively valued at $50 billion.[1] Uber owns no vehicles, but harnesses a real-time digital data stream of its drivers' cars and matches them with real-time demand for rides. Existing organizations are also successfully leveraging the real-time flow of "big data" for new value creation. Consider San Francisco's SFPark, a private-public partnership of the San Francisco Municipal Transportation Agency (SFMTA). By installing magnetometers to detect a vehicle in each of the city's paid parking bays, SFMTA creates a real-time flow of parking data. By doing so, it supplies real-time visibility of available parking spaces, reducing both the average time motorists spend searching for a parking spot (43 percent), the average cost of on-street parking (4 percent), and garage rates (12 percent). Furthermore, the city is also able to appropriate some of the value created by introducing demand-response pricing. There is a gradual and periodic block-by-block rate tweaking (both increases and

Federico Pigni is an Associate Professor of Information Systems at Grenoble Ecole de Management, France. <federico.pigni@grenoble-em.com>

Gabriele Piccoli is the Edward G. Schleider Chair for Information Systems in the E. J. Ourso College of Business at Louisiana State University, and the Director of the Digital Data Streams Lab at LSU. <gpiccoli@cct.lsu.edu>

Richard Watson is a Regents Professor and the J. Rex Fuqua Distinguished Chair for Internet Strategy in the Terry College of Business at the University of Georgia, and the Research Director for the Advanced Practices Council of the Society of Information Management. <rwatson@terry.uga.edu>

decreases) to meet availability goals. As a result, parking bay occupancy targets are reached 31 percent more often, and there has been a 30 percent reduction of greenhouse gas emissions because of the fewer miles travelled by drivers circling for parking.[2] The venerable company Macy's implemented a strategy for engaging millennial shoppers by harnessing their location digital data streams with Apple's iBeacon technology. Specifically, Macy's delivers personalized and contextually relevant content through Shopkick (a popular shopping app) including personalized department-level deals, discounts, recommendations, and rewards to directly influence in-store retail sales.[3] The success of the early pilots in increasing shoppers' engagement in 2013 resulted in the full-scale deployment in all 768 shops, the largest beacon technology implementation to date. The retail industry is following this move, and beacon technology is expected to influence over $4 billion worth U.S. sales in 2015.[4]

While these specific examples are striking, managers must still navigate the flood of new data streams and discriminate between uncontrolled hype and realistic opportunities for value creation.

Digital Data Streams

The Evolution of Business Data Processing

At the dawn of digital business data processing in the 1950s, transactions were batched and processed by a mainframe at periodic intervals—daily, weekly, or even monthly. As a result, the main opportunity for value creation was in the automation of clerical work and the standardization of routine decisions. The introduction of computer terminals, personal computers, and networks enabled the processing of online transactions. Value could now be created in novel ways, for example by improving customer service or shifting work to business partners, and eventually to customers (e.g., package tracking). A positive side effect of this trend was the recording of these transactions in databases. Aggregate analyses became viable, for example to identify market trends and detect customer preferences. The relational database, introduced in the 1970s, greatly expanded the support for data-driven decision making. The reality, however, is that many companies remained data rich and information poor as many were slow to develop the competencies needed to create information from the ever-increasing data repositories.[5] In the 1990s, Business Intelligence (BI) began to fill the void and enabled more organizations to effectively analyze large volumes of data to gain insights on internal operations and customers. Until very recently, the focus has been on mining larger, but relatively slow-changing, volumes of data. Even with the increasing attention to "Big Data," the spotlight has remained focused on the sheer quantity and variety of the data for potential analysis (Table 1).

TABLE 1. Digital Data Processing Eras

Era of the	Value Proposition
Batch	Automate and standardize clerical work
Transaction	Self-service for customers and suppliers to reduce cost and improve efficiency
Digital Data Stream	Wider view of the customer and new information services

It is now evident that there is a transition underway in digital business data processing with the velocity of data becoming increasingly prominent.[6] The computer mediation of everyday activities leads to continuous flows of real-time data captured in digital form at their inception—a phenomenon we call digital data streaming. A Digital Data Stream (DDS) is a continuous digital encoding and transmission of data describing a related class of events.[7] The transmission, or flow, of these digital representations of events is a DDS, which may be human-generated (e.g., a tweet, an instagram) or machine-generated (e.g., a CO_2 reading, a GPS location). DDSs allow managers to dissect events in real-time, to shorten the decision cycle, and to deepen their understanding of customers at the same time. A spliced set of DDSs can serve as a wide-angle lens on a transaction, providing pre-transaction details (such as when and where the customer entered a store) and post-transaction attributes (such as what the customer tweeted about a purchase). More than just analysis, DDSs enable real-time processing and response to an event (e.g., algorithmic trading). Thus, DDSs increase organizational visibility and accelerate the corporate metabolism.

Opportunities in the New Model

DDSs open new and broader opportunities for value creation for those organizations that adapt to the change. The laggards who fail to recognize and respond to this change may well find their business diminished or gone. The consequences are highlighted in a McKinsey & Co. report on the European banking industry.[8] It cautions banks who delay digitizing by stating that digital transformation will put upwards of 30 percent of the revenues of a typical European bank in play. Few businesses can ignore a change that imperils such a high level of their revenues.

Furthermore, the prior two shifts in data processing (to batch and to networks) had profound effects that transformed business. The Internet decimated industries (e.g., travel agents and music stores), and DDSs already threaten the taxi industry. For the last half-century, information systems have been the major force for productivity gains.[9] DDSs will continue to fuel this productivity engine and present the same major challenges for companies to adapt to and survive information technology disruptions. Organizations need to investigate how to tap into DDSs to develop real-time insights and actions to protect, and possibly enhance, their revenue.

Generating Digital Data Streams

Every time we visit a website, drive through a tollbooth, send an e-mail, or just tap an icon on a smartphone, we generate digital data. Every time we walk or run with a Fitbit on our wrist or Instagram a picture of our food from a restaurant, we create a digital representation of an event. Our smartphones are platforms packed with sensors able to gather real-time data concerning a multitude of parameters: sound and images, light, gestures, proximity, rotation, movement, magnetic field, temperature, humidity, atmospheric pressure, and location. In the near future, wearable devices will likely measure blood pressure, pulse, and other vital medical signs. Your smartwatch provider might soon, if not already, know more about your health status than your general practitioner. The list of opportunities to capture events in the digital space is seemingly endless. Consider the 2012 U.S. presidential election. People sent more than 31 million related tweets on Election Day, and the tweet-stream about the vote averaged 10,000 tweets per second in the most active hours.[10] Each tweet was an opinion, a point of view, a comment, or a critique on the unfolding of the election. All these events are born digital, captured, streamed, and therefore available for processing in digital form *as they occur*.

