Integrated Battle Command System, shown here on 1 December 2023 at Redstone Arsenal, Alabama, is the foundation of the Army's broader modernization efforts and provides transformational air and missile defense capabilities to the battlefield. (Photo by Nathaniel Pierce, Program Executive Office Missiles and Space)

Goldilocks Kill Chains and the Just Right Data

Maj. Michael G. Dunn, U.S. Air Force

The Department of Defense (DOD) faces a crucial challenge in achieving its goal of joint all-domain operations due to the inability to achieve foundational, fast-paced evolution of data storage, management, and analytics exemplified in the commercial sector. A shift began around 2000, when the commercial industry began outpacing defense in technological advancement, primarily because of its adaptable data strategies and computing capacity.

This analysis emphasizes the significance of data processing in achieving cost-effective kill chain development for joint all-domain operations, given its requirement for complex operations across multiple domains. It differentiates between big data, which necessitates complex machines for comprehension, and small data, which humans can understand naturally. Furthermore, it draws parallels between commercial and military operations, using the data-information-knowledge-wisdom (DIKW) pyramid as a decision model.

The analysis proposes the adoption of object-based storage to address the challenges of cross-domain data integration and presents a framework based on the DIKW pyramid, illustrated by an analogy of

rivers, streams, reservoirs, waterfalls, and lakes. This framework demonstrates how adopting commercial data strategies, particularly object-based storage, can enable the DOD to leverage data from various sources, enhancing knowledge for tactical and operational decision-makers. In essence, this research underscores the urgency for the U.S. government and DOD to embrace commercial data practices to facilitate advanced cross-domain algorithms, empowering decision-makers with a deeper understanding of complex situations and more effective decision-making capabilities.

Garbage In, Garbage Out

Information is the oil of the 21st century, and analytics is the combustion engine.

—Peter Sondergaard[1]

In 1640, John Graunt recorded the first use of the English word "data" while trying to provide the first description of data analytics.[2] The first calculations to create facts or types of data occurred as early as 19,000 BCE.[3] Since the seventeenth century, data has continuously expanded in complexity and application from agriculture to medicine to defense. The defense industry remains on the forefront of novel ways of applying data and decision-making formulas including the speed of data transmission; for example, pathways changing from a scouting party to a telegram to radio to computer speeds. Gordon Moore predicted in 1965 that the number of components per integrated function will increase at a logarithmic rate as technologies advance and the cost per component decreases in what is commonly referred to as "Moore's law."[4] The relationship of computing availability and complexity between the commercial and defense sectors has inverted since the 1960s.[5] In the 1960s, the military had the clear advantage of access to higher performance computing, leading the way in application of computer technologies to problem-solving. Today, the commercial-to-defense computer power application has inversed, leading to the amount of computing that occurs in the commercial world far outweighing that of a single system in the military.

As a prime example, a Tesla vehicle with full self-driving capability has roughly 180 times the computing power of an F-35 fighter.[6] While tactical edge-based computing, such as in an aircraft, a vehicle, or a handheld radio, must continually strive to increase safe and reliable computing in order to disaggregate computational locality and complicate enemy targeting, things the commercial world worries little about, the military can find advantage in the inverse relationship by focusing on commercial applications of data analytics. More data generation does not necessarily equal better decisions, and in the cost curve of acquiring new computing technology, the military can drastically increase its use of current data sets to enable decision space.

Defining commercial data, strategies, and dichotomies is necessary to determine what commercial advances in data analytics should find applicability in the defense sector. Moreover, this section presents a commercial viewpoint of so-called data layers to include transitioning from data to usable products or decisions. The etymology of the word "data" gives an insight into its formation. Data is the "plural form of the Latin word 'datum,' which means the 'thing given.'"[7] Classically used, datum is "a fact given as the basis for calculation in mathematical problems."[8] A data set, singular, expresses a block of data and allows for classification in generalities, such as big or small data. Generalizing things into data sets does not allow for proper data understanding, classification, curation, and management without acknowledging the individual datum types inside the larger data set. Thus, to keep things simple, data sets allow for the application of data strategies, but big or small data, when classified, constrains to the operational or tactical use of said data strategies.

