

# Analytics in the Pursuit of Knowledge: Adapting the Knowledge Pyramid

Ronald D. Freeze

University of Arkansas – Fayetteville

[rfreeze@walton.uark.edu](mailto:rfreeze@walton.uark.edu)

## Abstract

*Advances in storage leading to the Internet of Things (IOT) and Big Data has exponentially increased the Data aspect of the traditional Knowledge Pyramid – Data-Information-Knowledge-Wisdom (DIKW). This paper presents an adaptation of the Knowledge Pyramid as an Analytics Pyramid in which Time is posited to represent Wisdom as the pinnacle achievement when pursuing knowledge. Analogies of the DIKW are presented from the Analytics Pyramid as Description-Aggregation-Modeling-Time.*

*Implementing the premise of the Analytics Pyramid focuses on an iterative/repetitive movement of both individuals and organizations through all Description-Aggregation-Modeling-Time stages in order to build and obtain the Wisdom pursued in the traditional Knowledge Pyramid.*

*This model reinforces organizational learning and the importance of adaptability when pursuing knowledge. In addition, the wisdom gained from analytics is only recognized when monitored business processes are longitudinal in nature. Organizational analytics must rely on the recognition of a changing environment (Time) in order to adapt.*

## 1. Introduction

There are numerous articles citing the growth of data: 2.5 exabytes of data are created each day [1], the average company ... has more data stored than the Library of Congress [2], our universally available data is expected to exceed 8,000 exabytes by 2015 [3] and our digital universe of data will grow to 44 zettabyte by 2020 [4]. The growth of available data and the emphasis on extracting value from that data is echoed in the analytics expenditures in organizations and the expansion of analytic programs in universities.

Recent projections of expenditures on analytics within organizations include: business analytic software revenue increased from US\$17.5 billion in 2005 [5] to almost US\$35 billion in 2012 [6], global BI and analytics market will grow to USD\$20.8 billion by 2018 [7], big data and business analytics will grow to more than \$203 billion in 2020 [8] and the [big data] market will grow to \$92.2B in 2026 [9]. Analytic programs and the training of data scientists are rapidly expanding due to projected shortages of talent: big data

will need 4.4 million jobs by 2015 with only one-third being filled [10], U.S. may face a 50 to 60 percent gap in deep analytics talent [11], demand for employees in the analytics discipline surpasses the supply [12] only 26 percent of companies feel that their analytics needs are met [13] and business schools [need] to develop appropriate business analytics courses and programs for all majors [14]. With the rapid expansion and availability of data, a push to spend money to create value from the captured data and the recognition that greater talent is needed to realize value, the question arises on whether the predominant focus on the data is getting in the way of acquiring more knowledge. The question needs to be asked: How do we create knowledge through analytics, and what, in analytics, relates to knowledge?

The technological advances allowing the storage of big data lay the groundwork to more easily reveal relationships between the data captured. This capture of increasingly larger stores of data is being propelled forward by the expansion of the internet of things that allow both beneficial and questionable uses of the data from license plate readers, facial recognition and global positioning systems for example [4]. With respect to the Knowledge Pyramid (KP), the base (data) is expanding rapidly and is more easily connected. This does not necessarily allow faster decision making nor a building of greater knowledge. It seems obvious that increasing analytics skills and ability should allow a faster aggregation and movement to information, knowledge and wisdom. The goal of this paper is to focus on methods to increase the rate of knowledge creation by looking at a modified KP: The Analytics Pyramid (AP).

The following sections address the foundational aspects of the AP and the goal of knowledge creation within an organization. The first section will address organizational use of data for rapid decision making. Rapid decision making will be viewed through the analytic stages of the information value chain (IVC) and a knowledge management (KM) lens. Next will be a review of the historical KP along with some adaptations of the KP. The AP will be presented with examples from prior literature on how a view of the AP can benefit an organization. Finally, we will conclude with some research challenges and concluding remarks.

## 2. Rapid Decision Making

The focus in analytics has been the expansion of data capture in all technological advances. The result has been a focus on Big Data and the Internet of Things. Big Data analytics is directly tied to decision making [15] and organizations wish to capitalize on this expansion through rapid business decision making based on huge volumes of information [16]. This focus points to the unique and enduring purpose of the IS discipline: to understand and improve the ways people create value with information [17].

### 2.1. Analytic Stages

In order to create value with information, scholars and practitioners need to make sure they do not focus on the technical facets of big data to the exclusion of people and their institutional/social environments. This pitfall can create a lack of socio-technical harmony that IS implementation initiatives often need to succeed [18]. A recent review of the information value chain (IVC) provided sample research opportunities in the context of people-process-technology across the three dominant IS traditions of behavioral, design and economics [19]. The IVC is the cycle of converting data to information to knowledge in order to make decisions that initiates action. The five steps for the IVC are grouped into the categories of deriving knowledge (Data-Information-Knowledge) and decision making (decisions-actions) [15, 19, 20]. Data scientists and scripting-oriented programmers now perform knowledge activities that database managers and SQL programmers traditionally performed. While data scientists work closely with analysts and management in the knowledge derivation stage, there is a proliferation of real-time data-driven decision making that results in self-service analytics [21]. The rise of self-service analytics raises the question of whether the rapidly increasing pace possible for decision making is actually made based on completing the “deriving knowledge” stages of data-information-knowledge or decisions are made using only data (or information) with no application of knowledge.