The Basic Elements of a Digital Data Stream

A DDS can capture, and thus represent, up to six basic elements describing an event (see Table 2). These elements are "primitives," meaning that they cannot be described in terms of other elements, or inferred from them. These primitives derive from what are commonly known as the 5W+H of narrative (who, what, when, where, why, and how). Discovered and re-discovered several times throughout history,[11] they originate in rhetoric and the 2nd century BCE work of Hermagoras of Temnos, Gorgias of Leontini, Cicero, and Quintilian.[12] In rhetoric, 5W+H are the *circumstances,* the elements (*moria peristaseos*—particulars of circumstance) of an event, and in narrative they represent the basic grammar of a story. Computer scientists attempting to extract events from multimedia DDSs have revitalized the 5W+H model,[13] which is now at the heart of several provenance ontologies.[14] While the model can readily handle structured (e.g., a single numeric) or ill-structured (e.g., a text string) data, extracting value from ill-structured data is challenging and has galvanized considerable research on text mining.

TABLE 2. Elements of a digital data stream segment

Element	Description	Example
When	The time when the data segment was created	A timestamp with date, time, and time zone
Where	The location of the entity when the segment was created	Latitude, longitude, elevation
Who	The unique identifier of the entity that caused the data segment to be created	Person's customer number, RFID of a pallet, URL of a web site
What	The activity that caused the segment to be created	The identifier of an item in a sales transaction, the arrival of a ship in a port
How	The means by which the event was initiated, authorized, or completed	Credit card number for payment, status of arriving flight (e.g., safe landing)
Why	Motivation for the action related to data segment creation	Birthday gift, planned destination

In a digital data stream, an event can have one or more associated primitives. For example, a retail sale might have two *what* elements, a credit card number and a product code. Some elements might be unavailable (i.e., null values). For instance, credit card DDSs from retail transactions or GPS DDSs from traffic typically do not contain any data for the *why* element, not because the motives for purchasing a product or the purpose of a trip do not exist, but because there is today no available digital technology for capturing them unobtrusively.

The shift from desk-bound to mobile technology for electronic interaction means that *where* and *who* are often precisely measured. The GPS unit in a smart phone can report latitude and longitude in real-time and derive elevation from them. Smart phones and personal mobile devices are registered to a particular person, so *who* precision is also increasing rapidly. Other elements of a transaction can provide further identification, such as a credit card or customer number. *What* and *how* are common transaction elements. We think of a tweet as a text string limited to 140 characters, but a tweet segment contains metadata to indicate when the text was tweeted, where it was created, the author's URL, details of the author, and so forth.[15] *Why* is still a puzzle, as in most cases individuals don't reveal their motives directly. However, by looking at data streams surrounding a transaction, it can be possible to infer motive. Consider a sales transaction (see Figure 1). It is recorded (1) and a short time (2 hours and 20 minutes) after at a nearby location (~ 7 miles as calculated using latitude and longitude values) the same person (the retailer has purchased data linking Facebook and Twitter identifiers to people and addresses) makes a Facebook entry (2). Later that day, the same person posts a tweet (3). The timing and location data help to confirm that the identifiers match. The various elements are spliced to record the episode (4).

DDS splicing has far reaching implications. The more DDSs become available, the greater the opportunity for reconstructing complete episodes from single events and gaining further insights on business occurrences. Recent studies demonstrate that data, in public or anonymized data sets, can be combined to identify individuals. For example, the missing *who* element was re-identified from anonymized

FIGURE 1. Splicing DDSs to Infer Motive

	When	Where	Who	What	How	Why
(1) Sale	2014-03-05 12:11:31	33.903327, −83.3299229	6713458	8912345	1234 4567 7890 1234	
(2) Facebook	2014-03-05 14:31:56	33.9280696, −83.4466097	VegXYZ			Got Bruce a red tie for b'day
(3) Twitter	2014-03-05 21:09:03	33.958064, −83.378344	VegXYZ			
(4) Spliced episode	2014-03-05 12:11:31	33.903327, −83.3299229	6713458	8912345	1234 4567 7890 1234	Got Bruce a red tie for b'day

mobile phone data sets,[16] credit card transactions, medical records, and the Netflix Challenge dataset[17] by splicing external temporal and geolocation data (*when* and *where*) like the metadata contained in a tweet or in a Facebook status update.

The move to DDS business data processing creates a new set of tactical and strategic opportunities for firms. Tactically, managers have the opportunity to leverage real-time data about streamed events (e.g., tweets on a new product, number of people in a section of a store) and the availability of new classes of events (e.g., occupancy of a parking space). Tactical initiatives entail taking immediate action based on the content of a single stream—such as routing a request for a ride to the closest Uber driver—or integrating several streams in real-time to gain superior insights—deciding how to respond to a tweet based on a customer's social influence measured by Klout. Strategically, executives can confront major business risks or create value by designing new business models around the unique characteristics of emerging DDSs.

DDS Tactics: Creating Immediate Value

DDS tactics is about recognizing and extracting the incremental value afforded by a DDS, as well as being able to prioritize execution of initiatives. Previous research in the business intelligence tradition introduced the concept of *response time latency* (see Figure 2).[18] It recognizes the loss in business value associated with delays in: extracting needed data from an event of interest (*capture latency*); transforming data into usable information (*analysis latency*); and making a decision to act upon this new information (*decision latency*).

The tactical value of DDSs stems from a firm's ability to moderate response time latency to improve the timeliness of decisions and actions. Predictive analytics experts, such as Dan Porter (the Director of Statistical Modeling for the Obama campaign), have long recognized the importance of data streams: "While refining algorithms and statistical models provides incremental improvements, the real boost in predictive and explanatory power comes from adding new data sources."[19]

Our analysis of existing DDS initiatives shows that organizations extract value from events in a DDS via either process-to-actuate or assimilate-to-analyze tactics (see Figure 3).

Process-to-actuate occurs when a firm creates value by initiating action based on real-time DDS processing. An insurance company monitoring a weather forecast data stream and sending text messages to its customers in the area where hail is expected in the next 30 minutes illustrates the immediacy of process-to-actuate. The firm combines events that are currently streaming in a DDS (i.e., real-time location-specific short-term weather forecasts) and the results of a static database query and other contextual data in order to alert its potentially affected customers in a timely manner. The result is superior customer service and fewer insurance claims because customers have been able to garage their vehicles at the right time.