Big versus Small Data—FIGHT!

All data are blocks of facts in whatever size, shape, storage location, etc., and further attempts to separate the fact that big and small data are more than just

Maj. Michael Dunn, U.S. Air Force, is a member of the deputy chief of staff's studies group at Headquarters Air Force. He holds a BS in systems engineering-computer systems from the U.S. Air Force Academy. During his career, Dunn served with the 15th Reconnaissance Squadron, the 432nd Operational Support Squadron, the 26th Weapons School Squadron, the 17th Attack Squadron, and the 30th Reconnaissance Squadron. He also attended an Air University yearlong fellowship at the Defense Advanced Research Projects Agency.

A Tesla vehicle with full self-driving capabilities has roughly 180 times the computing power of an F-35 fighter. (Photo courtesy of Tesla)

classifications creates unevenness in arguments about data management. However, regardless of data purism or an etymology faux pas, keeping the separation between seemingly big or small data allows for targeted application of strategies, concepts of operations, and concepts of employment. The primary differentiation between big or small data derives from the measurement of four characteristics called the four Vs of data: (1) volume, (2) velocity, (3) variety, and (4) veracity.[9] Each "V," in and of itself, could drag a data set from small to big classification. Simply defined: (1) volume is "the amount of data," (2) variety is "the diversity of sources and types of data," (3) velocity is "the speed of data transmission and generation," and (4) veracity is "the accuracy and trustworthiness of the data."[10] A fifth "V," value, creates additional utility by providing an answer to the "why" question for businesses to apply information management techniques.[11] Unfortunately, value for a business model vice value for a military application creates an argument between subjective and objective value (value of decisions made versus dollar value); thus, this analysis abstains from applying the value classification.

From a simpler perspective, the business world simply classifies small data as "small enough for the human to comprehend both in terms of volume and format" and big data as "chunks of data that are too large and complex to be analyzed and processed by traditional data-processing techniques."[12] In order to classify what a human can process, one must assume that the human received training and a competency level in processing said data. In a reductionist example, an electrically optimized (EO) sensor, such as a daytime television camera, produces video imagery that a trained human can process and make decisions from. In contrast, a farm of EO daytime television and infrared cameras would create such a complex picture across multiple modalities (infrared and EO), including multiple sources, that a single human would struggle to process the raw imagery in a near instantaneous timeline.

Enter the New Model: DIKW

Why are data important? While data in and of itself are interesting, data generation for the sake of data generation should never be the end goal. Data

Two F-35 Lightning IIs bank after receiving fuel over the Midwest 19 September 2019. While tactical edge-based computing must strive to increase safe and reliable computing to disaggregate computational locality and complicate enemy targeting, things the commercial world worries little about, the military can find advantage in the inverse relationship by focusing on commercial applications of data analytics. (Photo by Master Sgt. Ben Mota, U.S. Air Force)

must have a downstream effect, and the effect it provides is wisdom to make a correct action. Therein lies the question, how does one get from data to action? Data analysts in the commercial world use an action pyramid model called the data-information-knowledge-wisdom (DIKW) pyramid (as depicted in figure 1), which starts with the foundational data layer, builds to an information layer, again onto a knowledge layer, and finally, ends with wisdom.[13] Action produced from the layers of knowledge and wisdom implies that the person or entity that consumes the wisdom generated from data brings predefined or pretrained institutional decision matrices that when married with wisdom produces the proper output. While the DIKW pyramid was introduced in the early 2000s in the information technology sector, the true beauty of it derives from its simplicity.[14] Because it is simple, the DIKW is data categorization agnostic, meaning it could apply to both big and small data. In a small data example, a person or computationally small computer—in this case, think tactical systems—could organically derive the information from the gathered data either by preprogrammed filters, algorithms, or human intuition, bring its own knowledge of the situation, and finally, make an action. Simplicity in data,

system, and algorithms equals a reduction in timeline for processing and decision-making.

