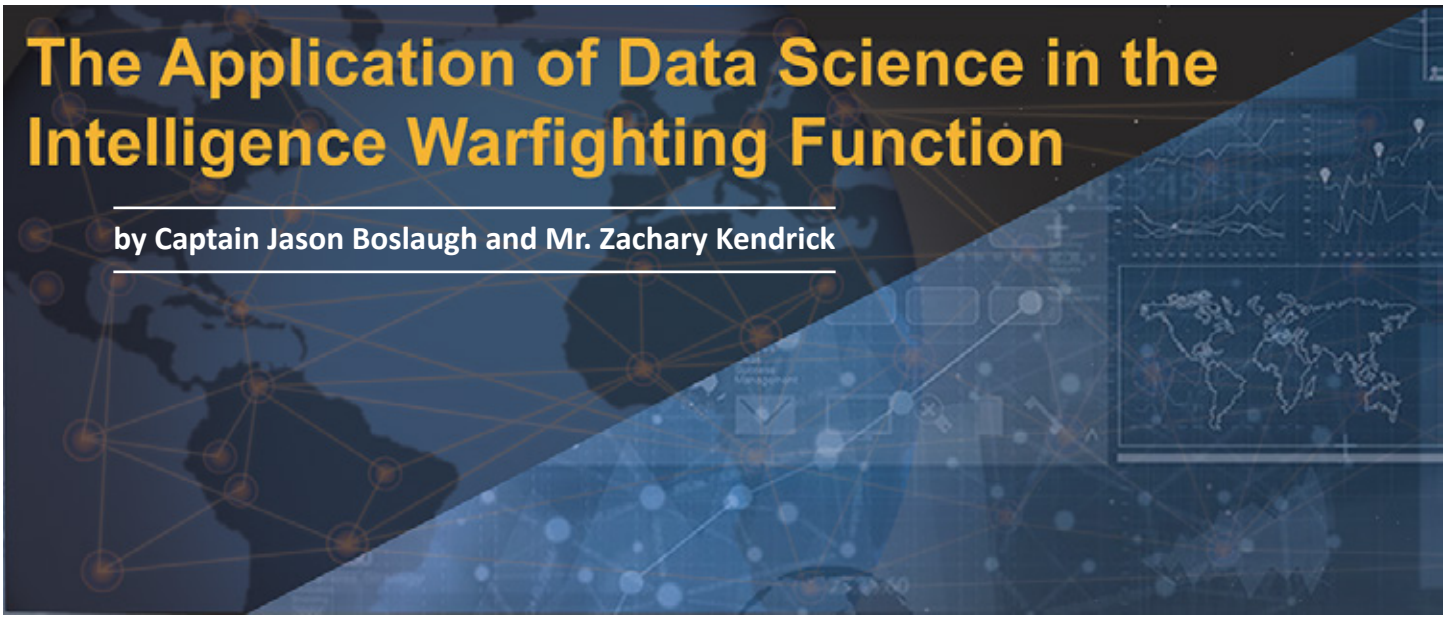


The Application of Data Science in the Intelligence Warfighting Function

by Captain Jason Boslaugh and Mr. Zachary Kendrick

Editor's Note: This article is the second in a two-part series on data science. The first part, by CPT Iain Cruickshank, titled "On Data Science and Intelligence Analysis," was in the July–September 2019 issue of* Military Intelligence Professional Bulletin*. CPT Cruickshank's article provided a basic foundational understanding of data science and its application in the intelligence community. Part two of the series, presented here, discusses how the U.S. Army can apply data science lessons learned from academia and industry to modernize the intelligence warfighting function.

Introduction

Data science—or the continual extraction of knowledge from data using advanced mathematics and coding—seeks to provide meaningful and actionable insights into a problem set based on the analysis of large volumes of complex data. Industry has used data science for years to provide relevant and useful business insights to increase profit margins, to attract and personalize offerings to customers, and to identify and reduce internal and external inefficiencies in organizational processes. In 2013, McKinsey & Company, an American worldwide management consulting firm, estimated that big data[1] initiatives could account for "$300 billion to $450 billion in reduced health-care spending, or 12 to 17 percent of the $2.6 trillion baseline in U.S. healthcare costs."[2] Cost savings occur as data scientists deliver insights that were previously indecipherable based on the large volumes of data available to organizations. Insights come in the form of risk mitigation, relevant products to support decision making, and streamlined organizational processes. When corporations use data science, it can result in increased profitability and efficiency, whereas for U.S. Army intelligence, it can provide the commander with an information advantage to enhance the lethality and survivability of U.S. forces in multi-domain operations.