With the advent of data visualization tools, the self-service analytic trend has enabled “nontechnical” users to “make effective use of data and reduce their time to insight” [22] with the assumption that the decisions and actions resulting from that insight can be termed knowledge. Datafication, making sense of big data in a complex world, has been termed a sensemaking process in order to derive value [23]. Sensemaking has also been indicated as a process to derive knowledge [19]. This would place value and knowledge on equal terms which contributes to the potential confusion of individual use of big data and self-service analytics. The complications of determining knowledge originating from the IVC expand further when other views are considered: 1) Individual consumers of data are able to make data-driven decisions at both the macro and micro level [24], 2) value has been

positioned as “speed to insight” and “pervasive use”. [25], 3) information intensity has been promoted as “more signals improve precision” [26] and 4) “unobtrusive” big data information sources (social media and Web clickstreams) facilitate realism [19]. While these are good attempts to measure the value of big data and the processes and tools designed to support analytics, the shortcoming is how knowledge is gained to support the long term use of analytics in longitudinal big data channels.

Big data has often been defined in terms of volume, variety, velocity and veracity – the four Vs. When applying these definitions, especially longitudinally, there is a disruptive effect that implicates the changing nature of knowledge gained through the analytic stages. Analytic tools allow patterns to be found in large volumes of data, but are business processes agile enough to recognize the change of pattern and implications of those changes. Organizations now deal with structured, semi-structured, and unstructured data (variety) from in and outside the enterprise [27], but has knowledge been confirmed by a consistent application of disparate data flows. The element of time, data in motion - velocity, impacts the acquisition of knowledge since patterns, insight and knowledge are now moving targets. Slow and fast moving data streams need to be joined to create situational awareness [28]. However, the merging of large volumes of data with disparate varieties and velocities is further complicated when credibility and reliability of those data sources vary (veracity). Together, the four V's represent the disruptive nature of the IVC [19] and indicate the difficulty in confirming that knowledge has been derived and appropriate action taken.

These issues invite the need for an increased focus on knowledge management. The new barrier has been how the data capture expansion rate can in reality be translated into knowledge and wisdom. The disruptive nature of the four Vs indicates that factual knowledge has a very short half-life [29]. A refocus on KM can improve what has been forgotten – improved knowledge from big data. Wisdom also adjusts based on any of the four Vs. An analogy to time-series speaks to adapting kernels of wisdom based on the modification of any V aspect. The traditional scientific model of research (constructing hypothesis and testing on carefully sampled data) needs to be modified to hypothesis designed for a stream of data with continual monitoring. Accurately predicting behaviors must be placed in time for greater predictive accuracy. The further in time the prediction is placed, the less accurate the prediction. Economist Herbert Simon once said, “A wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it” [2].

### 2.2. Knowledge Management (KM)

Knowledge Management activities, like most organizational activities, must deliver value to the firm

in order to procure a portion of the resources generated by the organizations success. KM initially followed the Technology-Push model, where knowledge could be captured, codified, stored and then transferred (sent) to a user in need of that knowledge. However, this approach failed to deliver business value to organizations [30]. The fundamental issue with transferring knowledge in this manner is due to the mechanistic, information-processing model [31] and is similar to a recommendation for improving business analytics value by building the foundation according to an information agenda [32]. The predominant focus on the data and information for both KM and analytics obscures and denies the socially constructed nature of knowledge [33]. Chief Knowledge Officers (CKOs) recognize the shortcomings of a technology-based KM strategy and have moved to a socialization-based KM strategy by recognizing the knowledge flow networks within their organizations [34].

Knowledge-intensive BPs confirm the relevance of, and the need for, a “Strategy-pull model of KM”. Essentially, “this model embodies organizational processes that seek a synergistic combination of data and information-processing capacity of information technologies and the creative and innovative capacity of human beings” (pg. 15) [30]. KM projects treated as IT projects will fail to recognize the social connection with knowledge. “One of the main reasons that knowledge management efforts are often divorced from day to day activities is that the people who design and build the systems for collecting, storing and retrieving knowledge have limited, often inaccurate view of how people actually use knowledge in their jobs” [35].

A recognition of merging knowledge (the knowledge supply chain) and business processes (the information value chain) is needed to extract full value from KM activities [36, 37]. The knowledge supply chain, similar to the IVC, is a sequence of related knowledge-based processes that together produce a product or service [38]. The process view of knowledge creation has been addressed in a number of research studies [39, 40, 41]. Analytics, similar to the history of KM, must not forget the process associated with knowledge creation and end up staying at the data/information level. Similar to treating KM projects as IT projects without involving users, treating analytics projects as IT projects without involving users does not recognize the differences in these approaches [18]. Where is the knowledge and is knowledge lost or gained when executing the organizational processes?

### 3. The Knowledge Pyramid

The KP provides a view of how the basic kernels of understanding can be combined into information, knowledge and finally wisdom (Figure 1). There have

been many papers citing the KP over the years [36, 42, 43, 44, 45, 46]. The initial KP has, at its base, Data which flows upwards to Information, then to Knowledge and finally to Wisdom as the pinnacle. Adaptations of the KP have consisted of reversing the flow of Data to Wisdom [43], a revised pyramid to reinforce organizational learning and the adaptation of knowledge management [45, 46] and inverting the pyramid to explain business process competitive advantage [36]. Each of these representations provide insight into the identification of knowledge but each also has shortcomings in how the accumulation and changing nature of knowledge is accomplished.