The "data" layer is the foundation of the DIKW pyramid, the beating heart pumping raw facts into the action model. Assume that for the generic action, without data, the action model collapses. In 1989, Russell L. Ackoff, an organizational theorist, defined data and information:

> Data are symbols that represent properties of objects, events, and their environments. They are products of observation. To observe is to sense. The technology of sensing, instrumentation, is, of course, highly developed. Information, as noted, is extracted from data by analysis in many aspects.[15]

Suffice to say, the raw facts of a situation, environment, or other observations form the data layer.

Information builds the next layer of the DIKW pyramid. Think of information as the answer to questions one might have about the data. The questions could drive specific answers or inferred answers that combine multiple pieces of data to hypothesize and answer the question. The question could also drive additional functions accomplished on the data to derive an answer. Consider the following examples and explanations of precise and derived data. If a data analyst queries for a specific person's birthdate or social security number from a list of attendees to an event, the analyst extracts precise information. If, instead, the data analyst wants to know the average age of everyone who attended, the analyst would have to first make sure everyone on the list attended, maybe by querying an attendance binary, and then execute an averaging function across the complete list of ages. This simple example expresses a few critical relationships between data and information. To derive information requested, the data set queried must contain the exact or derivable data requested; conversely, information could also reveal what is not contained within the data, including correlations of datums. Data not contained within the set queried helps find relationships between different data sets helping to derive answers or reveals a data structure problem.

Any data analysts trying to optimize information extraction must first analyze the relationship of information requests to data structure. According to the *Encyclopedia of Big Data*, "Data can be classified as structured, semi-structured, and unstructured based on how it is stored and analyzed."[16] Structured data is organized data, typically "in a strict format of rows and columns."[17] Semistructured data is a separate form of structured data, but because of its nature, whether raw or strict, it does not have an "underlying data model, hence cannot be associated with any relational database."[18] Finally, unstructured data, the most common type, has "no conceptual data-type definition," and the content is typically stored in some unique to the generating system type file, for example, a smartphone picture, a webpage, or a multispectral image.[19]

As described within the DIKW section of the *Encyclopedia of Big Data*, "As data sets increase in both structured and unstructured forms, analysis and management get more diverse."[20] In the commercial sector, multiple diverse types of networked storage and other wide-ranging technologies or techniques exist to
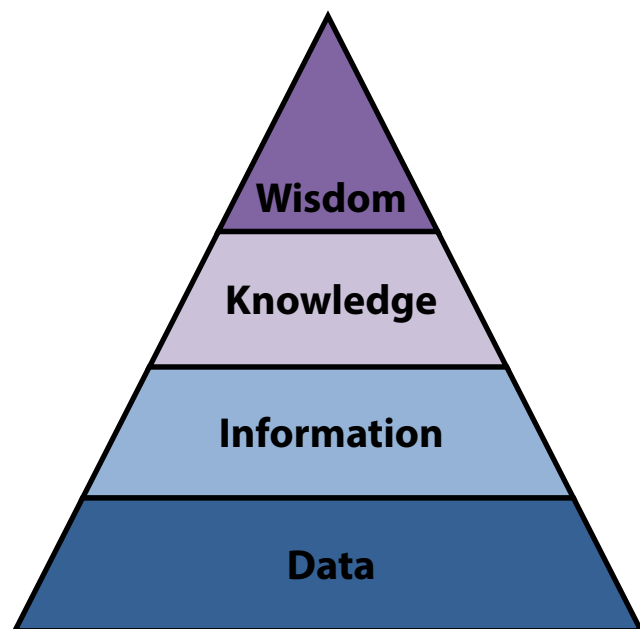
(Figure by author)

Figure 1. DIKW Pyramid

"analyze, manipulate, aggregate, and visualize big data," but one that keenly aligns with the defense sector is object-based storage.[21]

Object-based storage allows for managing, storing, and calling large swaths of unstructured data or semistructured data. It is a form of data curation, which is the "process of creating, organizing, and maintaining data sets so they can be accessed and used by people looking for information."[22] MySQL, one of the most "widely used open-source relational database management systems in the world," was created in 1995 using a codebase created in 1981.[23] Since 1981, the commercial and defense sectors alike have creatively matured and evolved the use of MySQL, among other tools, to leverage efficient and effective database management. However, to unlock the use of those creative, legacy techniques, the data must exist in some form of a structured database. Therefore, the key is to curate unstructured or semistructured data in such a way as to enable a multiplicity of data strategies while simultaneously preventing acquisition vendor lock.