This article describes the benefits that data science provides to the intelligence warfighting function and commanders, and it recommends near-term actions that must occur to successfully integrate data science practices into Army intelligence.

Data Science and the World's Largest Employer—Walmart

The U.S. Army is a massive organization, inundated with data, but the issue with implementing data science into the Army is not one of scale. Walmart Inc., the largest employer in the world by employee count and revenue,[3] has integrated data science principles into the majority of its processes. With technology-driven competitors like Amazon rapidly increasing their market share and the sheer scale of Walmart's operations, Walmart required integration of big data and data science practices to develop the solutions necessary to remain profitable.

An early example of Walmart's use of data science occurred in 2012, when Hurricane Sandy struck the most densely populated area of the United States. Everyone knew that people in the hurricane's path would need flashlights, emergency equipment, and other extra supplies, but Walmart's chief information officer successfully used statistical models to show what specific staple food items would be required, as well as quantities, to meet customers' immediate needs during an emergency. Because of these models, Walmart was able to divert logistical chains containing these types of items to the hurricane's impact zone.

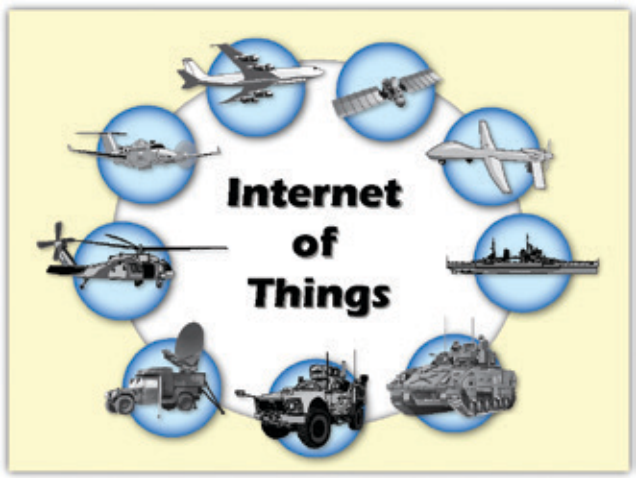
A more recent application of Walmart's use of data science is its Data Café, a state-of-the-art analytics hub that automatically detects sales anomalies in its inventory. When

items are not selling, analysts can contact responsible store merchandising teams that can provide immediate feedback leading to rapid data-driven solutions. This could result in re-positioning a product to high-traffic areas, reducing the price of the product, or discontinuing sales of the product. Walmart's Data Café system has led to a reduction in the time it takes from spotting a problem to proposing a solution—from an average of 2 to 3 weeks to approximately 20 minutes.[4]

Walmart is also currently using machine learning[5] and artificial intelligence[6] to determine customer satisfaction for in-store self-service kiosks (facial recognition) and online sales. Walmart was an early adopter of radio-frequency identification (smart labels) and wireless networking technologies to monitor logistics. Current and future capabilities include "integrating [Internet of things] IoT tags to products in order to monitor product usage, auto-replace products as necessary, and monitor expiration dates or product recalls."[7] Expedient analysis of real-time data is key to driving business performance in an industry context. This application in industry parallels a military application—meeting the commander's immediate information requirements to support mission command.

So What is the Internet of Things?

"This is the concept of basically connecting any device with an on and off switch to the Internet (and/or to each other). This includes everything from cellphones, coffee makers, washing machines, headphones, lamps, wearable devices and almost anything else you can think of. This also applies to components of machines, for example a jet engine of an airplane or the drill of an oil rig…If it has an on and off switch then chances are it can be a part of the IoT. The analyst firm Gartner says that by 2020 there will be over 26 billion connected devices…That's a lot of connections (some even estimate this number to be much higher, over 100 billion). The IoT is a giant network of connected "things" (which also includes people). The relationship will be between people-people, people-things, and things-things."[8]



Collaborating with Government, Industry, and Academia

Based on use cases and lessons learned from academic and industry partners, Army leaders have recognized a direct applicability of data science to processes within the Army warfighting functions. In 2018, senior Army intelligence leaders engaged in an effort to modernize the military intelligence (MI) workforce and optimize Army intelligence for future conflicts, particularly large-scale ground combat operations. To start the Army intelligence modernization process, the Deputy Chief of Staff G-2, the U.S. Army Intelligence and Security Command (INSCOM) Commanding General, and the U.S. Army Intelligence Center of Excellence (USAICoE) Commanding General directed the development and integration of data science capabilities into the intelligence warfighting function. Over the past 18 months, the Army Futures Command's Capability Development Integration Directorate - Intelligence at USAICoE has deliberately captured best practices, tactics, techniques, and procedures from government, industry, and academia to describe the ways, means, and ends to integrate data science into Army intelligence.

