

Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

**Please provide feedback**

Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

## INTEGRATING DATA ANALYTICS & KNOWLEDGE MANAGEMENT: A CONCEPTUAL MODEL

Chaojie Wang, [cjwang@mitre.org](mailto:cjwang@mitre.org)\*

### ABSTRACT

*Data analytics may be heavily reliant on technology such as statistical models, machine learning algorithms, big data, and cloud computing; however, its outcome depends largely on human qualities such as experience, intuition, value, and judgement. Human knowledge is at the core of data analytics and knowledge management plays a key role in the analytics process. This paper uses the Data-Information-Knowledge-Wisdom (DIKW) hierarchy as an overarching structure to examine the end-to-end process of data analytics and to illustrate a conceptual three-phase data analytics process model integrating knowledge management practices including the discovery, creation, and application of knowledge. Nonaka's knowledge conversion theory is applied to the analytics process to shed light on the easily and often overlooked human and organizational aspects that are fundamental to the effectiveness of data analytics. The alignment and synergy between data analytics and knowledge management help foster collaboration, drive innovation, and ultimately improve outcome.*

**Keywords:** DIKW Hierarchy, Data Analytics, Knowledge Management, Knowledge Conversion

### INTRODUCTION

Many similar terms have been used to describe the process of identifying patterns and gaining insights from raw data to inform decision making and support problem solving. *Data mining* and *data science* are two terms commonly used in technical fields. *Business analytics* and *business intelligence* are two terms commonly used in managerial fields. *Knowledge Discovery in Databases* (KDD) is predominantly used in academia (Piatetsky-Shapiro, 2000). While these terms may differ in mechanics, they are similar in meaning. This paper adopts the generic term *data analytics* for its balanced coverage of both technology and business. We will adopt the definition from Davenport and Harris (2007) that data analytics is “the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions” (p. 7).

Unlike data analytics, the term *knowledge management* does not have many synonyms. However, it certainly has no shortage of diverse definitions. Sims (2008) compiled a list of over 50 definitions, the majority of which are categorized either as management-oriented or process-oriented. From the management viewpoint, knowledge management is a conscious effort directed by the management of an organization. By contrast, the process viewpoint characterizes knowledge management as a process of human activities and interactions, either coordinated by the management or performed autonomously by the people. A frequently-cited, management-oriented definition from the Gartner Group defines knowledge management as “a discipline that promotes an integrated approach to identifying, capturing, evaluating, retrieving, and sharing all of an enterprise’s information assets. These assets may include databases, documents, policies, procedures, and previously un-captured expertise and experience in individual workers” (Duhon, 1998, p. 12). This definition enumerates the various forms of knowledge assets which include both the tangibles such as documents and policies and the intangibles such as human expertise and experience. Scarborough, Swan, and Preston (1999) provides a process-oriented definition which defines knowledge management as “any process or practice of creating, acquiring, capturing, sharing and using knowledge, wherever it resides, to enhance learning and performance in organizations” (p. 669). This definition emphasizes the value of knowledge management which is to enhance organizational learning and performance. This paper will demonstrate that the intangible human experience and the organizational learning process of knowledge management are the essential ingredients for a successful data analytics effort.

\* The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author. Approved for Public Release; Distribution Unlimited. Case Number 18-2274. ©2018 The MITRE Corporation. ALL RIGHTS RESERVED.

Although data analytics and knowledge management are two different disciplines with different approaches, they share the goal of improving organizational effectiveness and performance. From an ontological perspective, data and knowledge are two related and inseparable constructs on the same spectrum of human experience (Skovira, 2007). A gold ore is the source of gold, and the gold is the refined form of a gold ore; thus, gold and the gold ore are one entity taking different forms at different lifecycle stages. The same dual nature can be ascribed to data and knowledge.

