Data Collection, Analysis and Tracking in Industry

Daniel P. Bumblauskas
University of Northern Iowa

Herbert A. Nold
Polk State College

Paul D. Bumblauskas
PFC Services, Inc.

Current technologies allow organizations to collect vast amounts of data, creating the possibility of a data binge. More is not necessarily better as vast data pools make it difficult to convert data into actionable knowledge in a timely fashion. Effective data analysis is particularly important as product and service life cycles have shortened. The ability to analyze meaningful and relevant data then convert data to information and ultimately knowledge in time to drive decisions is a key competitive differentiator. We will explore reasons for this condition and offer practical suggestions for agile companies in the 21st Century business environment.

INTRODUCTION

Big data is an ambiguous term often associated with the collection and analysis of large datasets. A recent article in Forbes lists a variety of different definitions commonly used to describe what "big data," really means (Arthur, 2013). For the purpose of this article, data, information, and actionable knowledge are defined based on the published literature in this area. Analysis of data and action based upon that information is the key to the process of effectively utilizing big data. One particular aspect of data analysis that is needed in all businesses is accurate forecasting for sales, revenues, and production of goods and services. The Aberdeen Group discusses some of the challenges faced in this respect, one of which is having multiple or "no single" demand forecast (Aberdeen Group, 2004). One of the problems with forecasting remains groupthink as documented in Heizer and Render's account of jury of executive opinion forecasting method (Heizer & Render, 2013). Daniel Power references the work of Herbert Simon in saying: *"…the central problem will not be how to organize to produce efficiently, but how to organize to make decisions — that is, to process information. Big data means more processing of information and a greater need to organize to use the information in decision making* (Power, 2013)." Power, being a leader in Decision Support System theory, references seven reasons why managers do not maximize decision support systems, one of which is "*information overload*," which supports his reference to the work of Janis and Mann stating, "…*when the degree of complexity of an issue exceeds the limits of a person's cognitive abilities, there is a marked decrease in the adequacy of human information processing that is a direct effect of information overload and ensuing fatigue"* (Power, 2002). Sela and

Berger also discuss information overload in the context of getting weighted down by trivial decisions (Sela & Berger, 2011) and reference Jacoby et al.'s work which documented information overload back in the mid-1970's (Jacoby, et al., 1974).

In working with industry practitioners, we have identified that large databases have evolved into what is referred to today as "Big Data". The ability to assemble huge data sets can drive companies to make statements, such as "we can analyze anything," and often technology, particularly computer applications, allow us to have this ability. However, having this ability in an unstructured ad hoc basis can cause paralysis by analysis. This is particularly the case when one person or department creates the data and another must actually take action with the information provided. Access to more data does not necessarily lead to better decision-making.

In addition, there is risk associated with the availability of Big Data. Prior paper-based systems allowed for a physical internal control system. Today, automated data creates access risk inside and outside of organizations. Take, for example, the recent breach of controls associated with credit card information at Target and the notice that this is "*only the beginning"* (Miller, 2014), of such cybersecurity concerns with big databases. We are not sure yet whether a weak control environment caused this breach, but there is no question that the accumulation of such large data sets exacerbated the situation.

To draw some practical suggestions, we must first consider the evolution of literature in this area, after which recommendations are offered for ways to avoid being overwhelmed by data.

LITERATURE REVIEW

A chronological literature review has been provided in order to trace the evolution of big data historically. The evolution of big data begins with Fredrick Winslow Taylor and scientific management techniques in the early 1900s with his world famous work "The Principles of Scientific Management" (Winslow, 1911). Applying scientific management techniques required the accumulation and analysis of detailed work related data but was limited by the technology of the time. During the World War II and post-war era, the work of Deming, statistical quality control, the 14-point management method, followed by total quality management (TQM) during the 1940s & 1950s the need for data to support management decisions grew rapidly. Anderson et al. discuss this in their 1994 paper in the Academy of Management Review (Anderson, et al., 1994). The introduction of digital computers in the 1930s and 1940s (Burks, 1989) later led to mainstream deployment of computing technology during the 1970s to mid-1980s enabling the collection of vast quantities of data with limited ability to get information and actionable knowledge from the systems. During this period, it was also acknowledged that information overload was a problem in decision-making (Jacoby, et al., 1974). During the mid-1980s to mid-1990s the advance of distributed computing made data available across the organization and lead to the evolution of MRP and ERP systems. The Internet makes data available instantaneously across the organization, countries, and nations from the mid-1990s to present, driven by increased use of cloud based systems and applications (e.g., Microsoft Sharepoint & SkyDrive, Google Drive, etc.). In the late 1990's, the terminology "actionable knowledge," was introduced in both definition and meaning along with information theory based approaches (Cheng, et al., 1997).