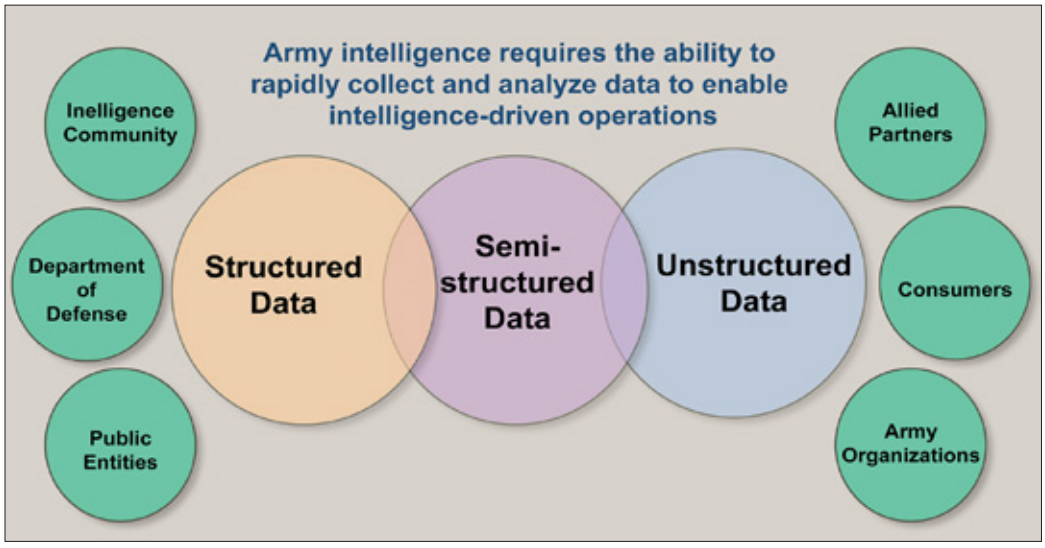
The assessment has been shaped by and nested with USAICoE's involvement in various Army analytic modernization efforts. This has included the USAICoE-sponsored data science study with the U.S. Army Research Laboratory and industry partner CUBRC, Inc., "to determine how to employ future data science and data scientists to maximize data exploitation and reduce the burden on Army Intelligence Analysts."[9] The study included a 2-day workshop at George Mason University. "The focus of the workshop was to gather experts from government, industry, and academia to discuss best practices and perspectives on utilizing data science within the [intelligence warfighting function] IWfF. In addition to government representatives from U.S. Army Intelligence and Security Command, U.S. Army Training and Doctrine Command, U.S. Army Forces Command, and U.S. Army Special Operations Command, the workshop captured the experience of scientists from Microsoft, BAE Systems, RAND, and Lockheed Martin."[10]

How Data Science Can Help

Army intelligence is at a tipping point for the next evolution in data-driven intelligence analysis. The information deluge is outpacing the analyst's ability to derive comprehensive insights from all the available sources. To put the vast amount of data into perspective, from 1992 to 2018 the Army's use of Multimedia Message Manager, a secure messaging capability, has increased an average of 1000 percent, from approximately 30,000 messages per year to the current flow of 35,000,000 per year.[11]

Within the Army MI context, data science applies to all intelligence disciplines, and a broad data solution must be resourced with the appropriate technology that is capable of scaling to meet continued exponential growth of data in a multilevel security domain environment. Data science capabilities have the potential to rapidly analyze high-volume data sets and automatically correlate entities to create comprehensive threat analytics products. This will enable the analyst to spend more time on quality control and critical analysis and less time on data search and discovery. The addition of these capabilities to an intelligence analyst's kit bag can sharply enhance the ability of intelligence sections to manage information and support the commander at the pace of war.