Assimilate-to-analyze occurs when a firm extracts value by merging multiple data streams and static databases and dissecting the composite data set. The focus is on extraction of insights rather than immediate action. To avoid the financial risks

FIGURE 2. Response Time Latency (adapted from Hackathorn, 2002)

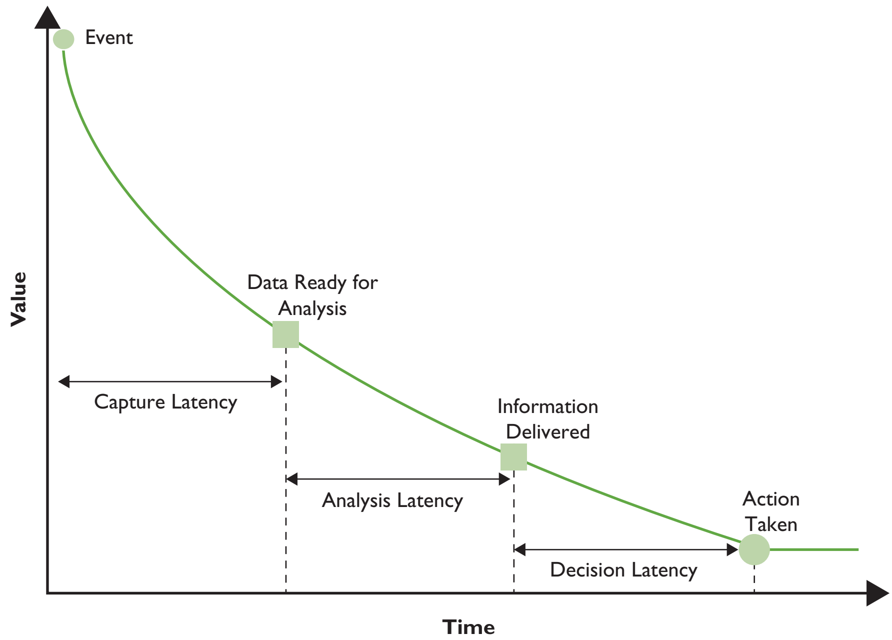
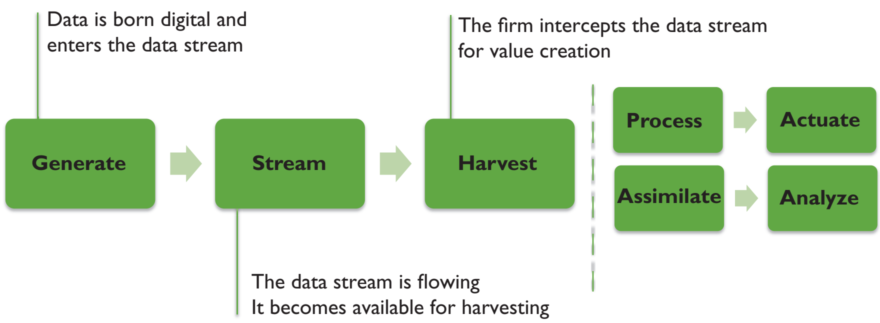


FIGURE 3. Tactical Opportunities to Leverage Digital Data Streams



associated with planning errors, some firms have integrated external DDSs in their demand forecasting system. For instances, Tesco and other retailers merge and analyze data from multiple digital data streams to generate forecasts to estimate demand. Predictions are based on information generated from store location, product characteristics, recent weather history, and weather forecasts. Note how the result of the analysis is not immediate automatic action, as in process-to-actuate, but rather the presentation of superior insight that enables better decision making.

Naturally, firms can combine these two approaches so that a prescribed action is the result of the analysis of multiple data streams and database records.

While managers intuitively appreciate that new DDSs have the potential to reduce response time latency, the challenge is to reconcile these opportunities with managerial and business ecosystem needs. More specifically, organizations are often overwhelmed by the potential of new data sources to the point that they are unable to decide what to pursue.[20] Moreover, there are yet many events of interest that may not be appropriately captured in a DDS. Thus, tactical decisions about what initiatives to pursue depend upon the potential that different events produce for value creation, and the ready availability of events in a DDS. We call this concept *event streamability*.

Event Streamability: Identifying Exploitable Events

Event streamability is the degree to which a class of events is amenable to encoding and channeling through a digital data stream. Streamability is a function of an event's characteristics and the current capabilities of available technology at a specific time. Event streamability is determined by three characteristics: detectability, measurability, and interpretability.

Detectability assesses whether each of the 5W+H elements of an event exceeds a minimum threshold magnitude. For instance, if a person is stealing a negligible amount of money from a company, the thefts might go unnoticed, as in the Salami slicing method of fraud.[21] Detectability is an essential condition for capture latency mitigation. A firm cannot harvest events that are not easily detectable and therefore cannot implement value extraction initiatives. Consider the presence of customers in a specific area of a store. The *who* and *where* elements of this event are detectable by surveillance cameras if they have enough resolution to discern individuals. Low detectability reduces the possibility of identifying an event, and thus the likelihood of creating an associated digital data stream. Imperceptible events cannot be streamed.

Measurability assesses whether the magnitude of the 5W+H elements of an event can be quantified with the needed precision. The higher an event's measurability, the greater the potential for capturing latency reduction. A firm can quickly make available for analysis those event elements that are easily measurable with high precision. Conversely, low-measurability event elements require significant transformation and manipulation to be identified, often imprecisely, before being analyzed. If the degree of measurement precision is insufficient, the elements are simply not usable because the resulting DDS produces a stream of unreliable data. For example, a firm's quarterly profits have high measurability, while pain is low on this dimension because it is self-reported, and individual pain tolerance varies subjectively. In the case of video surveillance cameras, the accuracy in measuring the presence of customers depends on the type of analysis performed on the video stream. Measurability thus determines the precision, and usefulness, of the individual elements within a stream.

Interpretability assesses whether a firm can achieve the needed understanding of the content of one or more elements in a stream or a portion of a stream.

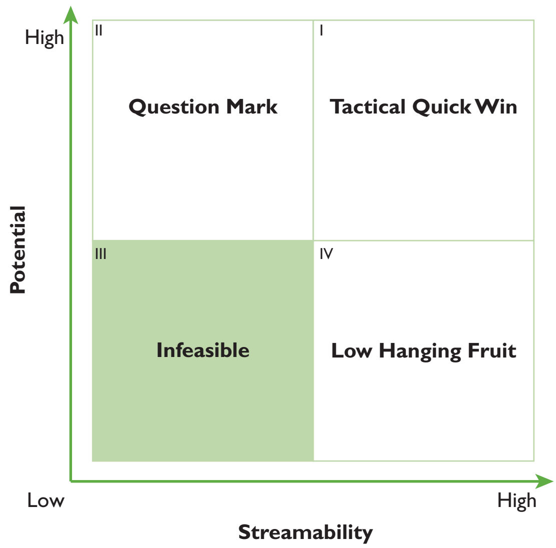
Interpretability is instrumental in mitigating analysis latency. Consider the Twitter stream, which among other things contains a tweet's origin and text. The location element is highly interpretable because it obeys the global standard for latitude and longitude. However, what can we make of the sentiment or mood of the tweet from the text? In its crudest form, sentiment analysis simply reduces a message to a list of words and computes a score for the tweet based on whether a word appears in a list of positive words (score +1), negative words (score −1), or neither (score 0). Thus, "Impressed and amazed: you are peerless in your achievement of unparalleled mediocrity" scores 4,[22] and is thus considered to express a positive sentiment. Not surprisingly, text within a stream is more difficult to interpret and comprehend than location. The identification of an object or a behavior in a video stream can be even more challenging. DDSs come in different forms: customers' opinions, voice registrations, video streams, and sensor readings are some examples of the variety of DDSs. Extracting value from these sources requires the capacity to manage and process heterogeneous data streams exhibiting varying degrees of structure. Interpretability will then vary depending on a firm's competencies in processing and identifying valuable events and extracting meaning from these streams. As for surveillance cameras, new and improved video processing algorithms allow the extraction of additional features such as assessment of gender, estimate of age, and categorization of behavior of a bar's patrons.