The remainder of this paper will start with the brief review of the Data-Information-Knowledge-Wisdom (DIKW) hierarchy and the concept of tacit vs explicit knowledge. These two constructs provide the theoretical foundation for the examination of the end-to-end data analytics process and its integration with knowledge management. Next, a conceptual three-phase process model for data analytics is proposed in terms of knowledge management practices (knowledge discovery, creation, and application). Lastly, the paper introduces Nonaka's knowledge conversion theory and applies it to the analytics process to illustrate the intertwined relationship between data analytics and knowledge management. The goal of this paper is to show that data analytics and knowledge management should be aligned and integrated to inform and reinforce each other so that organizational performance can be optimized.

### DIKW HIERARCHY

The DIKW hierarchy is a commonly-used construct representing the ascending nature of human understanding. It has many synonyms, such as information hierarchy, knowledge hierarchy, and wisdom pyramid. Rowley (2007) provided an extensive summary literature review of the structure and meaning of this construct.

Three definitions of DIKW from different perspectives have great relevance for the discussion of this paper as summarized in table 1.

**Table 1.** The Three Perspectives of Data, Information, Knowledge, and Wisdom

<b>DIKW Construct</b>	<b>Ackoff (1989) – Information process perspective</b>	<b>Davenport and Prusak (1998) – Knowledge management perspective</b>	<b>Zeleny (2006)</b>
<i>Data</i>	Symbols	A set of discrete, objective facts about events	Know nothing
<i>Information</i>	Data that are processed to be useful; provides answers to who, what, where and when questions	Data that makes a difference	Know what
<i>Knowledge</i>	Application of data and information	Derives from minds at work	Know how
<i>Wisdom</i>	Evaluated understanding	Higher-order concept lumped into knowledge for practical purpose	Know why

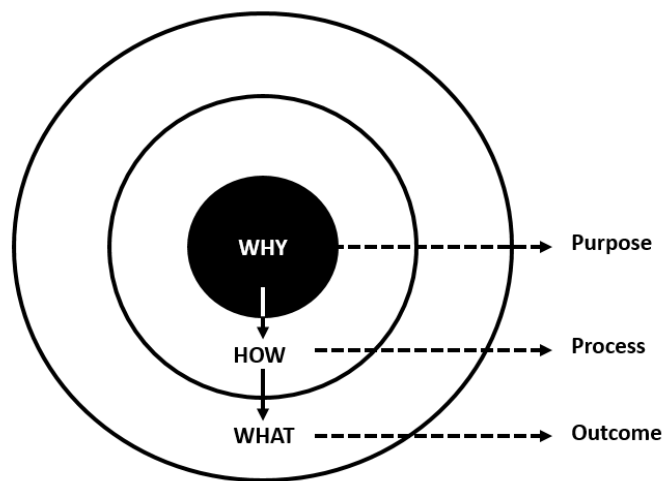
Ackoff (1989) defined DIKW from the information processing perspective, where data, consisting of symbols, are processed to produce information, which is then analyzed to derive knowledge and increase understanding. Data analytics follows this technology-oriented, information processing perspective, as is evident in the three leading analytics process models of CRISP-DM, SEMMA, and KDD (Azevedo & Santos, 2008; Piatetsky-Shapiro, 2014). The three models share similar limitations in that each was developed from a technical perspective and is for technical professionals; each concludes prematurely at the end of information processing; Each fails to extend into *knowledge creation*, a process of multidisciplinary collaboration of both technical and business experts, and into *knowledge application*, a process of engaging decisionmakers and stakeholders and applying knowledge to problem solving.

Davenport and Prusak (1998) defined DIKW from a knowledge management perspective, where data are “facts about events” (p. 2), information “has meaning” (p. 4) and “must inform” (p. 3), and knowledge “derives from minds at work” (p. 5). They provided a more elaborate definition of knowledge:

*Knowledge is a flux mix of framed experiences, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds 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)*

This definition goes beyond the technical mechanics of information processing and places knowledge in the context of human and organization as reflected in the key words used in the definition: “*minds*”, “*knowers*”, “*experience*”, “*insights*”, “*organizations*”, and “*organizational*”. Interestingly, Alavi and Leidner (2001) echoed that sentiment by stating that “knowledge is information processed in the mind of individuals. It is personalized information” (p. 109).