Object-based techniques can allow for data structuring by storing data based on their content and other attributes, using variable lengths and applying unique identification parameters for calling the data.[24]

By creatively applying simple algorithms to separate unstructured or semistructured data into objects with specific attributes and proper identification, data analysts can, with some requisite changes, apply legacy data mining algorithms to extract information swiftly and accurately. Coincidently, object-based management can allocate new object space for unforeseen or never-before-seen observations, and while it may not allow for immediate use, it can guide future use to adjust for any data class imbalances. (Note: Class imbalance is important for machine learning to prevent biased information output.) While object-based storage is not the panacea of unstructured or semistructured data management or inclusion into structured data sets, it does offer an avenue of organization that enables contemporary and evolutionary information generation strategies.

Data begets information leading to both knowledge and wisdom in the DIKW pyramid. Since each dataset will not have all the required data to answer an information call, optimizing storage and management systems enables increases in information returns. Object-based storage is an example of large dataset management, unstructured or semistructured, that would enable rapid data flexibility and information answerability. The next section breaks down how to apply the concept of the DIKW pyramid and object-based storage to both tactical and operational military constructs.

Break It Down—Build It Up

You can have data without information, but you cannot have information without data.

—Daniel Keys Moran[25]

Data management and information calling strategies have differing effects when applied to various levels of decision-making. Tactical and operational decision-making definitions have similar characteristics when compared between the commercial and defense sectors. The difference between the two resides on the tactical level, which is exemplified by immediate decisions needed to provide an in-situation effect against a specific goal. The operational level holds the grander scale decisions to provide long-term goal completion. This analysis focuses on one primary differentiator between operational and tactical levels, which is the timeline with which each follows.

Longer, operational timelines provide advantage in the opportunity for more data assembly and usage, but as a double-edged sword, the word "operational" implies larger force schemes of maneuver and thus, requires continuous, decisive, and contemporaneous action to affect the battlespace. Tactical timelines, while much more granular, offer simpler decisions and therefore, more precise data required to make the decision. If one imagines the DIKW pyramid as the total sum of all parts in or related to the battlespace, then operational actions should strive to account for the greatest chunk of the pyramid. Meanwhile, tactical actions should strive to optimize decision space by accounting for only that information, which relates to the next set of actions. As illustrated in figure 2, the DIKW pyramid could break down into varying shapes that exemplify different types of actions. The figure shows an example operational kill chain of understanding the environment, deciding on preferred commands, and acting within relative control to enable passing command-and-control actions along the seams of wisdom and knowledge. Those actions enter the segment of the pyramid wherein a tactical user enables their understanding of the intent or authority contained therein, decides on the correct effects and timeliness and acts, all the while relaying back to the operational segment both observations and effects.