In military operations, the sheer amount of data feeds available are exponentially greater and potentially more complex in multi-domain operations than in previous eras.[12] Army intelligence requires the ability to rapidly collect and analyze structured, semi-structured, and unstructured data. This data is created by the intelligence community, Department of Defense (DoD), public entities, allied partners, consumers, and the Army's own organizations to enable intelligence-driven operations, to "enable commander's situational understanding in the future complex environment using adaptive and innovative means."[13]



Simultaneously, Army intelligence must process and exploit the data collected by an increasingly large and varied array of sensors. Each sensor captures orders of magnitude more data at multiple levels of analysis and classification than earlier generation sensors. All of this data is collected to support a diverse range of commanders, operational missions, and automated mission command systems. Data science has the potential to be a force multiplier for the intelligence warfighting function; however, to integrate data science effectively into Army intelligence operations, the Army must address multiple priority findings from the application of data science within the intelligence warfighting function.

Invest in, Develop, and Retain Human Capital

As of 2018, more than 10,000 data science positions have been vacant throughout the industry sector. With the demand for data scientists, cloud engineers, and developers reaching an acute level, IBM and Google have separately launched new certification programs to fill their human capital gap by training from within their organizations. IBM has developed a data scientist certification to assess and validate data science skills that professional job candidates obtain through practical experience. Google's efforts are focused on four areas: cloud developer, cloud engineer, cloud security engineer, and a G(oogle) suite certification. Google also offers on-demand courses and short hands-on labs to develop its workforce. These are just two of many industry examples upon which Army MI can model its approach. These examples also indicate that industry is committed to developing the necessary workforce to meet long-term increasing data demands, showing the importance of this capability for organizations with large volumes of data.[14]

In order to build an effective data science capability in Army intelligence, the foremost priority is to develop and retain data scientists who can execute the mission. "Improving the resilience of leaders and Soldiers—the Army's most valuable capability—requires training, educating, equipping, and supporting them to execute [multi-domain operations] MDO in all of its intensity, rigor, and complexity."[15] Army MI must invest in and align its workforce to incorporate data science in the intelligence warfighting function to integrate enhanced tradecraft and intelligence analysis capabilities. A successful data science capability for Army MI requires human capital with the necessary training and education to conduct this mission.

As a starting point to scale this capability within Army intelligence, an identification mechanism and/or competency assessment can identify existing highly talented Soldiers with data science-related competencies in the current MI workforce. This assessment would include a scrub of Soldiers' records and the identification of skillsets with

relevance to data science, such as a civilian education in statistics or other data science-related fields. The Army's Human Resources Command or the Headquarters, Department of the Army's G-1 can conduct the assessment, and the Army can use it to assign, train, and leverage top talent to data science missions. The intent is to develop an intelligence data science capability and a more responsive MI workforce to meet commanders' needs. The Army must solve this problem set by leveraging data science to provide the "heavy lift" to intelligence analyst capability, enhancing predictive analysis, and providing a more complete intelligence picture that creates actionable intelligence to drive the commander's decision-making process. Today's operational environment presents a much more dangerous threat to the Army than that faced in Iraq and Afghanistan, and it requires rapid decision making to counter our adversaries in large-scale ground combat operations.

A data science capability will not replace human analysts. Single source and all-source intelligence analysts are essential to information discovery, processing, exploitation, and dissemination for all MI operations with direct access to combat information in near real time. This analytical workforce, trained on data science principles and capabilities, will remain the first to start the process of managing collected data and ensuring it is distributed appropriately across the intelligence community and DoD enterprise. The resident expertise in these occupational specialties will be instrumental in providing initial opportunities, identification of shortfalls, and feedback for data science teams. Army intelligence analysts must adapt to the challenge of identifying opportunities by managing and analyzing large data sets supporting mission command to increase lethality and survivability through rapid identification of critical and relevant combat information.

Trained Army MI data scientists and supporting intelligence elements can enable MI operations to better support Army, joint, and multinational operations by leveraging data science principles and tradecraft to information collection operations and the overall intelligence process. Supporting elements must have opportunities to obtain additional data science training to enable cross-functional team collaboration. Training additional supporting personnel in the MI workforce will enable quality results down to the lowest level and will provide the flexibility to meet mission command requirements at multiple echelons and supported organizations. The ability to acquire, educate, and train Army personnel will drive the timeline to achieve the desired capability end state. In order to identify candidates to become Army intelligence data scientists, the Army can conduct a service-wide assessment to identify Soldiers who already possess formal education and training required of a data scientist. These skillsets include but are not limited to computer software programming, computational social sciences, and statistics.