Zeleny (2006) defined the DIKW hierarchy in the most succinct and yet most profound way. According to this definition, data, information, knowledge, and wisdom represent four different types of human knowledge and different levels of human understanding. Data are like unearthed materials from a mine, which disclose nothing, and hence have limited value. Information is like coarsely processed mine materials which convey *what*. Additional “w” words like *who*, *when*, and *where* apply here as well. Knowledge is information further refined by human minds, which reveals *how*. Know-how is the most practical and useful type of knowledge. Wisdom is to know *why*; it sits at the pinnacle of human understanding. Wisdom is the most powerful and critical type of knowledge that drives human endeavors as reflected in the “Start with Why” Golden Circle model by Sinek (2011) where the *why* drives the *how* to achieve the *what* as illustrated in figure 1.



**Figure 1.** The “Start with Why” Golden Circle Model (Sinek, 2011)

## **POWER OF TACIT KNOWING & TWO TYPES OF KNOWLEDGE**

The recognition of the importance of personal experience in the pursuit of knowledge can be traced back to the scientist-turned philosopher Polanyi (1958), who was critical of the positivist notion of science and believed that knowledge is personal and requires personal commitment. Polanyi (1966) believed that “our body is the ultimate instrument of all our external knowledge, whether intellectual or practical” (p.15); emphasized the power of tacit knowing by distinguishing tacit knowledge from explicit knowledge; and believed that all knowledge is rooted in the tacit knowledge.

Polanyi (1966) defined tacit knowledge as “a knowledge that we cannot tell” (p. 5). Tacit knowledge constitutes most of human knowledge which is embodied within human bodies and minds and cannot be explicitly described or

codified. For this reason, Polanyi (1966) started his discussion on tacit knowing by famously stating that “we can know more than we can tell” (p. 4).

Explicit knowledge is the human knowledge that can be codified or described using languages and other conceptual means such as drawings, signs, gestures. Polanyi (1966) also used the term practical and intellectual/theoretical knowledge to differentiate the tacit and explicit knowledge. Tacit knowledge is practical while explicit knowledge is intellectual or theoretical.

Nonaka and Takeuchi (1995) contrasted the two types of knowledge from several aspects, as seen in table 2.

**Table 2.** Comparison of Tacit and Explicit Knowledge Adapted from Nonaka and Takeuchi (1995)

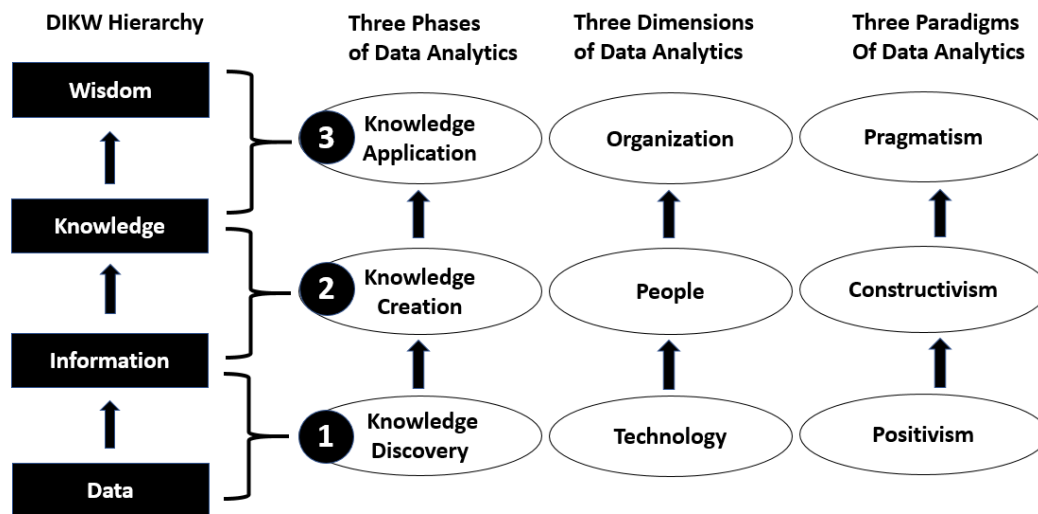
Tacit Knowledge	Explicit Knowledge
<ul style="list-style-type: none"> <li>• Subjective</li> <li>• Knowledge of experience (<i>body</i>)</li> <li>• Simultaneous knowledge (<i>here and now</i>)</li> <li>• Analog knowledge (practice)</li> </ul>	<ul style="list-style-type: none"> <li>• Objective</li> <li>• Knowledge of rationality (<i>mind</i>)</li> <li>• Sequential knowledge (<i>there and then</i>)</li> <li>• Digital knowledge (theory)</li> </ul>