When the Levee Breaks

Operational and tactical relationships, on a grander scale, necessitate that information and knowledge flow freely and bidirectionally across the inherent divide. Operational actions inherently encompass a series of tactical actions. Since the flow of data, information, knowledge, and wisdom is critical to operational and tactical success, analogize each layer of the DIKW pyramid as a body of water. Each body of water fills or flows at different rates. Consider a constantly streamed intelligence collector as a river of data and the information assertion results as streams filling a section of a knowledge reservoir. The knowledge from each individual collector coalesces to form the overarching situational awareness or knowledge reservoir. Similarly, knowledge produces pockets of battlespace awareness and understanding in the form of waterfalls. These waterfalls, in turn, help fill the situationally dependent wisdom lake that is already partially filled and sourced
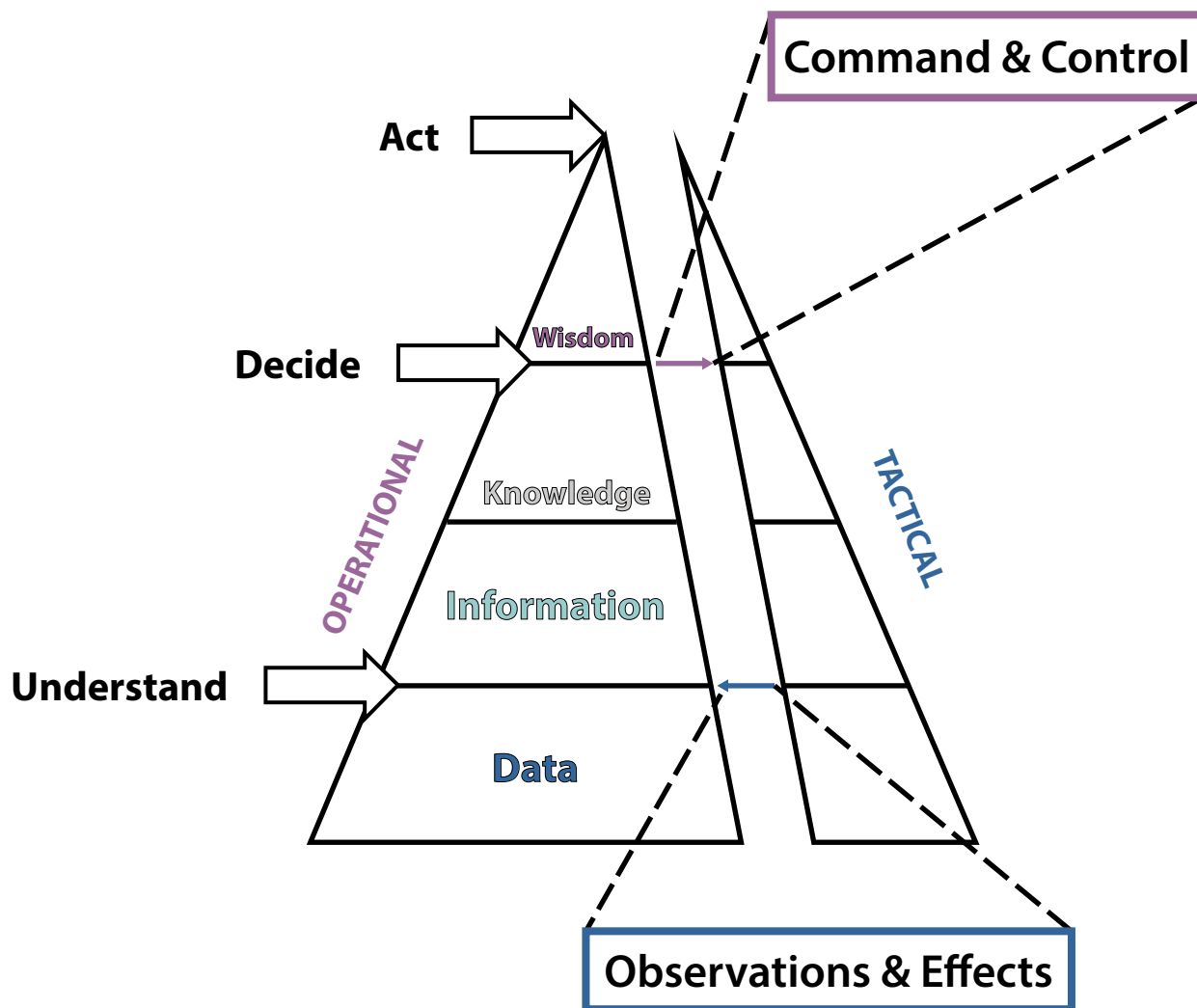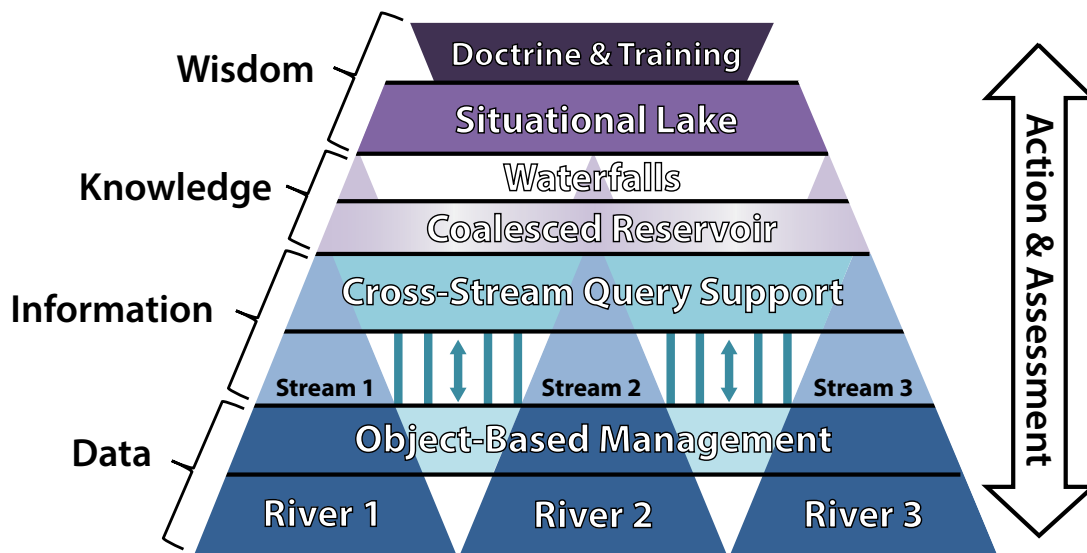
(Figure by author)