Last, the Army must also be prepared to identify retention and incentive mechanisms for Soldiers who become data scientists, such as additional duty service obligations or incentive programs. Many industry and government organizations are experiencing problems with hiring talented data scientists for a variety of reasons. The biggest hurdles to overcome are lack of qualified and skilled personnel, security clearance eligibility, and competitive salary constraints. While professional development incentives exist, monetary incentives will likely require policy changes. The Army Medical Department's Selected Reserve Incentive Program (SRIP) shows the potential for a data science recruitment/retention model because medical and data science skillsets are parallel, requiring formal education and competitive compensation. It is important to identify and adopt best practices from SRIP to minimize the compensation and benefit gaps between government and industry.

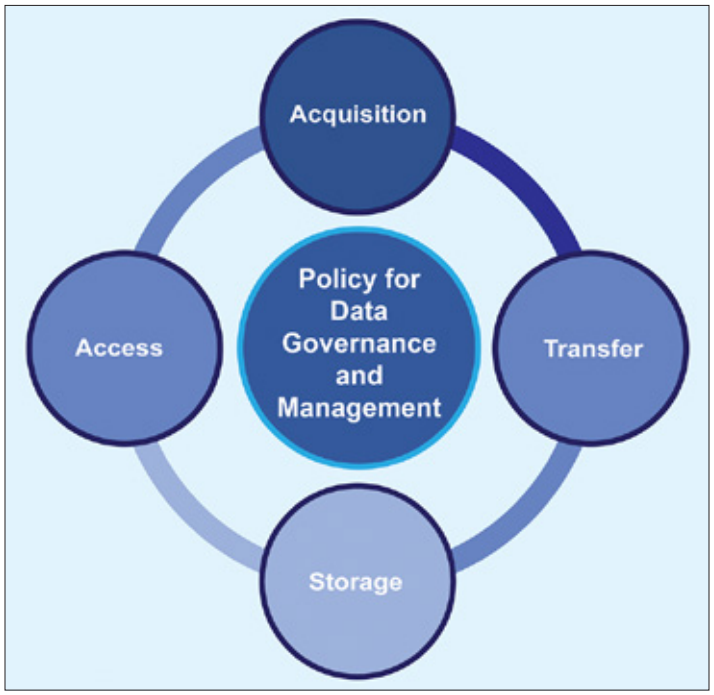
Modify Existing Policy for Data Governance and Management

Army MI must also establish policies and procedures for data governance and monitoring, to include functions involving acquisition, transfer, storage, and access to enable data science operations. Assigning an office of primacy for Army data science will be necessary to ensure data science professionals have access to the tools to operate seamlessly across multiple networks and classification levels. Army MI must also be involved in developing and managing policies that determine who should have access to data and specify/interpret the functions and operations of data science personnel. The specific policies to address in the near term are—

- Rapid accreditation of data science tools.
- Adjustment of information assurance policies to promote the rapid fielding of data science tools on government networks.
- Development of policy to enable an instantaneous or near-instantaneous multilevel security domain data transfer mechanism.
- Rapid, streamlined onboarding of Army personnel to data that is non-DoD owned and managed.

We must address these policy issues; otherwise a comprehensive data science capability is not feasible and maintaining relevance of the capability will become a

struggle because data science will not be able to keep pace with the commander's decision-making process in large-scale ground combat operations.



Adopt a Data-Centric Culture at all Army Echelons

At the Army leadership level, an overarching need exists for a top-down culture change in order to adopt data science processes to support military operations. Identifying, consolidating, and structuring data alone is just a small part of addressing MI data problems and shortfalls. Data science teams must provide immediate value through relevant, easily digestible, algorithmically generated intelligence products that a commander trusts and is comfortable integrating into his or her decision-making process. These products, derived from a multitude of intelligence community-approved sources, can increase the effectiveness of and value provided by intelligence sections with relative inexperience, particularly at echelons corps and below. Once intelligence personnel have corroborated the available data and analyzed it, the data can then provide the mission command intelligence that assists commanders in making operational decisions to mitigate risk and enable an advantage over an adversary.