Skovira (2012) viewed tacit knowledge as more aligned with the Eastern philosophical tradition, manifesting itself in the practices of the Zen Buddhism, whereas explicit knowledge is more aligned with the Western philosophical tradition, which manifested itself in the scientific methods influenced by the positivist worldview.

### THREE PHASES OF DATA ANALYTICS PROCESS

While there has been much hype about the technical capabilities of data analytics due to advances in technology like big data, cloud computing, and machine learning, the people and organizational aspects of data analytics are often overlooked and neglected. The DIKW hierarchy and the concept of tacit knowing provide a systemic framework to better understand the intricate end-to-end data analytics process.

The data analytics process is like climbing the DIKW pyramid, and can be summarized in three major phases. Each phase corresponds to one of three dimensions (technology, people, and organization), and represents one of three philosophical paradigms (positivism, constructivism, and pragmatism) as shown in Figure 1.



**Figure 2.** The Data Analytics Process Aligning with the DIKW Hierarchy

**1. Knowledge discovery phase (from data to information)**

During this phase, the raw data from various sources are blended and processed using statistical models and computational algorithms to derive useful information or analytics, which represents newly discovered explicit knowledge through the application of technology.

While this phase includes important human activities such as the initial project planning, the understanding of business needs, and the validation of data quality, it is mostly technology-oriented and involves the heavy use of statistics, machine learning, and software engineering. This phase represents the application of the positivist worldview where objective truth of knowledge is sought using scientific methods possessing quantitative, deterministic, and reductionist characteristics (Creswell & Creswell, 2017).

**2. Knowledge creation phase (from information to knowledge)**

During this phase, the resulting analytics are consumed and further processed by human minds which interpret the analytics using individual experience, insight, and the existing body of knowledge. As stated by Alavi and Leidner (2001), “information is converted to knowledge once it is processed in the mind of individuals” (p. 109). This phase also involves collaboration by experts from multiple disciplines to evaluate the explicit knowledge discovered from the previous phase. New knowledge, mainly in the form of tacit knowledge is created during this process.

This phase is mostly people-oriented and represents the application of the constructivist or social constructivist worldview where subjective truth of knowledge is created or constructed through social interaction, human collaboration, and sharing of experiences. The constructivist method of inquiry is qualitative and inductive in nature (Creswell & Creswell, 2017).