#### 3.1. A Reverse Flow K Pyramid

In the reverse flow KP, raw data does not exist and data emerges last only after knowledge and information are available [43]. Justification for this view, particularly in the context of information systems, resides in the meaning structure or semantics provided for the data that emerges or is captured. Data is not collected in a vacuum and therefore cannot be the building block for information, knowledge and wisdom. A human cognition cannot see simple facts without these facts being part of its current meaning structure. This view is reinforced by defining a conceptual model for a database. A specific location or data field in the structure is defined within the database and the value of the field may change but the meaning of the content is fixed. Data has been created by the description of this field. The field constructed to contain the data could not have been created without information/knowledge/wisdom as an antecedent. A further example is provided by a basic thermometer. The data (temperature) is determined by the instrument itself. The instrument was created using information/knowledge/wisdom. A thermometer is created with the possibility to observe temperature as data [43].

This line of reasoning prompts the question of where the wisdom originates. It is evident that the thermometer, a measurement device, has aspects of knowledge instilled into its structure. However, the meaning of the data is not determined by the instrument, context is not necessary for the information/knowledge/wisdom to exist and a thermometer is not created as the only possibility to observe temperature data. The temperature data was already in existence. The thermometer was constructed as a communication method between individuals and could be communicated via the temperature scales of Kelvin, Celsius, Fahrenheit or semantically with “it is really cold” or “really hot”. The context is provided as a method of transferring DIKW between individuals and not a means of constructing data. In an information system, captured data must have the

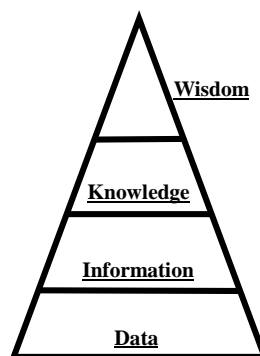


Figure 1 - Knowledge Pyramid

context for communication and explanation of coordinated events. It is equally important to recognize that missing data does not mean the data does not or did not exist. Missing data is simply data that was not captured in a manner in which it can be communicated. Whether a means of capture was provided or not, data exists first. This view is reinforced by the tacit/explicit view of knowledge.

Tacit knowledge is personal, context-specific, hard to formalize/communicate and explicit knowledge refers to knowledge that is transmittable in formal, systematic language [47]. Tacit knowledge therefore is analogous to the observation of temperature data with no thermometer as a cultural means of shared communication. System designers implicitly rely on the culturally shared meanings and the accumulated stocks of knowledge that have not necessarily been captured. The difficulty of capturing and accessing organizational knowledge has been explicated as organizational sensemaking. Sensemaking is defined as “an ongoing socio-cognitive activity that organizational actors initiate when seeking to understand and control their environment” [48]. When engaging in these activities, the sensemaker must approach the data as if the data were intended to mean something [43]. With the proliferation of data stored within an organization and the ability of the organization to connect to external data sources, flexibility to reinterpret any articulated relationship must be supported and practiced. Organizations must recognize that the flow of DIKW is not necessarily up or down but iterative. Movement between the DIKW levels must be enhanced in order to support organizational success. Confirmation of knowledge and wisdom is accomplished through an ongoing effort to confirm the data and information flows moving through an organization.

### **3.2. A Revised Knowledge Pyramid**

The revised KP [45, 46] brings in several critical aspects that are important to Big Data and the IOT. This extension reinforces the importance of organizational learning (OL) and the inclusion of KM activities in the generation of KM intelligence. At the base, data is represented as the results of sensors connected to “reality”. Social networks are positioned as the means to understand and move reality through the various DIKW stages. The social networks are taken in the broadest sense to include any communication which assembles data to information, information to knowledge or knowledge to wisdom. The primary emphasis of the reconceptualization is on learning which can be recognized as a change in behavior, expectation or enhanced organizational decision. An important aspect of the revised pyramid is that there is a bi-directional nature to the processing between the DIKW stages. The bi-directional nature reinforces that learning is not bottom up, but a continuous recognition of refinement of each stage of the pyramid. Information learned may not necessarily

impact knowledge immediately, but may be used to refine the data captured or make an adjustment of the sensors that are capturing reality. Emphasized in the revised KP is that organizations live in a dynamically changing landscape.

Similar to the reverse flow of the KP [43], context is necessary for OL to occur through the use of insight, analysis and sensemaking. The revised pyramid indicates an inversion due to the combinations of data that could make up information-knowledge-wisdom. These combinations are exponential in nature and can represent the expanding difficulty of organizations to coordinate the sensors and create the combinations necessary for organizational effectiveness. The filters in the revised pyramid [45, 46] were designed to “get the right DIKW to the right people at the right time”. However, the expansion of storage and the advent of in-memory computing will force a change in the movement of and methods of organizational storage of knowledge. Movement of DIKW will create waste in the organization. This waste occurs because when computers are used for knowledge management, they are primarily used as media for decontextualized communication, and not as tools for automatic data processing” [43]. However, the revised pyramid has also removed the apex which supports the notion of the lack of an ultimate point for an organization. The reality of expanding sources of data (sensors) confirms the notion that there is no final point for knowledge and wisdom and the goal of organizational effectiveness is not an end goal but an iterative repetitive process.

### **3.3. Inverted K Pyramid for Business Process**

Business process management (BPM) has historically been focused on obtaining competitive advantage for organizations through their core business processes. Knowledge management activities have been focused on achieving organizational goals and creating value through their stores of knowledge. An inverted KP was presented with a focus on merging these two streams of research through the recognition of knowledge-intensive business processes (KIBP) [37]. The historical nature of a business process had greater rigidity and consistency due to the defined nature of the processes. These were predominantly simple procedural processes. Attempts to identify and define within BPM more complex or very complex processes recognized KIBPs. These processes needed knowledge workers or experts and were hard to impossible to automate [49].