Data science techniques applied to previously collected data can provide quantitatively grounded insights into the most likely threat courses of action based on previous activity. This serves as a starting point for determining future threat courses of action, improving on analyst-driven, qualitatively derived insights that lack the full complement of data sources available. To provide value, these data science-driven products and tools must be injected into the commander and staff's decision-making processes to inform decisions based on both the operational instinct and the plethora of data sources available. In the interim, Army leaders must be educated on the benefits that an intelligence data science team can bring to intelligence analysis. Improvements to the vision and strategy for the data science mission would result from providing Army leaders with a better understanding of this powerful capability, including its limitations.

Tailor and Scale the Capability to the Warfighter's Needs

An enterprise approach to integrate data science effectively into Army intelligence at each Army echelon requires a unique data science capability based on the envisioned threat to that echelon, decision-making timelines, and the data science expertise of the Soldiers organic to that echelon. Challenges will exist as the echelon approaches the tactical edge, for example, challenges resulting from intermittent network connectivity, truncated decision-making timelines, and limited familiarity with the capabilities of data science. Soldiers at the tactical echelon will likely function as customers of products or applications created at higher echelons, such as INSCOM's organizations or corps-level intelligence elements. The data science tools provided to the tactical organizations will automate the standardized and tedious steps of the intelligence process. Automation tools can support enemy course of action development as part of intelligence preparation of the battlefield, assist with terrain/mobility analysis, and provide situation development, which allows insight into enemy decision making. These tools will reduce the "heavy lift" of the analysts who are providing direct support to tactical commanders and maneuver elements, whereas higher echelons will have the time and space to develop unique problem set-specific tool instances to monitor complex problems.

Integrating data science within the intelligence warfighting function has the potential to produce capabilities that support information dominance by modernizing the training, organizations, analytical toolsets, and architecture related to the most critical and underlying element of intelligence analysis—"data." By using industry-proven processes, teams of highly skilled data experts will work with the current intelligence workforce to build enhanced tradecraft, big data analytics, and data science capabilities into Army formations. Incorporating data science into the intelligence warfighting function's core missions will have a considerable return on investment when used to successfully address intelligence requirements and inform decisions

across all time horizons and echelons. Data science teams will not replace the need for intelligence analysts and their specialized critical-thinking skills; instead, these teams will provide new analytic capabilities to support intelligence missions across the force.

Army MI capabilities can be improved through an iterative lessons learned process that documents potential analytic opportunities. This will allow the intelligence staff to have the resources to assess where current analytic capabilities exist and where legacy intelligence workflows can be improved. Army units can identify optimal organizations that can integrate and support data science capabilities with the appropriate resources, tools, architecture, and leadership understanding of the intelligence problem to ensure success of the MI data science program. Organizations will mentor highly skilled Soldiers and leaders to provide commanders with accurate intelligence and assessments of the information environment, enhancing situational understanding and relevant, timely decision making.

To provide the success needed for integrating data science into intelligence units or sections, the Army must address changes across all echelons through recommended intelligence data science doctrine, organization, training, materiel, leadership and education, personnel, facilities, and policy (DOTMLPF–P) solutions. It is necessary to develop a concept of employment guided by FM 3-0, *Operations*; FM 2-0, *Intelligence*; the Army Operating Concept; approved Army strategies; lessons learned; and the intelligence data science DOTMLPF–P. The concept of employment will describe the functions and roles that data science will play for the intelligence warfighting function across the range of military operations, with a focus on large-scale ground combat operations.

Organizations that are part of the data science community of interest, and responsible for leading the integration of data science into Army intelligence, will identify senior leader decision points, establish timelines, and, if required, initiate requisite Joint Capabilities Integration and Development System documentation to implement changes (e.g., initiation of a DOTMLPF–P integration change recommendation or execute order). The Army data science community must identify the breadth of resources necessary to execute the tasks, missions, and functions of an integrated data science capability to support large-scale ground combat operations in environments that are complex, uncertain, and rapidly changing.

Conclusion

By institutionalizing data science within the Army's workforce, culture, and policy to address the warfighter's requirements, the Army can incorporate robust new information capabilities across the DOTMLPF–P. The result will enable the Army to dominate the information environment and effectively enhance the lethality and survivability of U.S. and coalition forces in multi-domain operations.