**3. Knowledge application phase (from knowledge to wisdom)**

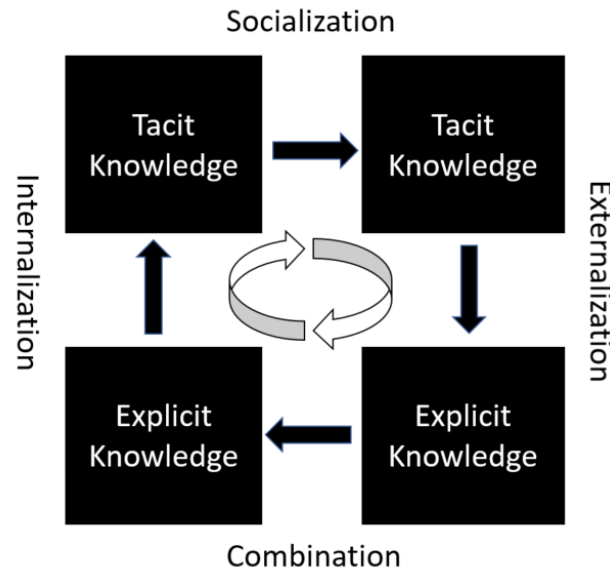
During this phase, decisionmakers and stakeholders are involved, and the new tacit knowledge created from the previous phase is assessed against organizational values and business strategies. The application of knowledge leads to informed decisions and plan of actions aimed at improving situations and solving problems. The resultant decisions and actions can be viewed as explicit knowledge stemming from collective wisdom of experts, decisionmakers, stakeholders, and anyone else involved.

In this phase, which is mostly organization-oriented, organizational values and business strategies are the key considerations. This phase represents the application of the pragmatist worldview that emphasizes the practical application of knowledge (Creswell & Creswell, 2017). This phase takes the explicit knowledge discovered through the positivist approach, and the tacit knowledge created through the constructivist approach, and applies them in practical problem solving.

The three phases of the data analytics process align well with the three eras of knowledge management: Era one, leveraging explicit knowledge; era two, leveraging experiential knowledge; and era three, leveraging collective knowledge (Dixon, 2018).

## **FOUR MODES OF KNOWLEDGE CONVERSION**

Nonaka and Takeuchi (1995) drew upon both the concept of tacit knowledge from Polanyi (1966) and the tradition of Eastern philosophy and incorporated them together in the knowledge management process. This was contrary to the Cartesian dualism of subject and object or body and mind, which has been the basis for scientific methods. The Eastern philosophy stresses “oneness of humanity and nature” (p. 27), “oneness of body and mind” (p. 29), and “oneness of self and other” (p. 31). This unity provides the basis for the authors’ theory of knowledge creation known as SECI (Socialization, Externalization, Combination, and Internalization). The SECI model describes the dynamic and continuous conversion process between tacit and explicit knowledge resulting from direct human interactions and collaboration. The authors define *knowledge conversion* as the process in which “human knowledge is created and expanded through social interaction between tacit knowledge and explicit knowledge” (p. 61). The SECI model presents four modes of knowledge conversion in a spiral process as shown in figure 2.



**Figure 3.** Knowledge Conversion Process Adapted from Nonaka and Takeuchi (1995)

The four modes of knowledge conversion are summarized in table 3 using the clear and succinct definitions from Nonaka, Toyama, and Konno (2000).