Figure 2. Example DIKW Breakdown

from pretraining, doctrine, and recent events. This lake ebbs and flows as knowledge of a situation changes, but as it ebbs, it will reach decisive fill points that necessitate action. Once an action occurs, it inherently reduces the lake's waterline while waiting on the results of said action in the form of assessment. This water analogy appears in figure 3 as a DIKW water table.

The DIKW water pyramid is agnostic of operational or tactical systems. It exemplifies how multiple different collectors from a single tactical system might create tactical understanding, decisions, and actions, or applied operationally, how multiple different tactical systems might feed operational understanding,

decisions, and actions. The critical factor is how objected-based management of data fills the gaps of informational streams by allowing informative queries to transcend any individual data river. Furthermore, this factor highlights that a cohesion of multiple streams and cross-stream information fills a coalesced reservoir of knowledge in which individual pieces of knowledge can enable situational awareness that activates doctrine and training. Where doctrine and training lack, however, it also creates decision space for atypical actions that, when properly informed, might create the optimal solution to the current situation. Removing data cohesion eliminates the ability

Figure 3. DIKW Water Table

for pulling information from multivarietal sources, reducing the knowledge gained in any one situation, and creating ill-informed actions. Thus, data management enables data processing and subsequent information garnering, the cheapest and most cost-effective way to improve kill chain dynamics.

Conclusions

Without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway.
—Geoffrey Moore[26]

The defense sector lies at an inflection point for applying leap ahead technologies to exploit data in all forms. The commercial world has exploded with data applications from personalized advertisements to machine learning language models such as ChatGPT to market research to data storage and calling. The DIKW pyramid provides a simple data growth framework that, when applied correctly, could take future kill chain concepts and make them tenable. The critical key in making webs of sensors that feed webs of shooters is data management. In a world where communication, especially high-bandwidth, low-latency versions, cannot be guaranteed, data management can provide a continuum of successful decisions in a more future-proof, forecastable way. The best way to achieve

data management at an infinite scalability with reliability and resilience in mind is object-based storage and management. Use any search engine, and find solution after solution advocating for object-based storage, from Google to Amazon Web Services to RedHat and more. "Developed in the late 1990s by researchers at Carnegie Mellon University and the University of California–Berkeley, object storage software today can store and manage terabytes (TBs) or petabytes (PBs) of data in a single namespace with the trifecta of scale, speed, and cost-effectiveness."[27]

The DOD should lead the next generation of kill chain dynamics in Joint All-Domain Operations by adopting object-based storage solutions within its intelligence apparatuses. First, it should accomplish an analysis of all its sources of data, specifically looking for where and how the source stores data. Then, it should look for where object-based storage solutions could, when inserted correctly, adapt current data streams into objects. They must accomplish this step both at tactical edge nodes and big data facilities, an underdeveloped operation. Finally, it should experiment with different information calling algorithms to ensure data usability. At completion, the DOD will have created a framework for all portions of the U.S. government to adopt, and it will have laid the groundwork for joint all-domain command and control and

future design methodologies. Again, object-based storage is not the panacea, but it is one example of how the government could take advantage of the commercial sector's efforts to find, extract, and implement the most cost efficient and useful elements. Understanding the way data feeds the overarching machine is critical to the government as it would enable better decisions now, using legacy investments, optimizing data workflows, and ultimately, provide tools and knowledge when and where required. ■