Streamabilty depends on the inherent characteristics of the event; it is what information system theorists call a soft-primary characteristic,[23] because it is relative to advances in information technology. Thus, detectability, measurability, and interpretability of the same event improve over time due to technological progress in sensors, computational power, and software algorithmic efficiency. Automatic, for instance, developed an adapter that plugs into the diagnostic port of most cars manufactured since 1996 to intercept and stream real-time data on engine functioning, fuel consumption, and other car performance data. Using custom algorithms, and data from the onboard GPS and accelerometer sensor, Automatic coaches the driver to adopt fuel-efficient behaviors. It makes the driver aware of hard braking and excessive acceleration and speed. Automatic unlocks detectability and measurability to create environmental benefits. It illustrates how the increasing pervasiveness of sensors and networks makes more events potentially detectable and hence streamable.

Tactical Priorities

The tactical opportunities of DDSs are tightly related to the assessment of streamability of events of interest. Managers should first identify promising events that are currently not used for analysis or processed to make decisions and take action. However, firms have to prioritize initiatives under resource constraints, and can't unquestioningly attempt to extract value from all potentially worthy events. Streamability aids in this critical prioritization step because it helps in estimating how difficult it would be to reduce response time latency and take advantage of the data generated by the events under consideration. Evaluating potential value extraction in the context of streamability enables a prioritization analysis (Figure 4).

FIGURE 4. The DDS Prioritization Matrix



Low-potential/low-streamability events are infeasible and should be ignored. On the contrary, high-potential/high-streamability events can provide tactical quick wins. More interesting questions surround the approaches on the off-diagonal of the prioritization matrix. Low-potential/high-streamability events are "low-hanging fruit" that provide opportunities for proof of concept initiatives, even when they have limited upside potential. In fact, the DDS may already exist and may be easily harvested by the firm. While nobody has examined this issue in the context of DDSs, related IS research has shown that prototypes and proofs of concept produce unintended positive effects with limited investments. For example, they enable organizational learning, cultural change, and can provide evidence to convince sceptical executives.[24] They could assist firms to move up the DDS exploitation learning curve. The most challenging decisions managers need to make about tactical DDS initiatives pertain to events with high potential but low streamability. The perception of the upside potential implies that the organization is able to identify the likely gain of processing the DDS. However, limited streamability can make value extraction costly and difficult to realize. Significant investments are needed to overcome both capture and analysis latency, making the return on investment for the initiative unclear. Keep in mind that over time, as technology evolves and a particular event's streamability changes, initiatives that were not feasible may become attractive.

DDS Strategy: New Avenues for Value Creation

Our analysis in the previous section guides managers in discovering and prioritizing DDS tactical initiatives. Within existing business models, these are incremental venues for extracting value from a small set of specific DDSs. However, much of our research has focused on the strategic implications of DDS. We systematically analyzed 70 organizations—both startups and mature firms—that innovatively leverage DDSs.[25] Their ingenuity ranges from using DDSs to enable business model improvements to outright DDS-enabled industry transformation. These firms embrace DDS data processing to create new products and services, and to seize growth opportunities.

Our research led us to abstracting five distinct processes for value creation,[26] which we call DDS Value Creation Archetypes. They differ in their value creation focus and mechanisms. While visible initiatives and projects by the firm may have elements of various archetypes, each archetype hinges on defining value creation activities directly related to the value proposition the firm seeks to offer based on DDSs. Thus, understanding the five archetypes can help managers better frame their strategic objectives and challenge their current business model in an effort to seize opportunities afforded by the emerging DDS data processing framework (see Figure 5).[27]

Managers have historically used information technology to address three types of strategic risks: demand, efficiency, and innovation risk (see Table 3).[28] Three of the archetypes we uncovered directly address these issues, and the other two can be considered as "pre-strategic." They generate general capabilities or resources that, once applied to a specific problem, aid a firm in addressing one of the three types of strategic risks.

Value Creation Archetypes

The *Generation* archetype creates value by originating one or more data streams. These are developed by organizations that recognize (or stumble upon) events with high streamability. The events may have been streaming in the past,

FIGURE 5. Integrated View of the DDS Data Processing Framework

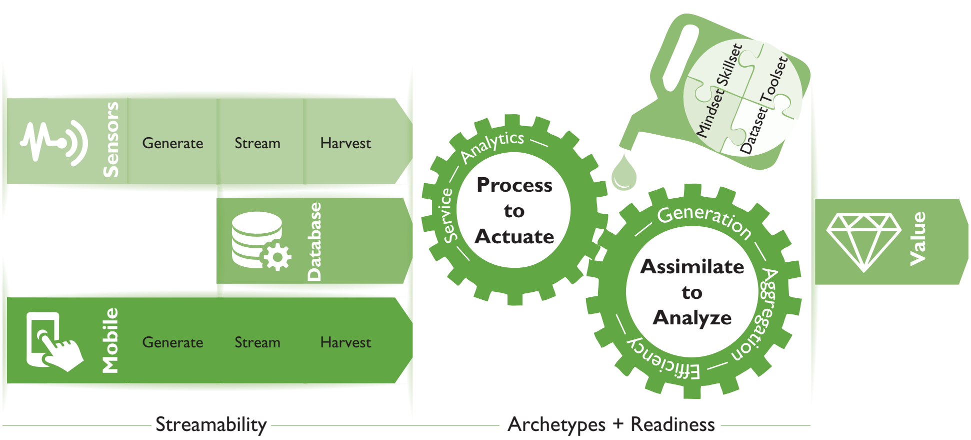


TABLE 3. Archetypes and Strategic Risks

Archetype	Strategic Issue	Meaning
Generation	Pre-Strategic	An investment in a resource that has not yet been applied to address a strategic risk
Aggregation	Pre-Strategic	An investment in a resource that has not yet been applied to address a strategic risk
Service	Demand Risk	The threat of wide swings in or overall declining demand
Efficiency	Inefficiency Risk	The danger of a higher operating cost structure than competitors
Analytics	Innovation Risk	The failure to create new products and services as quickly as competitors