The inverted pyramid placed the creation of processes in the context of DIKW with the goal of creating a competitive advantage [36]. Similar to the revised pyramid, the inverted pyramid recognized the multiple combinations of data sources that could impact a business process and reinforced the concept that only a few of the combinations could result in competitive advantage. Competitive advantage was viewed as achieving wisdom for a business process, but achieving the level of wisdom for a business process

did not necessarily mean achieving wisdom. When constructing KIBPs in the context of the inverted pyramid, an exploratory approach was emphasized to identify that many combinations of the KIBP may need to be considered prior to confirming the process that achieves the highest place on the DIKW pyramid.

**4. The Analytics Pyramid**

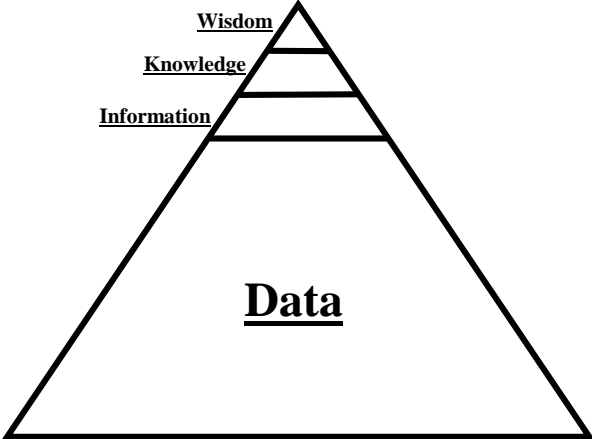
While each of the KP adaptations further the understanding of the DIKW flow, each does not emphasize the speed at which data is expanding and the impact this expansion has on the ability to achieve the IKW portion of the pyramid. There exists an imbalance within organizations with respect to the quantity of captured Data that impacts the efficient flow to Information-Knowledge-Wisdom. A KP that reflects reality would be closer to Figure 2 - Knowledge Pyramid Reality.

The imbalance identified in Figure 2 is a representation of the quantity of data being captured by organizations and emphasizes the imbalance of the usage of that data. The measurement scale is practically problematic and not addressed historically when assessing prior representations and adaptations of the KP. The prior KP representations were pictorially balanced and addressed only anecdotally the measurement scale aspects. If a measurement for the area of each segment of the pyramid were provided by the bytes of data representing each DIKW, the data aspect would potentially be even larger than represented. Whereas, the pinnacle of the pyramid, wisdom, may seem to be non-existent due to the fact that the wisdom generated by the flow through the DIKW may not be captured as explicit knowledge. The wisdom generated may reside implicitly in the minds of organizational leaders.

Conversely, stating the area for each DIKW as value to the organization, the representation of Figure 2 would be balanced improperly as the knowledge and wisdom sections of the pyramid should greatly outweigh the information and data sections of the pyramid. Success in an organization depend on the ability of the organizational leaders to extract knowledge and wisdom from the information and data sections. The addition of a single piece of data to an organization adds little to no value to the organization.

Even the aggregation of many pieces of data adds little value to the organization. Only when the movement to knowledge and wisdom occurs does value begin to be added to the organization. The goal of the AP is to address methods and processes that rebalance the difficulties presented by Big Data and the IOT.

The construction of an AP needs to take into account the ever

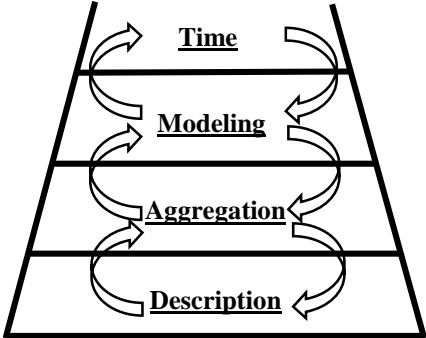


**Figure 2 - Knowledge Pyramid Reality**

growing store of data associated with Big Data and the IOT. For organizations, obtaining knowledge and wisdom in order to facilitate decision-making and improve organizational efficiency is important for success. With expenditures and the need for talent on the rise, a rigor is needed to ensure that decisions are being made on as high a level as possible on the KP. The AP is patterned after the DIKW flow to allow organizations to ensure that a complete analytical flow is followed to avoid missteps when applying analytical decision-making to their organizations.

As a short description of the AP, the base starts with Descriptions and is analogous to Data in the KP. The second stage is one of Aggregation and parallels Information. Modeling comes at the third stage and is where Knowledge begins to emerge from the application of Analytics in organizations. Decision-making, especially automated decision-making, should not be made at a level lower than Modeling to be a data-driven organization. Finally, Time must be taken into account as the fourth stage of the AP in order to parallel Wisdom from the KP. Each stage of the AP needs to contain a recursive process with the prior stage and the following stage. This is to insure that any Aggregation is properly supported by the appropriate Description and properly supports the Modeling of the third stage (Figure 3).

A further explanation of each stage is provided in the context of analytic examples provided in prior research. These examples include: 1) Netflix’s business model adaptation focusing on media recommendations [23], 2) GUESS’s creation of GMobile on the iPad platform [25], 3) Kroger’s Infra-red sensors used to reduce customer wait time [50], 4) Best Buy’s increased store revenue based on employee engagement and 5) Target’s highly accurate marketing efforts to a teenage pregnant daughter [51].