Notes

1. Peter Sondergaard, "Keynote Address" (conference presentation at Gartner IT Symposium/Xpo, Orlando, FL, 2011).

2. Sarah E. Shatby, "The History of Data: From Ancient Times to Modern Day," 365 DataScience, 1 June 2022, https://365data-science.com/trending/history-of-data/.

3. Ibid.

4. Gordon E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics* 38, no 8 (April 1965): 114–17.

5. Ibid.

6. Charles Morris, "Is Tesla's Onboard AI Chip Smarter than USAF's F-35 Fighter Jet?," InsideEVs, 21 May 2021, https://insideevs.com/features/508879/tesla-ai-chip-crazy-powerful/.

7. *Online Etymology Dictionary*, s.v., "data (*n.*)," accessed 29 January 2024, https://www.etymonline.com/word/data.

8. Daniel Barker, "How We Use the Word 'Data' Has Changed—and It's Dangerous," Towards Data Science, 23 February 2018, https://towardsdatascience.com/how-we-use-the-word-data-has-changed-and-it-s-dangerous-b7b6278a8e09.

9. Maddalena Favaretto et al., "What Is Your Definition of Big Data? Researchers' Understanding of the Phenomenon of the Decade," *PLoS ONE* 15, no. 2 (25 February 2020): Article e0228987, https://doi.org/10.1371/journal.pone.0228987.

10. Hui Luan et al., "Challenges and Future Directions of Big Data and Artificial Intelligence in Education," *Frontiers in Psychology* 11 (October 2020): Article 580820, https://doi.org/10.3389/fpsyg.2020.580820.

11. Ibid.

12. Sagar Khillar, "Difference between Big Data and Small Data," DifferenceBetween.net, 14 July 2020, http://www.differencebetween.net/technology/difference-between-big-data-and-small-data/.

13. Martin Frické, "The Knowledge Pyramid: A Critique of the DIKW Hierarchy," *Journal of Information Science* 35, no. 2 (April 2009): 1–13, https://doi.org/10.1177/0165551508094050.

14. Ibid.

15. Russell Ackoff, "From Data to Wisdom," *Journal of Applied Systems Analysis* 16 (1989): 3–9.

16. Laurie A. Schintler and Connie L. McNeely, eds., "Big Data Workforce," in *Encyclopedia of Big Data* (Cham, CH: Springer, 2022), 110, https://doi.org/10.1007/978-3-319-32010-6_208.

17. Ibid.

18. Ibid.

19. Ibid.

20. Ibid.

21. Ibid.

22. Mary K. Pratt, "Data Curation," TechTarget, 20 January 2022, https://www.techtarget.com/searchbusinessanalytics/definition/data-curation.

23. Emmanuelle Rieuf, "History of MySQL," Data Science Central: A Community for Big Data Practitioners, 16 December 2016, https://www.datasciencecentral.com/history-of-mysql/.

24. Schintler and McNeely, "Big Data Workforce," 110.

25. Corinne Lenherr, "'Patch Everything, All the Time' Is Out – Today Is Threat Intelligence," *Security Awareness* (blog), InfoGuard, 11 April 2019, https://www.infoguard.ch/en/blog/patch-everything-all-the-time-is-out-today-is-threat-intelligence.

26. Geoffrey Moore (@geoffreyamoore), "Thoughts from the week #1: Without big data analytics, companies are blind and deaf, wandering out onto the Web like deer on a freeway," Twitter, 12 August 2012, 10:29 p.m., https://twitter.com/geoffreyamoore/status/234839087566163968?lang=en.

27. Sudipto Paul, "What Is Object Storage? It's Crucial for Managing Cloud Data," G2, 5 November 2021, https://www.g2.com/articles/object-storage.