but their value went unrecognized or, more often, their streamability has increased recently due to technology advances. The stream could arise as a byproduct of business operations (e.g., a sale) or as a deliberate action to capture data (e.g., RFID tags in gambling chips to track a player's bets). Firms that embrace the generation archetype realize (or think) that capturing these highly streamable events unlocks previously hidden value. The ensuing data stream can be made accessible to external partners or used internally. Generators aim at directly applying process-to-actuate for creating value. They take a potential data source, such as a sensor reading or a click on a web site, and convert it into a data stream element. Generation creates a resource, and thus can be considered pre-strategic until applied to a critical problem. Consider the early days of Twitter. The startup allowed people to share moments of their life with people who cared. The founders did not have a clear idea of how they would one day monetize their idea. Twitter is a good example of the generation archetype. This DDS captures humans' daily life, and it has proven valuable to organizations seeking to mitigate all three kinds of strategic risk. Service providers like Jet Blue, for instance, use the Twitter DDS to mitigate demand risk by keenly monitoring their customers. Other examples of this archetype range from startups, like Instagram and Foursquare, or major firms like the Ford Motor Company that generates real-time data on over four million vehicles through onboard sensors.

Aggregation occurs when a firm collects, accumulates, and repurposes digital data streams to create value through information services. Note that while some degree of aggregation of data is a prerequisite for all DDS initiatives, it represents the core value proposition for those organizations that fit the aggregation archetype. These organizations create platforms providing access to DDSs generated by other companies, thus exploiting the recognized value in the streamability of certain specific events with widespread value creation potential. This is a similar model to that of established financial data providers, such as Bloomberg or Thomson Reuters. Socrata takes advantage of the Open Data and data.gov initiatives to aggregate data streams generated by government agencies, and it makes them available to the public. DataSift coalesces real-time and historical data from the major social platforms and offers access to them through a common API or direct streaming to custom applications. A key value proposition of aggregators is to link the disparate data streams, in order to make them useful to customers seeking to moderate one of the three strategic risks. In these cases, it is up to the customer of the aggregator's service who extracts additional value from the DDSs.

Firms exploit the *Service* archetype when they merge and manipulate DDSs to provide new services or to improve existing ones. For instance, Disney's wireless-tracking wristband MagicBand[29] enables guests to unlock their hotel rooms, enter theme and water parks, check in at dedicated entrances, and purchase food and merchandise during the stay. The innovative MagicBand streams such events from simple RFID readings. The DDS allows Disney to collect valuable visitor data such as real-time location, purchase history, profile data, or ride patterns for popular attractions. MyTaxi puts taxi drivers and passengers in direct contact without requiring a traditional dispatch center. When a passenger hails a cab via the MyTaxi app, all taxis in the proximity (shown on the "Taxi Radar") receive the request. In turn, drivers can accept the call to pick up the traveler. Customers can personalize their preferences, review drivers, and pay for the trip via the app. MyTaxi links the cab's location DDS and passenger's transportation needs (a stream of events) to deliver a new value-adding service. The GPS sensor in smartphones increases the streamability of both the cab's and the passenger's locations making the business model feasible. The key insight here is that the service archetype emerges when the firm's value proposition consists of analyzing or processing DDSs with the intention of improving customer service, thus addressing demand risk.

The *Efficiency* archetype optimizes internal operations or tracks business performance using both internal and external DDSs. These organizations already recognize the value intrinsic in event streamability, within or outside their organizational boundaries. The efficiency archetype also applies assimilate-to-analyze, but the goal in this case is to harvest efficiency gains. Higher performance, lower prices, and cost savings are typical examples of how greater organizational efficiency translates into customer value. As the name implies, the efficiency archetype addresses inefficiency risk.

Coca-Cola's sensor-enabled Freestyle fountain drink dispenser,[30] as well as dispensing over 100 different flavors, gathers and reports consumption data for market analysis. Customers can choose from among these many flavors and drinks, and create and share (through a dedicated app) custom mixes. Thus, these new fountains are also a platform for experimenting with flavors, without a commitment to major bottling and marketing investments. Coca-Cola has an opportunity to tap into changes in beverage consumption as they occur and respond quickly.

The *Analytics* archetype processes DDSs to enhance decision making by producing superior insights through mechanisms such as dashboards, data mining, and visualization. To create value, these organizations leverage the increased streamability of events by innovating on their interpretability. This archetype focuses on the analysis end of assimilate-to-analyze to create value. It merges streams to create the breadth of information necessary for supporting high-level value-creation opportunities and thus reduce innovation risk. It typically has a higher value impact than the other archetypes because it can have implications for a larger number of products or customers, as the following examples illustrate.

Coca-Cola has a proprietary algorithm, Black Book, to standardize the flavor of Minute Maid orange juices, because the source of oranges can vary seasonally.[31] The model considers the taste of the source oranges, consumer preferences, and matches them with the attributes of each batch of raw juice. Several external DDSs ranging from satellite imagery to regional consumer preferences are combined for advanced supply planning.

Semantria is an example of a class of companies leveraging the analytics archetype. It performs sentiment analysis on DDSs. Customers can integrate the results into their applications and decision processes. Several companies are now offering similar DDS-based services, such as machine learning, neural networks, or the increasingly discussed cognitive systems.

Our review of firms that are innovating with digital data streams suggests that business model improvements and transformation with DDSs requires a new set of capabilities. Some firms opt to develop internally the skills and the knowledge necessary for value extraction, whereas others partner with firms specializing in harvesting data streams. Irrespective of the approach, DDS exploitation is an educational process. Firms find new opportunities by learning from those streams they currently exploit and then expanding their range and depth of exploitation. However, our work suggests that organizations differ widely in their ability to take advantage of the emerging DDS-rich environment. We have developed a framework, called the DDS Readiness Framework, to help organizations benchmark their potential ability to take advantage of strategic opportunities created by new DDSs.