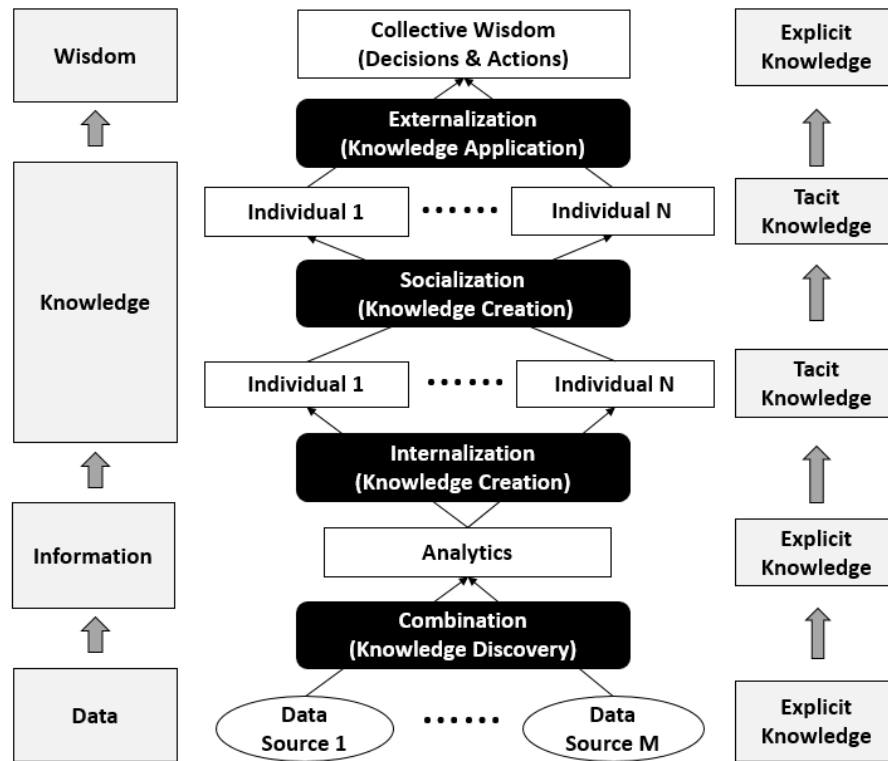
**Table 3.** The SECI Model

Conversion Mode	From and To	Definition (Nonaka et al., 2000)
Socialization	Tacit to Tacit	“Socialisation is the process of converting new tacit knowledge through shared experiences” (p. 9).
Externalization	Tacit to Explicit	“Externalisation is the process of articulating tacit knowledge into explicit knowledge” (p. 9).
Combination	Explicit to Explicit	“Combination is the process of converting explicit knowledge into more complex and systematic sets of explicit knowledge” (p. 9).
Internalization	Explicit to Tacit	“Internalisation is the process of embodying explicit knowledge into tacit knowledge” (p. 10).

### APPLYING KNOWLEDGE CONVERSION TO DATA ANALYTICS

If we look at data analytics in a holistic way, we will be able to see that Nonaka’s knowledge creation model can be applied to shed light on its intricate process. Here again we use DIKW hierarchy as a convenient construct to show the three phases of the data analytics process, and map them to the four modes of knowledge conversion as illustrated in Figure 3.





**Figure 4.** Applying the Knowledge Conversion Model to Data Analytics

The DIKW hierarchy is a convenient conceptualization of knowledge which encompasses data, information, knowledge and wisdom. We consider data as explicit knowledge since they represent codified, explicit measures of properties of entities or events of interest. This can be applied to any type of data including structured (tabular), semi-structured, and unstructured (textual, audio, video, etc.). Analytics resulting from the application of statistical models and computational algorithms are considered information which is also considered explicit knowledge.

#### 1. Knowledge discovery through combination

The use of statistical models and computational algorithms to uncover trends and patterns of data from various sources represents the technical dimension of the data analytics process. It is the process of knowledge combination by systemizing data (explicit knowledge) from diverse sources to form a broader base of explicit knowledge of the problem at hand. This is the process of combination.

#### 2. Knowledge creation through internalization

After the technical step has been completed, the process moves to the people dimension, where both technical and business professionals take the combined explicit knowledge and interpret it based on their professional experiences and intuitions and internalize them into tacit knowledge. This is the process of internalization.

#### 3. Knowledge creation through socialization

The process continues with team collaboration, where individuals communicate and collaborate to share their tacit knowledge that they just gained from internalizing the explicit knowledge discovered from data. This collaboration produces new and enhanced tacit knowledge for everyone involved. This is the process of socialization.

#### 4. Knowledge application through externalization



The final process is for the team to brainstorm and come up with shared understanding of the problem and potential solutions. This process also involves the engagement of decisionmakers and stakeholders and the incorporation of their perspectives. The collective wisdom in the form of tacit knowledge from all participants leads to the decisions and actions which are considered explicit knowledge. This is the process of externalization.