METHODOLOGY: ACTIONABLE KNOWLEDGE

Data by itself has little value and knowledge without action has little value to organizations. Davenport and Prusak (Davenport & Prusak, 1998) offered a useful description of the differences between data, information, and knowledge suggesting that increased knowledge has the potential to improve decision-making. Understanding the differences between these three constructs and the transformational process of changing meaningless raw data into knowledge that drives action, as shown in Figure 1, is essential for success or failure in big data analytics.

FIGURE 1
DATA TO KNOWLEDGE CONVERSION PROCESS

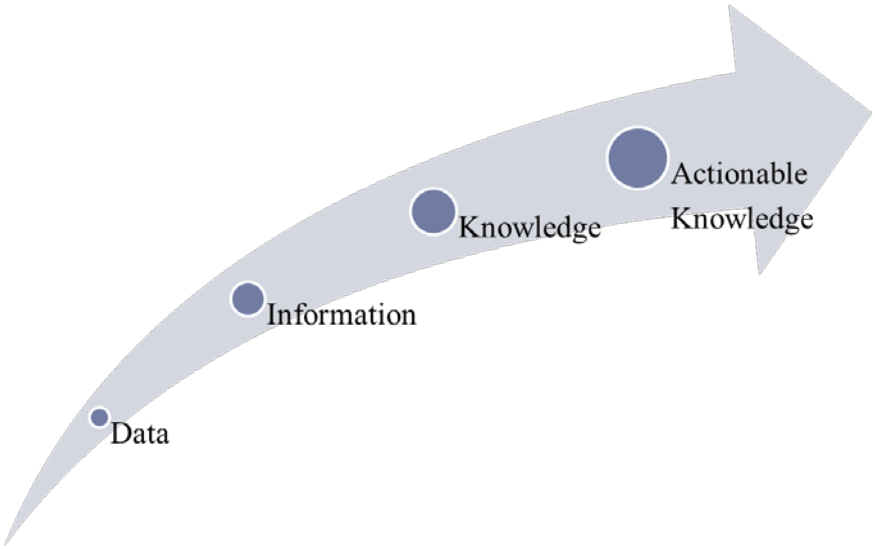

Figure 1 also provides an excellent and practical reference point for techniques that mitigate risk and create efficiency with the ultimate goal of identifying impactful ways to positively influence the organization. Therefore, we suggest that at each interface point, responsible parties should consider the controls, timeliness of movement between each conversion step and who and how impactful ideas can be prioritized and implemented. It is only through this process that knowledge conversion can be optimized.

Data consists of facts about some event with little relevance or purpose. Data without context or reference point has no meaning, but is essential for the creation of information. Humans give data meaning by adding context and reference points that are relevant and purposeful then communicating new information to a receiver. Interpretations by the receiver decide whether the information has value (Davenport & Prusak, 1998). Individuals then combine and synthesize multiple pieces of information to create a higher level of understanding that adds value through action. Davenport and Prusak defined knowledge as:

> *"A fluid mix of framed experience, values, contextual information, and expert's insight that provides a framework for evaluating and incorporating new experience and information. It originates and is applied in the mind of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms" (p. 5).*

The primary difference between data and information remains that data is collected and interpreted to help make informed decisions that drive action. Bellinger et al. defined the differences of data, information, knowledge, understanding, and wisdom (Bellinger, et al., 2011). The annual review of the International Journal of Knowledge, Culture, and Change in Organizations expressed the relationship between data, information, and knowledge as follows:

> "*Knowledge is the process of connecting the stuff of the mind and the stuff of the world. It is not a recorded thing (data, information), or at least, it is not just that. Knowledge is a form of action"* (Editors of International Journal of Knowledge, Culture and Change in Organizations , 2012).