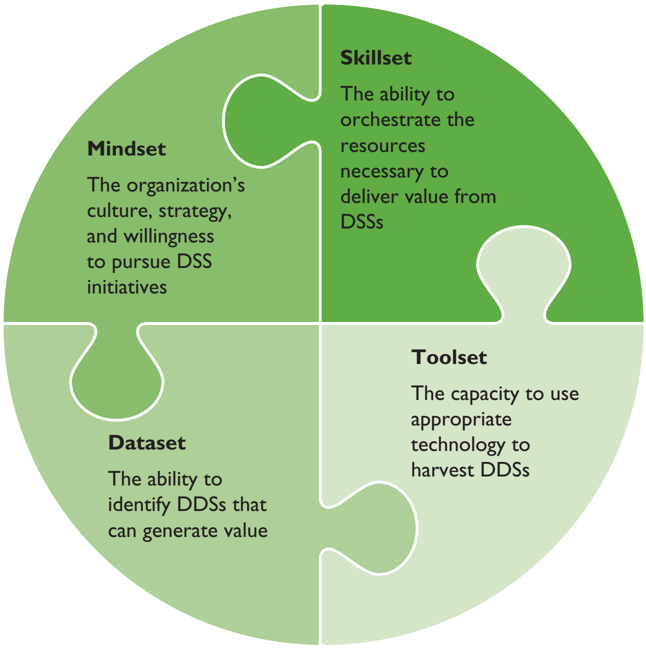
DDS Readiness Framework

The strategic opportunities introduced previously hinge on a firm's ability to build the capabilities necessary for their execution. Previous research has put forth four dimensions around which big data capabilities coalesce: Mindset, Skillset, Dataset, and Toolset (see Figure 6).[32] We have adapted these to develop and test a DDS readiness framework. The readiness framework components are the key resources and activities companies need to profit from a DDS. However, DDS readiness is just a potential, rather than the likelihood of the successful deployment of an initiative. Profiting from DDSs is not merely a problem of having the right "ingredients"—the right stuff—but instead demands astute management of their systematic interaction in an organizational setting.

Mindset

Mindset refers to an organization's willingness to invest in data-driven initiatives and assume the associated risks. It represents both an appreciation for the potential value of DDSs, as well as the keenness of the organization to pursue a DDS strategy through one or more of the value creation archetypes. A hallmark of the successful initiatives we have studied is the attitude of the organization to exploring DDS value potential, along with a strategic focus. Moving beyond DDS tactics entails significant risk, and a supportive cultural context for experimentation and development is critical. Consistent with previous research, we find

FIGURE 6. The DDS Readiness Framework Components



that business value is extracted from data only if a specific strategy is in place.[33] A data culture is manifested in attitudes concerning the relevance of data for decision making and the role streaming data have in understanding the business and appraising business opportunities created by DDSs.[34]

A change in mindset requires organizations to understand their DDSs, trust them, and organize their value creation architecture accordingly. This involves a change in established decision-making habits. A DDS strategy thus requires managers to consider the potential of real-time data streams to achieve business objectives. There is no value in the mere availability of a DDS (i.e., stalled in generation or aggregation) or strong data and analytical skills without a data-oriented mindset that can appreciate and envisage how to release the value of DDSs.

Skillset

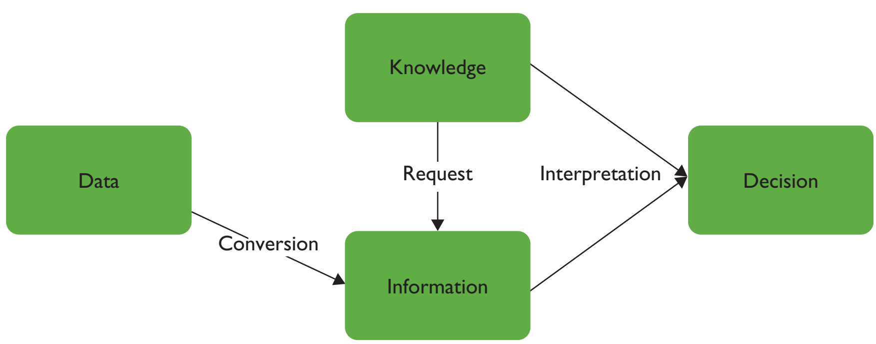
Skillset denotes the ability of an organization to manage DDS strategic initiatives. Deploying initiatives of the magnitude described by the value creation archetypes entails the acquisition and orchestration of all the resources, technical and complementary[35] to the DDS initiative, and assembling them[36] in the form of new products, processes, and decision-making routines.[37] A skillset is then manifested in the success in leveraging the strategic opportunities of DDS. It requires both the coordination of business and technological capabilities in order to act on all dimensions of streamability. Events have to be detected, measured, and interpreted before they can be a source of value. This requires strong coordination mechanisms among business functions and the ability to create and institutionalize new practices underpinning a specific capability of coordination within and across organizational boundaries.[38]

A skillset also incorporates the notion that an organization has the knowledge base to convert data into improved decision making. Knowledge is the capacity to recognize what information would be useful for making decisions. Knowledge also means that a person can interpret information and use it in decision making. Thus, the conversion of data to knowledge starts with a knowledgeable person requesting relevant information, followed by processing appropriate data to generate this information, assuming it is available. The resulting information is then interpreted by the original requestor, or others, to support a decision (see Figure 7).[39]

Dataset

Dataset is the capacity to effectively identify, intercept, and access real-time data streams that match organizational needs for value creation. In turn, effective access and use implies a sound understanding of the streamability factors discussed earlier, and how they relate to business needs.[40] For example, the decision to use a particular DDS for creating a new service needs to account for the specific characteristics of the event and the resulting stream, such as its intrinsic quality. Social media DDSs capture a multitude of events with different grades of streamability, and they are very "noisy" data sources that require a proper assessment and cleansing before being analyzed and integrated. Conversely, earnings calls and letters to shareholders are text DDSs of a much higher quality, without any, or minimal, sarcasm, irony, grammatical errors, or implied references to pop culture. Data coming from sensor network readings or RFID systems are intrinsically less "noisy," but may, for example, not measure all the desired event elements. These aspects of a dataset are generally reflected in a specific data and information governance configuration.[41]

FIGURE 7. The Relationship between Data, Information, and Knowledge



Toolset

Toolset represents the ability to use appropriate software and hardware to intercept a DDS and harvest its content. It is the most technology-oriented of the four DDS capabilities, encompassing both technical competencies and resources that are necessary to tap into streaming data. The toolset dimension influences directly the ability of a firm to profit from the increasing streamability of events. There are four main elements that, at the technical and architectural level, are required for operating a DDS:[42]

- Message-orientated middleware supporting standardized communication among heterogeneous systems.
- An advanced analytics engine to extract value from DDSs.
- A business process modeling engine that enables flexible and deep integration into an organization's human workflow, which significantly assists DDS value consumption and dissemination.
- A rules engine capable of executing business logic in real time and a related rules repository, separating the business and the application component, are required,

Some commentators suggest that firms follow an evolutionary path from "dataset to toolset to skillset to mindset."[43] So, the maturity of the organization should influence the relative importance of readiness components. In our research, we found an association between the extent of an archetype's exploitation and the levels of competency in mindset, skillset, dataset, and toolset. More specifically, the strategic and cultural dimensions captured in mindset significantly affect the value archetypes. The only exception is aggregation, which relies only on dataset and toolset. Mindset is the most significant antecedent of the analytics archetype together with dataset. The service archetype is associated with skillset and toolset capabilities, while efficiency relies on toolset and dataset. These findings support the notion that value exploitation based on analytics requires cultural, managerial, and data skills. However, this does not exclude the technological dimension; on the contrary, firms confirm that toolset is the asset on which their value strategy is built. Mindset and toolset are the core on which DDS initiatives are built. They are the current antecedents to DDS value creation.