**Figure 3 - The Analytics Pyramid**

## 4.1. Description

Description, as the analytic pyramid foundation's initial stage, emphasizes the importance of understanding data structure prior to moving forward to subsequent stages. The importance of the Description stage is recognized by the generally accepted 80/20 rule of data analytic projects [52, 53, 54]. Essentially, for any data analytic project (data mining, predictive modeling, regression analysis, cluster analysis, etc.) 80% of the time is devoted to data preparation and 20% of the time is actually devoted to analyzing the data. There are projections that the data preparation time can be as high as 90% of a project's time [52]. Big Data and the IOT necessitate that data collection be structured accurately and clearly. However, as organizations look to supplement their internal data with external sources, insuring the consistent meaning of data that an organization did not collect directly is a requirement. Descriptions of the core building block for decision-making must be solid.

Adding external sources to organizational data is not the only source of variance addressed in data preparation. An organization's business model can be evolving through the creation of additional categories (for a single variable) or the creation of new variables necessary to enhance the value of the organization's offerings. The changes at Netflix from a more static business model to a streaming model indicate the importance of attention to the data structures – Description [23]. Netflix's initial disc rental model was focused around the subscriber queue and the disc content itself. This model required the management of the queue and not the disc itself. The selection of the disc by the customer was distant in time from viewing and so there was no feedback during viewing. As Netflix's business became a streaming model for the media content, the addition of viewing statistics required the careful preparation of the data structure to enhance the usefulness to a greater extent. The growth of data captured increased to more than a 1000 facets associated with each media title [23].

The success of GUESS's implementation of their GMobile app on the iPad platform is evidence of the importance of a standardized data model [25]. While GUESS allowed varying local business models and POS systems, they required regional ERP systems. There were three regional data warehouses in different countries (Asia, Europe and the U.S.) that used the same data model. That data model was built from individuals that "worked...at four different retailers". The data model was considered the "best of breed". The consistent Descriptions associated with a stable data model are credited with achieving a fast speed to insight for GUESS [25].

## 4.2. Aggregation

Knowing the Descriptions of the data captured can provide the information surrounding the Aggregation of each organizational variable. This stage represents the identification of issues associated with the data

preparation cleanup and is indicative of the first recursive loop between Descriptions of the data and the Aggregation of that data. This process is a cleanup procedure that is designed to separate the signal from the noise [52] for movement to the next stage of the AP – Modeling. At a basic level, the Aggregation stage is looking for data outliers by the utilization of basic statistical concepts: mean, mode, max, min, confidence intervals, frequencies, etc.

A multitude of issues can arise at this stage based on data Descriptions and the planning associated with moving to the Modeling stage. By way of example, identifying an outlier for a numeric variable of age can range from being relatively easy to requiring specific context of the intended model in order to determine the course of action. The simplest outlier may be when there is a "negative" age or an extremely large age of 150 years. These two situations could mean the removal of that data. However, a complete use of the AP would require that the data capture associated with Description be analyzed as to how these values entered the data set and potentially what data type was used for the capture of age (Date or number). Obviously accurate ages, like 35 years, could also be considered an outlier if the purpose moving forward to Modeling is a focus on college age students (typically between 18 and 25 years). Finally, questions focusing on children ages 0 to 5 years may not be granular enough for the purposes of the project. Bringing the age and month may be needed in order to obtain the results necessary for the business questions being asked.

Frequencies are a form of Aggregation particularly useful for understanding categorical variables. However, as organizations segregate their data to more granular categories, the question of data usefulness must be raised with respect to historical data. Netflix's movement to a streaming model highlights potential issues of comparison with respect to recommendations made over a long time period (queue management) versus a shorter time period (removing the time between recommendation and viewing).

## 4.3. Modeling

The Modeling stage of the AP begins the creation of knowledge for the organization and represents the point at which organizational decisions can be made. Ideally, the data preparation has been completed and the analysts are using clean data. However, the process of modeling could easily identify additional Aggregations or Descriptions that need to be addressed. This is a reinforcement of the recursive movement from Modeling to Aggregation and back. The Modeling stage is most easily identified with some kind of dependent variable that is associated with multiple independent variables. Examples include regression, ANOVA, clustering, decision trees, etc.

A significant question, associated with the Modeling stage that requires a review of Aggregation, is when to pool data. Issues of measurement equivalence can provide false positive indications if

data is improperly pooled. This issue can arise even between two apparently similar groups, such as top and middle management, where leadership, strategic planning, customer focus, information & analysis and process management are significantly different when predicting customer satisfaction [55].

Increasing employee engagement can demonstrate a significant increase in revenue for an organization. Best Buy found that a 0.1 percent increase in employee engagement resulted in a \$100,000 annual increase in revenue at a particular store [56]. Modeling uncovered this gem, but a full analysis of the model variables would need to be addressed prior to launching a program to increase employee engagement across the organization. Potential pitfalls include the interaction of other variables in the model (i.e. – regional average income, time frame of the data, employee demographics, etc.). There is also the issue of the measurement scale for employee engagement. Any model breaks down at the edges of the input variables. Once employee engagement is being measured close to the top of the scale (a 7 on a 1 to 7 Likert scale), the model no longer measures accurately. This indicates the danger of continuing to apply the knowledge learned when the knowledge has degraded. This is a common mistake in analytics where a metric is kept alive when there is no longer any business reason for the metric [56]. Adaptation in the Modeling stage is necessary to review appropriate variables through the Description and Aggregation stages, but a recognition of the need for the final AP stage of Wisdom when it becomes necessary to rebalance the Model.