The concept of actionable knowledge is not new, having been discussed extensively in various contexts including behavioral science (Argyris, 1996), business management (Argyris, 1993), organization science (Cross & Sproull, 2004), and other related disciplines. However, the concept of actionable knowledge has taken on additional meaning and importance in recent years with the vast expansion of data availability and the need for quick, effective, decision-making (Cao, 2012). Unlike academics who prize and are recognized for generating knowledge for the sake of knowledge, business executives are evaluated based on solid results that are outcomes of management decisions. Management decisions translate into results only if they generate action with positive outcomes.

Chris Argyris, who has been a key figure in popularizing the concept of actionable knowledge, defined actionable knowledge as: "*information that actors could use, for example, to craft conversations that communicate the meanings they intend. Actionable knowledge has to specify how to produce meanings but leave actors free to select the specific words" (p. 2)*, (Argyris, 1995). What this means is that "actors" who are decision makers must be able to derive meaning from data or information driving decision-making that can be translated into specific action to be communicated to others. Confronted with vast amounts of data, 21st century leaders must find those bits of data that provide information leading to actionable knowledge. This is no easy task because the environment is constantly changing. Essential elements needed to generate actionable knowledge include:

- Having valid and timely information
- Making informed choices
- Vigilantly monitoring of both the validity of input information and implementation of decisions (Argyris, 1995)

In a world that is characterized by rapid change, having valid and timely information is not necessarily an easy task. Information that may be valid, meaningful, and useful today may not be so a year or two years in the future. Changes in the internal or external environment or results from management decisions may, and probably will, make valid, meaningful, and useful information today of questionable value in the future. Therefore, it becomes critically important to constantly monitor both data and information input as well as output in the form of results and make adjustments as needs change. The result of not doing so will be sound decision-making based on flawed information. Actions based on invalid or flawed information may lead to unintended negative outcomes.

The process of interpreting data in context, transporting relevant information to key decision makers to take action based on the knowledge gained is critical to the success of organizations of all size, type, and industry. Lukas Michel (2013) described an intricate, dynamic, system in the form of a performance triangle consisting of culture, leadership, and systems. As shown in Figure 2 which was co-developed by the authors and Michel (Michel, 2013), the system is powered by people through relationships, collaboration, and purpose to drive success of the organization.

The process of effectively, efficiently, generating actionable knowledge from data becomes dependent on intricate and complex interactions of people working within the triangle. Using the analogy of a living organism, Michel suggested that a virus infecting the organization at any place would inhibit the flow of data and ultimately knowledge. Unseen viruses creep into an organization through in an infinite number of ways such as obsolete data gathering systems, capture and display of irrelevant data, a culture where people lack trust so do not share what they know, or leaders who use industrial age management practices with knowledge workers to name a few (Michel, 2013). The result illustrated in Figure 3 is less success due in part to ineffective decisions based on irrelevant, untimely or lack of information. Actions taken based on perceived but inherently flawed knowledge will rarely yield expected results.

FIGURE 2
THE PERFORMANCE TRIANGLE – FOR OPTIMAL PERFORMANCE

(*Adapted From* (Michel, 2013))

FIGURE 3
THE PERFORMANCE TRIANGLE WITH VIRUSES

(Adapted from (Michel, 2013))

METHODOLOGY: DASHBOARDS

With the explosion of data over the past 30 years, capturing the most critical data, formatting the data, and getting it in front of key decision makes has become increasingly difficult. Dashboards have become a popular way to make key data sets available to overloaded executives and managers at a glance. Each dashboard is unique for the customer and manager typically containing flashy gauges, charts, tables, meters, and graphics intended to draw the viewer's attention to key elements of the organization that might demand action. Different definitions exist for dashboards but after a lengthy search for a definition, Stephen Few (Few, 2006) developed a definition that seems to capture the unique essence of dashboards:

> *"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance" (p.34).*

While the concept of having all critical data displayed on one page so the manager can appraise the performance of the organization at a glance is appealing, most dashboards fail to communicate the right data efficiently and effectively (Few, 2006). This failure is not primarily due to inadequate technology but rather because the dashboard is poorly designed thereby not communicating essential information as effectively as the unsuspecting manager believes. Software designers, many times, become enamored with creating glitzy, flashy displays while failing to recognize the basic purpose of the dashboard as a tool for making actionable decisions. Once deployed, many cute displays lose their luster in a few days becoming annoying and forgotten.