A New Era of Value Creation

Organizations and ecosystems are both major forms of economic value creation. Organizations use information systems to integrate data across the enterprise so that they have a holistic view of customers, suppliers, and other key stakeholders. They also want a single-view of enterprise data to ensure high levels of cooperation among the various divisions. Marketing and production, for example, need to coordinate their actions to ensure the right products are in the right place in sufficient quantities. Thus, it is not surprising that many organizations deploy enterprise resources planning (ERP) systems to support an integration of business processes and data. ERPs are a product of the transaction era of information systems.

Integration requires capturing all transactions and storing the resulting data in a single enterprise data repository.

In contrast, an ecosystem consists of many interdependent and autonomous organizations that cooperate to create value. Microsoft, for instance, grew rapidly and profitably by creating an extensive partnership with many software developers who use its products to create and market applications for enterprises. Apple nurtures a large ecosystem of developers to create apps for its iPhone and iPad, which helps fuel the sales of both Apple's products and the associated apps. Apple claims the value of the app ecosystem will exceed $150 billion worldwide by 2018.

Independent organizations want to maintain control, especially over access to their data. Because of interdependencies, they might well see the value of sharing data, but they typically want to decide what, with whom, and when to share. Thus, a dilemma that many ecosystems face is that they want to collaborate to reduce costs, but individual organizations don't want to cede control of their data to create an industry-wide database. In this article, we make the case that DDS data processing opens new opportunities for value creation for those organizations that adapt to the changing environment. DDSs provide a foundation for improving ecosystem efficiency because they enable the controlled sharing of data. For example, there are many players in the shipping industry, each with high autonomy because of the history of seafaring. There is a need for episodic tight coupling, whereby the parties in an ecosystem coordinate their actions through the sharing of data about those episodes when they want to interact (piloting a ship into a harbor, using tugs to berth a ship, and loading and unloading a ship's containers). The two entities want to coordinate their actions for an episode or an event, and then go their separate ways. One of the shipping industry's efficiency concerns is that a ship can arrive outside a harbor and then wait several days for access to a berth. The ship's captain has wasted energy arriving too early. In an ecosystem that includes DDS aggregators, shipping companies could fully embrace an efficiency archetype allowing each captain to "right steam" the ship and arrive just in time. Lack of information leads to poor management of episodic tight coupling and wasted resources. If each of the participants in a forthcoming collaboration were to securely stream details of its current and likely future status to each other (e.g., a ship's expected time of arrival and a container terminal's berth availability), coordination would be enhanced.

The processing of data, manual or digital, has been a driver of value creation throughout the information age, and even before. Cost accounting emerged in Great Britain at the beginning of the industrial revolution because managers needed to determine the cost of their inputs in order to set profitable prices. For thousands of years, business has been conducted with limited data and limited capacity to process the available data. Data were usually lacking, and if they were not, they were often delivered too late or in forms too difficult to analyze in a timely fashion to influence value creation decision making. There was also a lack of processing capability to handle more than a small volume of data. Developments such as financial and cost accounting were initial attempts to extract value from data, but the major breakthrough came with the introduction of digital computers to business. Data availability, timeliness, and processing capability have gradually increased in the last half-century. DDSs are moving business to the point where all assets can continually report every state change (e.g., a parking spot shifts from occupied to unoccupied) and all events can be captured in digital format (e.g., the landing of a particular flight for a specified airline at a uniquely identified airport). While the first order effect of DDSs will be to improve real-time decision making, we can already foretell the second order effect of enabling greater real-time sharing of data within and across ecosystems. Data sharing between parties will facilitate greater efficiency in episodic tight coupling.

Twenty years ago, a low-cost public network, the Internet, introduced the era of mass consumer transactions and galvanized a value creation wave (e.g., Google, Facebook, iTunes, cloud computing). DDSs will have a similar impact, as early examples such as the emergence of the sharing economy demonstrate. Organizations should begin now to prepare for the new environment by embracing digital data streams.

Notes

1. Chris Myers, "Decoding Uber's Proposed $50B Valuation (and What It Means for You)," *Forbes*, May 13, 2015, <http://www.forbes.com/sites/chrismyers/2015/05/13/decoding-ubers-50-billion-valuation-and-what-it-means-for-you/>, accessed July 15, 2015.
2. M.P. McDonald and A. Rowsell-Jones, *The Digital Edge: Exploiting Information & Technology for Business Advantage* (Stamford, CT: Gartner, 2012); San Francisco Municipal Transportation Agency, "SFpark: Pilot Project Evaluation," June 2014, <http://direct.sfpark.org/wp-content/uploads/eval/SFpark_Pilot_Project_Evaluation.pdf>, accessed July 15, 2015.
3. S. Halzach, "Is the New Technology at Macy's Our First Glimpse of the Future of Retail?" *The Washington Post*, September 25, 2014, <http://www.washingtonpost.com/news/business/wp/2014/09/25/is-the-new-technology-at-macys-our-first-glimpse-of-the-future-of-retail/>, accessed July 15; 2015.
4. C. Smith, "How Beacons—Small, Low-Cost Gadgets—Will Influence Billions in US Retail Sales," *Business Insider*, July 8, 2015, <http://uk.businessinsider.com/beacons-will-impact-billions-in-retail-sales-2015-2>, accessed July 15, 2015.
5. Often the terms data, information, and knowledge are used interchangeably, but they are distinctly different. Data are raw, unanalyzed facts. Information is data that have been processed into a meaningful form for the problem at hand. Knowledge is the capacity to identify required problem-related information and interpret it.
6. G. George, M.R. Haas, and A. Pentland, "Big Data and Management," *Academy of Management Journal*, 57/2 (April 1, 2014): 321-326, doi:10.5465/amj.2014.4002.
7. G. Piccoli, J. Rodriguez, and R.T. Watson, "Leveraging Digital Data Streams: The Development and Validation of a Business Confidence Index," paper presented at HICSS, Kauai, HI, 2015.
8. T. Olanrewaju, "The Rise of the Digital Bank," *McKinsey & Company Insights* (July 2014), <http://www.mckinsey.com/Insights/Business_Technology/The_rise_of_the_digital_bank?cid=DigitalEdge-eml-alt-mip-mck-oth-1407>, accessed July 15, 2015.
9. K.J. Stiroh, "Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say?" *American Economic Review*, 92/5 (December 2002): 1559-1576.
10. Mazen Rawashdeh, "Bolstering Our Infrastructure," Twitter Engineering Blog, November 7, 2012, <https://blog.twitter.com/2012/bolstering-our-infrastructure>, accessed July 15, 2015.
11. Roberto Franzosi, "On Quantitative Narrative Analysis," in James A. Holstein and Jaber F. Gubrium, eds., *Varieties of Narrative Analysis* (Los Angeles, CA: Sage Publications, Inc., 2012), pp. 75-96.
12. Ibid.
13. Lexing Xie, H. Sundaram, and M. Campbell, "Event Mining in Multimedia Streams," *Proceedings of the IEEE*, 96/4 (April 2008): 623-647, doi:10.1109/JPROC.2008.916362.
14. Li Ding, Jie Bao, James R. Michaelis, Jun Zhao, and Deborah L. McGuinness, "Reflections on Provenance Ontology Encodings," in Deborah L. McGuinness, James R. Michaelis, and

Luc Moreau, eds., *Provenance and Annotation of Data and Processes* (Berlin: Springer, 2010), pp. 198-205.