The journey rebalancing their marketing Model for more effective decision-making became very apparent to Target when the father of a pregnant teenage daughter received pregnancy coupons in the mail [51]. Target was able to identify the teen was pregnant before she had told her parents based on her buying patterns. Target, however, had sent only baby coupons to the customers they identified as pregnant. The father, angry at Target for promoting that his daughter “get pregnant”, called and complained to the manager of the store. As it turned out, his daughter was pregnant. However, this incident lead Target to understand, increasing their knowledge, that even though their analytics (Modeling) was correct, they had failed to address the social impact (Time & Wisdom) of their campaign. Knowing things about individuals and sending them congratulations on “your first child” made them uncomfortable. Target “got sneakier” and mixed additional coupons with the baby item coupons to make the baby items look random. This again, increased the use of the baby coupon items. However, the question still remains as to whether Target moved to understanding the next stage of the AP – Time where they would also be applying Wisdom to their analytics. Is it ethical to use subterfuge when developing a marketing plan to make sales to your customers? This question brings into play the final stage of the AP of Time. Understanding causality in the application of

analytics creates the possibility of long term advantages for organizations.

#### 4.4. Time

Wisdom as the apex of the KP is analogous to Time for the AP. Organizations are no longer dealing with data that can be considered relatively static. Big Data and the IOT at their core are essentially data flows. In order to move from the Modeling stage of the AP, Time must be considered and organizations must recognize the changing nature of data collection and work towards models that are sampling data flows in order to detect change. This change can be placed in context of many organizational structures. The relationship to wisdom can be addressed through a more refined attempt to determine when an exact point of action is required. Achievement of wisdom, or the application of Time for the AP, is recognized when the use of models effectively takes into account both causality and the snap shots of a stream of data to recognize both changes in that stream and the need for modified decisions.

Causality, for improved customer service, was correctly and beneficially identified by the U.S grocery chain Kroger when they used overhead infra-red sensors to count customers and anticipated the number of currently needed checkout lanes in the next 30 minutes. Customer wait times were reduced from four minutes to 26 seconds [50]. The application of Time to the issue of customer service was anticipated and addressed to the benefit of the customer with little impact on individual efficiency of their employees. It should be noted here that while the application of Time was considered by Kroger, the result may have still only been a good Model. Knowledge was obtained but was there necessarily Wisdom associated with this implementation.

Target demonstrated a highly accurate model, but the progression of the marketing efforts indicate that Time (Wisdom) continues to elude the decision-making efforts. While the short term efforts of mixing coupons keep the customer from being creeped out [51], the subterfuge associated with the mixing of coupons that Target knows the customer targeted does not want may have additional causal effects to their sales. Ethical issues must be addressed when evaluating the impact of Time on the analytic processes. Situations for organizations that do not address this issue create circumstances where, as Netflix indicated, 75% of content choice is now influenced by recommendation and where Google is happy to match ads to content without ‘knowing’ anything about either [23].

The apex of Time for the AP has been removed and contain a similar aspect explicated in the revised KP [45, 46] and the inverted KP [36]. For the revised KP, the apex was removed to indicate that there is no end point for an organizations pursuit of knowledge. In the inverted KP, a business process may achieve an optimal level, but that did not mean the business

process was a competitive advantage. In both of these KP adaptations and the AP presentation, organizations must continually adapt and reconfirm their processes. The analytic process must continually move between the Model and Time to reconfirm the knowledge and wisdom has not been changed.

## 5. Research Challenges

The goal of this paper was adapting the KP to an AP to address the explosion of data that has resulted from Big Data and the IOT. The challenge in analytics is to know when an organization has achieved knowledge and/or wisdom when they have a good Model. The admonition is that Time must always be taken into account to determine the longevity of the decision and whether the Model can remain consistent through the organization's life. The application of the AP brings forth several research challenges that are addressed with the following questions at both the AP stages as well as their recursive interaction.

- ✓ *Overall – What are the measures for each AP (KP) stage that can help an organization most effectively iterate through the process of applying Time (Wisdom) to the pyramid?*
- ✓ *Descriptions – How to reduce the organizational resources devoted to data preparation in the context of internal and external data capture and data merging?*
- ✓ *Aggregation – What methods can be associated with the goal of pooling data to identify outliers and insure the accuracy of the Model inputs?*
- ✓ *Modeling – Can the actual knowledge elicited from a Model contribute to accurate decision-making for an organization?*
- ✓ *Time – Are the constructed Models consistent through time and what changes in the base level of the AP – Descriptions – contribute to a causal change?*

The overall research question associated with how to measure each section of the AP can address organizational effectiveness in their iterative movement through the AP stages. As discussed in section three, a “storage” vision vs. a “value” vision can provide two separate views of success. For Kroger's success using infrared cameras, each step of the AP and KP can be quantified. For data (descriptions), the individual sampling of each infrared camera can be quantify in terms of data storage. The value of the storage and energy required can also be quantified in terms of resource expenditure. At this stage, the expectation would be a very large data capture in terms of storage with a relatively small organizational cost in terms of monetary value. The information (aggregation) generated by the captured data (description) may have an associated sampling cost, with little to no data storage usage due to no capture of the samples generated. When the sampling process triggers the

knowledge (modeling) that calls for additional employees at the cash register, the trigger should be captured for review of the model. The wisdom (time) realized, “when the trigger for a call of additional employees to the register occurs”, provides a large value to the organization, but may have no use of organizational resources that is comparable to the data storage indicated with the data (description) stage. Even the value to the organization must have additional investigation as the result for Kroger's customers, measureable as reduced wait time at registers, does not indicate how this has increased value for Kroger's. The example here implies the need to further link relatively easy measures of data storage with the relatively difficult translation of customer satisfaction to the value associated with a company's success.

