Joint Special Operations University
College of Special Operations

Exploitation of Big Data for Special Operations Forces

by Tammy Low

JSOU Press
Occasional Paper
November 2018

On the cover: Airmen wait for approaching aircraft during close air support training as part of Emerald Warrior 17 at Eglin Air Force Base, Florida, 2 March 2017. Emerald Warrior is a U.S. Special Operations Command exercise during which Joint Special Operations Forces train to respond to threats across the spectrum of conflict. AIR FORCE PHOTO BY TECH. SGT. BARRY LOO/DVIDS

JSOU Press Publications are available for download at:
http://jsou.libguides.com/jsoupublications.

ISSN 2572-9020 (print)
ISSN 2572-9039 (online)

About the Author

U.S. Air Force Lieutenant Colonel Tammy Low is currently the deputy science and technology advisor to the United States Indo-Pacific Command. She conducted her research for this report as a graduate student assigned to the Air War College, Air University, Maxwell AFB, in Alabama.

Prior to that, Lieutenant Colonel Low served as commander of the Air Force Drug Testing Laboratory, deputy director of the Asian Office of Aerospace Research and Development, assistant professor of chemistry at the United States Air Force Academy, branch chief at the Air Force Research Laboratory, fuels and energy branch, as well as a variety of other leadership and technical assignments. She earned her Ph.D. in organic chemistry from the University of Florida.

Exploitation of Big Data for Special Operations Forces

The exploitation of big data, analytics, and machine learning in the civilian sector grows exponentially each year. In comparison to the civilian sector, the U.S. military is not capitalizing on big data to its fullest potential. Big data analytics have broad applicability across the Department of Defense (DOD), however, this paper will focus on big data's use within the United States Special Operations Command (USSOCOM). The Joint Special Operations University *Special Operations Research Topics 2018* highlights the interest USSOCOM has in big data.[1] As USSOCOM is a unique combatant command that encompasses a variety of core activities across the range of military operations, its missions often require the use of special techniques, skills, and capabilities to operate in dynamic, austere environments and situations not ideal for conventional forces.[2] Core activities include unconventional warfare (UW), counterinsurgency (COIN), military information support operations, and civil affairs operations as examples where big data analytics and mining of public information bolster the understanding of population-centric issues. Because the core activities of USSOCOM vary widely in nature, identifying big data requirements and solutions across the Special Operations Forces (SOF) enterprise remains challenging. These challenges include types of data to collect, analysis of data, data management, integrating information with other intelligence, changing culture and governance, establishing policies for big data use, development of a data-capable force across multiple domains, and recruitment of talented data scientists and analysts. The priority should be to establish the strategy, and then identify the big data and analysis requirements. Developing a big data strategy and policy for USSOCOM is beyond the scope of this paper. Rather, this paper will focus on big data exploitation—specifically unclassified, open source data for SOF pre-conflict activities.

In certain operations like COIN, where SOF function in a permissive or semi-permissive environment with the host nation and are embedded in the local population, SOF can verify and validate potentially actionable information produced from big data. On the other hand, in denied environments of UW, the analysis of big data significantly assists SOF by enhancing situational awareness, bolstering indications and warnings, providing real-time feedback about a potential uprising area, and helping with pre-crisis decision-making, which is where SOF are most effective. According to the U.S. Army Special Operations Command (USASOC) unclassified report on Russia's modern UW titled *"Little Green Men": A Primer on Modern Russian Unconventional Warfare, Ukraine 2013–2014*, the Russians employed a variety of kinetic, non-kinetic, asymmetric, nonlinear actions and expanded the typical instruments of power to include informational, economic,

1. Joint Special Operations University, *Special Operations Research Topics 2018* (Tampa, FL: JSOU Press, June 2017), 4, 9.
2. *Special Operations*, Joint Publication (JP) 3-05 (Washington, D.C.: The Joint Staff, July 2014), ix-II–18.

financial, and cultural activities.[3] The authors elaborate further explaining how informational warfare across the domain (e.g., internet, newspapers, and blogs) encompasses manipulation and deceit.[4] Observations from open source information produce real-time evidence and reasoning on activities, events, and emerging issues, especially in nontraditional armed conflict requiring better agility to adapt to ambiguous, complex, and disruptive security environments.