Data analytics is a process of increasing understanding of reality, starting from the observations of a phenomenon to reaching the pinnacle of wisdom, empowering informed decisions and practical actions to influence or change the phenomenon. However, the real world is complex, dynamic, and ever changing. Data analytics with the goal of informing decision-making and supporting problem solving is not a onetime effort. The data analytics process is incremental and iterative in nature. The whole process of knowledge discovery, creation, and application continues in a spiral, and forms the basis for continuous process improvement, much like the Deming Cycle of Plan-Do-Check-Act (PDCA) (Tague, 2005).

### SUMMARY

Data analytics is an indispensable tool to help people overcome information overload, help organizations improve performance, and help society tackle social problems and improve human conditions. To achieve better outcomes and higher impacts, professionals and organizations should look beyond just technology, stress the need for engagement and collaboration, and emphasize social values and organizational strategies as critical forces driving data analytics. The concept of the DIKW hierarchy, coupled with the power of tacit knowing and the theory of knowledge conversion, provides an illuminating theoretical framework that integrates the three phases (knowledge discovery, creation, and application), the three dimensions (Technology, People, and Organization), and the three paradigms (positivism, constructivism, and pragmatism) of data analytics into a meaningful whole.

The theoretical analysis and conceptual model presented in this paper contribute to the body of knowledge by increasing the understanding of the data analytics process and its relationship with the knowledge management theory and practice. Practitioners can use the insights from this paper to guide their data analytics efforts to improve collaboration and outcome. Educators can use the insights to guide their data analytics program and course design to improve instructional and pedagogical quality. Academics can conduct further research to validate and improve the proposed theory and model.

### REFERENCES

- Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16(1), 3-9.
- Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1), 107-136.
- Azevedo, A., & Santos, M. F. (2008). *KDD, SEMMA and CRISP-DM: A parallel overview*. Paper presented at the IADIS European Conference Data Mining.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.): SAGE Publications, Inc.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*: Harvard Business Press.
- Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*: Harvard Business Press.

- Dixon, N. M. (2018). The three eras of knowledge management. In J. P. Girard & J. L. Girard (Eds.), *Knowledge management matters: Words of wisdom from leading practitioners* (pp. 19-47). Macon, GA: Sagology.
- Duhon, B. (1998). It's all in our heads. *Inform*, 12(8), 8-13.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*: Oxford university press.
- Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba and leadership: a unified model of dynamic knowledge creation. *Long range planning*, 33(1), 5-34.
- Piatetsky-Shapiro, G. (2000). Knowledge discovery in databases: 10 years after. *ACM SIGKDD Explorations Newsletter*, 1(2), 59-61.
- Piatetsky-Shapiro, G. (2014). CRISP-DM, still the top methodology for analytics, data mining, or data science projects. Retrieved from <http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>
- Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. London: Routledge and Keon Paul.
- Polanyi, M. (1966). *The tacit dimension*. London: The University of Chicago Press.
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of information science*, 33(2), 163-180.
- Scarborough, H., Swan, J., & Preston, J. (1999). *Knowledge management - The next fad to forget people?* Paper presented at the Proceedings of European Conference on Information Systems, Copenhagen.
- Sims, R. (2008). Holistic: KM definition evaluation. Retrieved from <http://wiki.guruj.net/Holistic!KM+Definition+Evaluation>
- Sinek, S. (2011). *Start with why: How great leaders inspire everyone to take action*. New York, NY: Penguin Group.
- Skovira, R. J. (2007). Ontological grounding of a knowledge mapping methodology: Defining data, information, and knowledge. *Issues in Information Science*, VII (2), 258-264.
- Skovira, R. J. (2012). *Japanese way, western way: Two narratives of knowledge management*. Paper presented at the Management, Knowledge and Learning International Conference 2012.
- Tague, N. R. (2005). *The quality toolbox*. Milwaukee, Wisconsin: ASQ Quality Press.
- Zeleny, M. (2006). From knowledge to wisdom: On being informed and knowledgeable, becoming wise and ethical. *International Journal of Information Technology & Decision Making*, 5(04), 751-762.