## 6. Conclusion

The presentation of the AP is intended to reinforce the parallel aspects of the analytic process with the construction of Knowledge and Wisdom in the KP. With Time at the top of the AP, success is never emphasized as an end point for an organization. The idea that the data flows are expanding reinforces the concept that a company must always reinvent itself in order to survive. As part of this reinvention, each stage of the AP has been positioned as recursive in nature with each prior and following stage. Measures are needed that reflect the value of each stage of the AP, but single measures will invariably emphasize one stage over another. A measure in terms of data storage will emphasize the data (descriptions) stage of the process, while value generated will emphasize the knowledge (modeling) and wisdom (time) stages. And some measures, such as customer satisfaction, must be linked with other more directly measureable store success measures. Therefore, multiple measures must be developed to fully explain each stage's contribution to the organization as a whole. Many organizations have obtained good Models (Knowledge) but the danger to avoid is implementing those decisions without evaluating how Time (Wisdom) effects the outcome.

## 7. References

- [1] McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," *Harvard Business Review*, pp. 59-68, October 2012.
- [2] McKinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity," McKinsey & Company, 2011.
- [3] D. Q. Chen, D. S. Preston and M. Swink, "How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management," *Journal of Management Information Systems*, vol. 32, no. 4, pp. 4-39, 2015.



- [4] K. E. Martin, "Ethical Issues in the Big Data Industry," *MIS Quarterly Executive*, vol. 14, no. 2, pp. 67-85, June 2015.
- [5] IDC, "Worldwide business intelligence tools 2009 vendor shares," Framingham, MA, 2010.
- [6] IDC, "IDC forecasts business analytics software market to continue on its strong growth trajectory through 2017," Framingham, MA, 2013.
- [7] L. James, "What's the one 2014 business intelligence prediction that all experts agree on?," 2014. [Online]. Available: <http://www.yellowfinbi.com/YFCCommunityNews-What-s-the-one-2014-BusinessIntelligence-prediction-that-all-experts-agree-on-154047>
- [8] G. Press, "6 Predictions For The \$203 Billion Big Data Analytics Market," 20 Jan 2017. [Online]. Available: <https://www.forbes.com/sites/gilpress/2017/01/20/6-predictions-for-the-203-billion-big-data-analytics-market/#6489cd062083>.
- [9] R. Finos, "2016-2026 Worldwide Big Data Market Forecast," 30 March 2016. [Online]. Available: <https://wikibon.com/2016-2026-worldwide-big-data-market-forecast/>
- [10] Gartner, "Big data creates big jobs: 4.4 million IT jobs globally to support big data by 2015," 2012. [Online]. Available: <http://www.gartner.com/newsroom/id/2207915>
- [11] McKinsey Global Institute, "Deep analytical talent: Where are they now?," McKinsey Global Institute, 2014
- [12] P. Solomon, "Demand for analytics skills outstrips supply in all sectors," 2014. [Online]. Available: <http://www.ft.com/cms/s/0/58d19eca-adb6-11e3-9ddc-00144feab7de.html#axzz4BDdBfj4d9>
- [13] N. D. Smith, "Technology-driven, must-have marketing skills," *DM News*, vol. 37, no. 10, pp. 20-22, 2016
- [14] O. Turel and B. Kapoor, "A Business Analytics Maturity Perspective on the Gap between business Schools and Presumed Industry Needs," *Communications of the Association for Information Systems*, vol. 39, no. 6, pp. 96-109, July 2016
- [15] P. Goes, "Big Data and IS Research," *Management Information Systems Quarterly*, vol. 38, no. 3, pp. iii-viii, September 2014
- [16] H. Chen, R. H. L. Chiang and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165-1188, December 2012
- [17] R. O. Briggs and J. F. Nunamaker, "Special Section: Creating Value with Information," *Journal of Management Information Systems*, vol. 28, no. 4, pp. 7-10, Spring 2012
- [18] D. A. Marchand and J. Peppard, "Why IT Fumbles Analytics," *Harvard Business Review*, pp. 105-112, January-February 2013
- [19] A. Abbasi, S. Sarker and R. H. L. Chiang, "Big Data Research in Information Systems: Toward an Inclusive Research Agenda," *Journal of the Association for Information Systems*, vol. 17, no. 2, pp. i-xxxii, February 2016
- [20] R. Sharma, S. Mithas and A. Kankanhalli, "Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations," *European Journal of Information Systems*, vol. 23, pp. 433-441, 2014
- [21] N. Chandler, B. Hostmann, N. Rayner and G. Herschel, "Gartner's business analytics framework," Gartner, 2011
- [22] D. Stodder, "Data Visualization and Discovery for Better Business Decisions," TDWI Research, 2013
- [23] M. Lycett, "'Datafication': Making sense of (big) data in a complex world," *European Journal of Information Systems*, vol. 22, no. 4, pp. 381-386, 2013
- [24] K. Gillon, S. Aral, C.-Y. Lin, S. Mithas and M. Zozulia, "Business Analytics: Radical Shift or Incremental Change?," *Communications of the Association for Information Systems*, vol. 34, pp. 287-296, January 2014
- [25] B. H. Wixom, B. Yen and M. Relich, "Maximizing Value from Business Analytics," *MIS Quarterly Executive*, vol. 12, no. 2, pp. 111-123, June 2013
- [26] X. Zhao, "Does Information Intensity Matter for Stock Returns? Evidence from Form 8-K Filings," *Management Science*, vol. 63, no. 5, pp. 1382-1404, May 2017
- [27] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales and P. Tufano, "Analytics: The real-world use of big data," IBM Institute for Business Value, 2012
- [28] M. Castellanos, C. Gupta, S. Wang, U. Dayal and M. Durazo, "A platform for situational awareness in operational BI," *Decision Support Systems*, vol. 52, pp. 869-883, 2012
- [29] O. Marjanovic, "Using the revised Bloom's Taxonomy to Scaffold Student Learning in Business Intelligence/Business Analytics," in *European Conference on Information Systems*, 2012
- [30] Y. Malhotra, "Integrating knowledge management technologies in organizational business processes: getting real time enterprises to deliver real business performance," *Journal of Knowledge Management*, vol. 9, no. 1, pp. 7-28, 2005
- [31] Y. Malhotra, "Why Knowledge Management Systems Fail? Enables and constraints of knowledge management in Human Enterprises," in *Knowledge Management Lessons learned: What Works and What Doesn't*, Information Today Inc., 2004, pp. 87-112
- [32] MIT Sloan Management Review and the IBM Institute for Business Value, "Analytics: The New Path to Value: How the Smartest Organizations Are Embedding Analytics to Transform Insights Into Action," Massachusetts Institute of Technology, 2010
- [33] S. Newell, M. Robertson, H. Scarbrough and J. Swen, *Managing Knowledge Work*, Palgrave, 2002
- [34] S. Dong, M. Johar and R. Kumar, "Understanding key issues in designing and using knowledge flow networks: An optimization-based managerial benchmarking approach," *Decision Support Systems*, vol. 53, pp. 646-659, 2012
- [35] J. Pfeffer and R. Sutton, *The Knowing-Doing Gap*, Boston, MA: Harvard Business School Press, 2000