Essential to the dashboard concept is idea of key performance indicators (KPI) that drive business performance. KPI proponents advocate identifying data that indicates operational effectiveness then cascading those or related performance indicators to succeeding lower levels in the organization. Fundamental to the effectiveness of the KPI approach is the belief that there is a cause and effect relationship with the KPIs and financial performance, which is not necessarily true in many cases. Choosing KPIs that have a cause and effect relationship, particularly for lower levels of the organization is difficult. Additionally, surveys indicate that on the average, organizations track nine times more KPIs than are actually needed. The reason for capturing so much data is that traditional approaches to KPIs follow a "more-is-better" philosophy (Battista & Shea, 2007).

As if choosing relevant KPIs was not difficult enough, consider that the world is changing at an ever-faster pace. Assuming that relevant KPIs are identified today and presented in an appealing visual format that stimulates management action, those same KPIs may not be relevant in the future. Relevant KPIs today may become irrelevant due to the management action that resulted or simply because of changes in internal or external environment. Either way, the need exists to constantly monitor and question the cause and effect and relevance of KPIs on a continuous basis to avoid the trap of making good decisions with bad data.

While dashboards can help with distilling data and help to create actionable knowledge, caution and cynicism is advised as time passes. Upon launch, dashboards are often coveted and embraced with excitement, with the exception of those resistant to change, but lose momentum as users start to become immune to warnings (e.g., flashing red indicators all the time, so who cares). Both presentation and relevance should be challenged periodically to verify that valuable information is being communicated in a way that leads to effective action. As Rick Warren observed "familiarity breeds complacency" which can be deadly (Warran, 2002). Questions that could be asked in the dashboard design process include:

- Dashboards are great in theory but do they work for our organization or unit?
- Is the return on investment worth the effort to compile the data?
- How will data be collected and how accurate is the data?
- Is any of the historical data flawed of poor quality?

- Are the people that need to see the data seeing it, processing the information, and making informed decisions?
- Can we trace and track that positive change is occurring with each item contained in the dashboard?

Dashboards are only one tool used in analyzing data to build information and knowledge. While there are some reports and papers on dashboard design (Few, 2006), specific applications such as software team productivity (Biehl, et al., 2007), etc. this is a rapidly growing and transforming field.

CONCLUSION AND FUTURE WORK

While actionable knowledge and dashboards are important theoretical constructs, it successful organizations must efficiently convert data to information to actionable knowledge. This process can vary from industry to industry and should be reviewed for practical execution on a case-by-case basis. In future work, we intend to explore the use of balanced scorecards and management by objective for applicability. In addition, we will address ways to prioritize alternative big data initiatives, including incorporating return on investment (ROI) concepts.

Final recommendations include ensuring that the data collected, acted upon knowledgably, and conveyed in dashboards are relevant, timely, and informative. Important considerations include when to refresh, what story the data tells, and whether the data can be used predictively as a forward looking indicator (e.g., to perform maintenance, etc.). The top two "Best Business Jobs," as noted by (U.S. News and World Report, 2014) were Market Research Analyst and Operations Research Analyst, both of which require the translation of data to information to actionable knowledge (using a dashboard or some other visualization tool).

If we assume that the availability of data drives our ability to interpret and consequently make decisions, we must further consider how the availability of massive amounts of data has evolved to facilitate more effective decision-making. In future work we plan to introduce the concept of relevant range theory which considers the changes which an individual or organization must under-go when operating outside of normal operating conditions. This would include changes required to operate in high growth or rapid decline scenarios such as the contingency planning undergone at organizations such as Caterpillar Inc. since the start of the Great Recession in 2008. When applied to data, the relevant range theory would indicate that when an organization or individual are below or above a "normalized" quantity or quality of data we are unable to properly interpret that data we must channel that data to the place where the data can be correctly interpreted. It is assumed, oftentimes improperly, that the recipient of the data has the time and ability to convert the data into information and actionable knowledge. With the absence of an appropriate filtering and directional system and when the effect of timing enters the equation, once outside the relevant range, the data diminishes in effectiveness to the point that it is useless.

One of the authors is often quoted as saying, "If knowledge is not applied or transferred is it really knowledge at all?" This statement is intended to address knowledge (data that has been converted to information but is as of yet unapplied) in people, computers and systems. Data that is not acted upon is trapped in the human mind (or computer) and is wiped clean, either at death (deleting the data on a computer) or by some other brain trauma (computer trauma). It is therefore critical that actionable knowledge by disseminated to all decision makers we anticipate being involved today and in the future at our organizations.