Endnotes

1. Big data is "extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviors and interactions." *Oxford English Dictionary,* s.v. "big data (n.)," accessed 20 May 2019, https://en.oxforddictionaries.com/definition/big_data.

2. Basel Kayyali, David Knott, and Steve Van Kuiken, "The big-data revolution in US health care: Accelerating value and innovation," McKinsey & Company, April 2013, 5, https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care.

3. "Global 500," Fortune, 2019, http://fortune.com/global500/list/.

4. Bernard Marr, "The Most Practical Big Data Use Cases of 2016," Forbes, 25 August 2016, https://www.forbes.com/sites/bernardmarr/2016/08/25/the-most-practical-big-data-use-cases-of-2016/#50be9a6a3162.

5. Machine learning is "the application of artificial intelligence (AI) that provides systems the abilities to automatically learn and improve from experience without being explicitly programmed." "What is Machine Learning? A definition," Expert System, accessed 20 May 2019, https://www.expertsystem.com/machine-learning-definition.

6. Artificial intelligence is "the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages." *Oxford English Dictionary,* s.v. "artificial intelligence (n.)," accessed 20 May 2019, https://en.oxforddictionaries.com/definition/artificial_intelligence.

7. Bernard Marr, "How Walmart Is Using Machine Learning AI, IoT and Big Data To Boost Retail Performance," Forbes, 29 August 2017, https://www.forbes.com/sites/bernardmarr/2017/08/29/how-walmart-is-using-machine-learning-ai-iot-and-big-data-to-boost-retail-performance/#224128996cb1.

8. Jacob Morgan, "A Simple Explanation Of 'The Internet Of Things,' " Forbes, May 13, 2014, https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#300d0b141d09.

9. Robert Ganger, John Coles, Jacob Ekstrum, Timothy Hanratty, Eric Heilman, Jason Boslaugh, and Zachary Kendrick, *Army Intelligence Data Science Workshop Summary Report* (Aberdeen Proving Ground, MD: U.S. Army Research Laboratory, November 2018), ii.

10. Ibid.

11. U.S. Army, Deputy Chief of Staff for Intelligence, G-2, *The Army Intelligence Enterprise Digital Concept of Operations,* 31 October 2018, 6.

12. Garrett Hopp, Glenn Gleason, Nick Rife, and Ashley Muller "Data Analytics to Win in a Complex World," *Military Intelligence Professional Bulletin* 44, no. 4 (October-December 2018): 59-61.

13. Department of the Army, "Terms of Reference Between the Army's Senior Intelligence Officers," n.d., 2.

14. Matt Hamblen, "IBM and Google Create New Certifications for Data Scientist and Cloud Skills," SDxCentral, January 29, 2019, https://www.sdxcentral.com/articles/news/ibm-and-google-create-new-certifications-for-data-scientist-and-cloud-skills/2019/01/.

15. Department of the Army, Training and Doctrine Command (TRADOC) Pamphlet 525-3-1, *The U.S. Army in Multi-Domain Operations 2028* (Fort Eustis, VA: TRADOC, 6 December 2018), 20.

CPT Jason Boslaugh is currently attending the College of Strategic Intelligence at the National Intelligence University, Bethesda, MD. He previously served as the branch chief for single-source intelligence capability development in the Requirements Determination, Capability Development Integration Directorate-Intelligence, Future Concepts Center, U.S. Army Futures Command at Fort Huachuca, AZ. In his position as branch chief, he worked with a team that conducted doctrine, organization, training, materiel, leadership and education, personnel, facilities, and policy (DOTMLPF-P) assessments. The team also determined future Army requirements for signals intelligence, geospatial intelligence, space, counter-unmanned aircraft system, human intelligence, counterintelligence, and data science to help prepare the Army intelligence warfighting function for the envisioned threat. Before this assignment, CPT Boslaugh served at 7th Special Forces Group (Airborne) as a military intelligence detachment commander and battalion S-2.

Mr. Zachary Kendrick is a capability developer in the Requirements Determination, Capability Development Integration Directorate-Intelligence, Future Concepts Center, U.S. Army Futures Command at Fort Huachuca, AZ, and currently works as a member of the Data Science and Processing, Exploitation, and Dissemination teams. He formerly served in the 82nd Airborne Division and Joint Special Operations Command deploying to Operation Iraqi Freedom three times as a signals intelligence noncommissioned officer. Mr. Kendrick has a bachelor of science in business administration.