- [36] O. Marjanovic and R. Freeze, "Knowledge Intensive Business Processes: Theoretical Foundations and Research Challenges," in *Hawaii International Conference on System Sciences*, 2011
- [37] O. Marjanovic and R. Freeze, "Knowledge-Intensive Business Process: Deriving a Sustainable Competitive Advantage through Business Process Management and Knowledge Management Integration," *Knowledge and Process Management: The Journal of Corporate Transformation*, vol. 19, no. 4, pp. 180-188, Oct/Dec 2012
- [38] H. S. Cha, D. E. Pingry and M. E. Thatcher, "Managing the Knowledge Supply Chain: An Organizational Learning Model of Information Technology Offshore Outsourcing," *MIS Quarterly*, vol. 32, no. 2, pp. 281-306, June 2008
- [39] Y.-J. Chen, Y.-M. Chen, H.-C. Chu and h.-Y. Kao, "On technology for functional requirement-based reference design retrieval in engineering knowledge management," *Decision Support Systems*, vol. 44, pp. 798-816, 2008
- [40] K. Linderman, R. G. Schroeder and J. Sanders, "A Knowledge Framework Underlying Process Management," *Decision Sciences*, vol. 41, no. 4, pp. 689-719, November 2010
- [41] I.-L. Wu and Y.-P. Hu, "Examining Knowledge Management Enabled Performance for Hospital Professionals" A Dynamic capability View and the Mediating Role of Process Capability," *Journal of the Association for Information Systems*, vol. 13, no. 12, pp. 976-999, December 2012
- [42] R. L. Ackoff, "From Data to Wisdom," *Journal of Applied Systems Analysis*, vol. 16, pp. 3-9, 1989
- [43] I. Tuomi, "Data is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory," *Journal of Management Information Systems*, vol. 16, no. 3, pp. 103-117, 2000
- [44] M. Fricke, "The Knowledge Pyramid: A Critique of the DIKW Hierarchy," *Journal of Information Science*, 2007
- [45] M. Jennex, "Re-Visiting the Knowledge Pyramid," in *Proceedings of the 42nd Hawaii International Conference on System Sciences*, 2009
- [46] M.E. Jennex and S.E. Bartczak, A Revised Knowledge Pyramid, *International Journal of Knowledge Management*, vol. 9, no. 3, 2013
- [47] I. Nonaka and H. Takeuchi, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford: Oxford University Press, 1995
- [48] M. Namvar, J. L. Cybulski and L. Perera, "Using Business Intelligence to Support the Process of Organizational Sensemaking," *Communications of the Association for Information Systems*, vol. 38, no. 20, 2016
- [49] P. Harmon, *Business process change*, vol. second edition, Burlington, MA: Morgan Kaufmann, 2007
- [50] A. Coolidge, "New technology helps Kroger speed up checkout times," 2013. [Online]. Available: <http://www.usatoday.com/story/money/business/2013/06/20/new-technology-helps-kroger-speed-up-checkout-times/2443975/>
- [51] K. Hill, "Forbes," 2012. [Online]. Available: <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>
- [52] M. Bhagde, "The Role of Data Preparation in Predictive Modeling," 28 July 2016. [Online]. Available: <https://www.merkleinc.com/blog/role-data-preparation-predictive-modeling/>. [Accessed 13 June 2017]
- [53] R. Perline and T. Woodfield, "Text Analytics Using SAS Text Miner Course Notes," SAS Institute Inc., Cary, NC, 2017
- [54] A. Palmer, "3 Reasons Why Analytics Projects Are Falling short - And What We Can Do About It," 12 October 2015. [Online]. Available: <https://www.tamr.com/3-reasons-why-analytics-projects-still-fall-short-and-what-we-can-do-about-it/>. [Accessed 13 June 2017]
- [55] M. Rungtusanatham, C. H. Ng, X. Zhao and T. S. Lee, "Pooling Data Across Transparently Different Groups of Key Informants: Measurement Equivalence and Survey Research," *Decision Sciences*, vol. 39, no. 1, pp. 115-145, February 2008
- [56] T. H. Davenport, J. Harris and J. Shapiro, "Competing on Talent Analytics," *Harvard Business Review*, pp. 53-58, October 2010