REFERENCES

Aberdeen Group, 2004. *The Demand Management Benchmark Report,* Boston: Aberdeen Group, Inc..

Anderson, J., Rungtuanatham, M. & Schroeder, R., 1994. A Theory of Quality Management Underlying the Deming Management Method. *The Academy of Management Review,* 19(3), pp. 472-509.

Argyris, C., 1993. *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change.* San Francisco: Jossey-Bass.

Argyris, C., 1995. Action science and organizational learning. *Journal of Managerial Psychology,* 10(6), p. 20.

Argyris, C., 1996. Actionable knowledge: Design causality in the service of consequential theory. *The Journal of Applied Behavioral Science,* 32(4), pp. 390-406.

Arthur, L., 2013. What Is Big Data?. *Forbes*, 15 August.

Battista, P. & Shea, T., 2007. From Policing to Performance. *Corporate Finance Review,* 11(4), pp. 31-35.

Bellinger, G., Castro, D. & Mills, A., 2011. *Data, Information, Knowledge, and Wisdom.* [Online] Available at: http://courseweb.lis.illinois.edu/~katewill/spring2011-502/502%20and%20other%20readings/bellinger%20on%20ackoff%20data%20info%20know%20wisdom.pdf [Accessed 1 January 2014].

Biehl, J., Czerwinski, M., Smith, G. & Robertson, G., 2007. *FASTDash: A Visual Dashboard for Fostering Awareness.* San Jose, CHI 2007 Proceedings • Distributed Coordination.

Burks, A., 1989. *The First Electronic Computer: The Atanasoff Story.* Ann Arbor: The University of Michigan Press.

Cao, L., 2012. Actionable knowledge discovery and delivery. *WIREs Date Mining Knowledge Discovery ,* Volume 2, pp. 149-163.

Cheng, J., Greiner, R. & Kelly, J., 1997. *Learning Bayesian Networks from Data: An Information-Theory Based Approach.* [Online] Available at: http://www.researchgate.net/publication/222667610_Learning_Bayesian_networks_from_data_An_information-theory_based_approach/file/5046351a60ed23f2fc.pdf [Accessed 1 January 2014].

Cross, R. & Sproull, L., 2004. More Than an Answer: Information Relationships for Actionable Knowledge. *Organization Science,* 15(4), pp. 446-462.

Davenport, T. & Prusak, L., 1998. *Working knowledge: How organizations manage what they know.* Boston: Harvard Business School Press.

Editors of International Journal of Knowledge, Culture and Change in Organizations , 2012. Annual Review. *International Journal of Knowledge, Culture and Change in Organizations ,* Volume 12, p. 7.

Few, S., 2006. *Information Dashboard Design: The Effective Visual Communication of Data,* s.l.: O'Reilly.

Heizer, J. & Render, B., 2013. *Operations Management.* 11th ed. s.l.:Pearson Prentice-Hall.

Jacoby, J., Speller, D. E. & Berning, C. K., 1974. Brand Choice Behavior as a Function of Information Load: Replication and Extension. *Journal of Consumer Research,* 1(1), pp. 33-42.

Michel, L., 2013. *The performance triangle: Diagnostic mentoring to manage organizations and people for superior performance in turbulent times.* London: LID Publishing, Ltd..

Miller, G., 2014. Cybersecurity experts warn Target data breach only the beginning. *Providence Journal*, 1 February, pp. 1-3.

Power, D., 2002. *Decision Support Systems: Concepts and Resources for Manager.* Cedar Falls: Daniel J. Power, pp. 43.

Power, D., 2013. *Creating a Data Driven Global Society.* Milan, SIGDSS Workshop on Reshaping Society through Analytics, Collaboration, and Decision Support: Role of BI and Social Media, pp. 1-11.

Sela, A. & Berger, J., 2011. Decision Quicksand: How Trivial Choices Suck Us In. *Journal of Consumer Research,* 39(2), pp. 360-370.

U.S. News and World Report, 2014. *Best Business Jobs.* [Online]
Available at: http://money.usnews.com/careers/best-jobs/rankings/best-business-jobs
[Accessed 2014 February 2014].
Warran, R., 2002. *The Purpose Driven Life: What on Earth Am I Here for?.* Grand Rapids, Michigan: Zondervan.
Winslow, F., 1911. *The Principles of Scientific Management.* New York and London: Harper & Brothers Publishers.