15. Raffi Krikorian, a developer on Twitter's API/Platform team, provided in 2010 a now famous map of all the metadata present in a tweet showing how a tweet is much more than just 140 characters. See <http://www.scribd.com/doc/30146338/map-of-a-tweet>.
16. Yves-Alexandre De Montjoye, Cesar A. Hidalgo, Michel Verleysen, and Vincent D. Blondel, "Unique in the Crowd: The Privacy Bounds of Human Mobility," *Science Reports*, 3 (March 2013), doi:10.1038/srep01376.
17. Yves-Alexandre De Montjoye, Laura Radaelli, Vivek Kumar Singh, and Alex "Sandy" Pentland, "Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata," *Science*, 347/6221 (January 2015): 536-539, doi:10.1126/science.1256297.
18. Richard Hackathorn, "Current Practices in Active Data Warehousing," Bolder Technology, Inc., 2002, <http://assets.teradata.com/resourceCenter/downloads/WhitePapers/bolder_110102.pdf>; Greta L. Polites, "From Real-Time BI to the Real-Time Enterprise: Organizational Enablers of Latency Reduction," *ICIS 2006 Proceedings*, Paper 85, 2006, <http://aisel.aisnet.org/icis2006/85>.
19. Speaking at the LSU Analytics Speakers Series on January 21, 2015.
20. Nick Heudecker, Lakshmi Randall, Roxane Edjlali, Frank Buytendijk, Douglas Laney, Regina Casonato, Mark A. Beyer, and Merv Adrian, "Predicts 2015: Big Data Challenges Move From Technology to the Organization," Gartner, November 28, 2014.
21. For example, where the interest calculation for a deposit is modified to round down the cents and accumulate the excess to the thief's account (see <http://all.net/CID/Attack/papers/Salami2.html>).
22. Impressed, amazed, peerless, achievement, and unparalleled all score +1, and mediocrity scores -1. See, for example, this interesting R tutorial by Jeffry Breen to demonstrate the use of R in the context of text mining Twitter <http://www.slideshare.net/jeffreybreen/r-by-example-mining-twitter-for>.
23. Robert G. Fichman, "The Diffusion and Assimilation of Information Technology Innovations," in R.W. Zmud, ed., *Framing the Domains of IT Management: Projecting the Future through the Past* (Cincinnati, OH: Pinnaflex, 2000), pp. 105-128.
24. Morten Hertzum, Jørgen P. Bansler, Erling C. Havn, and Jesper Simonsen, "Pilot Implementation: Learning from Field Tests in IS Development," *Communications of the Association for Information Systems*, 30/1 (2012): 313-328.
25. Gabriele Piccoli and Blake Ives, "IT-Dependent Strategic Initiatives and Sustained Competitive Advantage: A Review and Synthesis of the Literature," *MIS Quarterly*, 29/4 (December 2005): 747-776.
26. Mutaz M. Al-Debei and David Avison, "Developing a Unified Framework of the Business Model Concept," *European Journal of Information Systems*, 19/3 (2010): 359-376.
27. Alexander Osterwalder, Yves Pigneur, and Christopher L. Tucci, "Clarifying Business Models: Origins, Present, and Future of the Concept," *Communications of the Association for Information Systems*, 16/1 (2005): 1.
28. J. Child, "Information Technology, Organizations, and the Response to Strategic Challenges," *California Management Review*, 30/1 (Fall 1987): 33-50.
29. For a description of the initiative, see <https://disneyworld.disney.go.com/plan/my-disney-experience/bands-cards/>.
30. See <https://www.coca-colafreestyle.com>.
31. See <http://www.businessweek.com/articles/2013-01-31/coke-engineers-its-orange-juice-with-an-algorithm>.
32. Andreas Weigend in Mike Barlow, "The Culture of Big Data," O'Reilly Media Inc., 2013, see <http://chimera.labs.oreilly.com/books/1234000001713/ch01.html#fitting_in>.
33. Thomas H. Davenport, Jeanne G. Harris, David W. De Long, and Alvin L. Jacobson, "Data to Knowledge to Results: Building an Analytic Capability," *California Management Review*, 43/2 (Winter 2001): 117-138.
34. Ibid.
35. Piccoli and Ives, op. cit.
36. Michael Wade and John Hulland, "The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research," *MIS Quarterly*, 28/1 (March 2004): 107-142.
37. Rajeev Sharma and Graeme Shanks, "The Role of Dynamic Capabilities in Creating Business Value from IS Assets," in *AMCIS 2011 Proceedings*, presented at the 17th Americas Conference on Information Systems, Detroit, MI, USA, 2011.

38. Ibid.; Constance E. Helfat, *Dynamic Capabilities. Understanding Strategic Change in Organizations* (Oxford, UK: Blackwell Publishing, 2007), p. 65.
39. R.T. Watson, *Data Management: Databases and Organizations*, 6th edition (Athens, GA: eGreen Press, 2013).
40. Shvetank Shah, Andrew Horne, and Jaime Capellá, "Good Data Won't Guarantee Good Decisions," *Harvard Business Review*, 90/4 (April 2012): 23-25; Michael S. Hopkins, "Are You Ready to Reengineer Your Decision Making? Interview with Thomas H. Davenport," *MIT Sloan Management Review*, 52/1 (Fall 2010): 1-6.
41. K. Weber, B. Otto, and Hubert Österle, "One Size Does Not Fit All-A Contingency Approach to Data Governance," *Journal of Data Information Quality*, 1/1 (June 2009): 1-27; Paul P. Tallon, Ronald V. Ramirez, and James E Short, "The Information Artifact in IT Governance: Toward a Theory of Information Governance," *Journal of Management Information Systems*, 30/3 (2013): 141-178.
42. Zubin Dowlaty, "Real-Time Data Streams: An Analytics Practitioner's POV," *Cutter Benchmark Review*, 12/3 (2012): 11-14.
43. Andreas Weigend as cited in Andrew McAfee and Erik Brynjolfsson. "Big Data: The Management Revolution," *Harvard Business Review*, 90/10 (October 2012): 60-68.

California Management Review, Vol. 58, No. 3, pp. 5–25. ISSN 0008-1256, eISSN 2162-8564. DOI: 10.1525/cmr.2016.58.3.5.