The intelligence gaps during the 2012 Arab Spring uprising illustrate another example of the importance of exploiting big data. David Shedd, former deputy director of the Defense Intelligence Agency, acknowledged this intelligence failure and the need to discuss "how to take advantage of the enormous amount of open-source information that is out there, and draw inferences of where a trend may be."[5] General Raymond Thomas III, 11th Commander of USSOCOM, also noted during his address to the Senate Armed Services Committee in May 2017, that SOF are most effective before a conflict.[6] Big data contributes to the information environment and all-source intelligence, and provides greater understanding of socio-culture, educational, economic, and political issues prior to a conflict.

To better understand exploitation of big data for SOF needs, this paper begins with a definition of big data, followed by the evaluation of SOF big data requirements, opportunities, and challenges focused on pre-conflict situations and environments. Finally, the paper analyzes how the United Nations (UN) Global Pulse program, which uses data science and analytics for humanitarian actions and sustainable development, yields insights and lessons learned on the exploitation of big data for SOF. As with any new technology focus area used by the military, operational, technical, ethical, and legal considerations and limitations exist. Although critically important, this paper will not address ethical and legal issues due to the limited scope of this research.

Thesis

This research paper advocates for the exploitation of big data for SOF pre-conflict activities despite some operational and technical challenges and limitations employing big data. SOF are currently using big data in support of their missions, but not to the full extent of technological capabilities available for predictive analysis, decision-making, and expeditious integration with other sources. The UN Global Pulse program offers different perspectives and lessons learned from big data analytics since their establishment in 2012. To support the recommendation of exploitation of big

3. United States Army Special Operations Command, *"Little Green Men": A Primer on Modern Russian Unconventional Warfare, Ukraine 2013–2014* (Fort Bragg, N.C.: U.S. Army Special Operations Command, 2016), 5–6. For the purpose of clarity and uniformity within this study, the authors use the term *unconventional warfare* to embrace the wide variety of military, informational, political, diplomatic, economic, financial, cultural, and religious activities Russia employed in Ukraine; *Special Operations*, 2.

4. United States Army Special Operations Command, "*Little Green Men*," 15.

5. Cheryl Pellerin, "DIA Five-Year Plan Updates Strategic Warning Mission," *DoD News*, 18 July 2012, http://archive.defense.gov/news/newsarticle.aspx?id=117160.

6. General Raymond A. Thomas, III, *Statement Before the Armed Services Committee*, 4 May 2017, 10, https://www.armed-services.senate.gov/imo/media/doc/Thomas_05-04-17.pdf.

data more aggressively, this paper will analyze big data needs and challenges for SOF, and discuss lessons learned and opportunities from the UN Global Pulse big data initiatives as case studies. The exploitation of big data from open sources (e.g., news media, financial data, social media, and radio broadcasts utilizing data mining, machine learning, and readily available public big data programs) assist SOF operations by providing greater situational awareness and acts as an early warning detection system through increased understanding of sociocultural, population sentiments, political, and economic issues and events. Considering limited SOF resources, funding, and personnel, big data delivers insights on targeted areas that SOF should or should not focus. Moreover, it contributes by sorting through the vast amount of data, translating it into structured, usable formats, and obtains value-added information for processing, exploitation, and dissemination systems and processes. However, big data from public, open sources alone will not be a stand-alone solution to erase the fog and friction inherent in conflicts and must be validated prior to integration with other sources to obtain actionable intelligence.

Technological innovations like machine learning help overcome big data challenges. PHOTO BY NICOELNINO/SHUTTERSTOCK

Defining Big Data

A quick search on the internet results in multiple definitions of big data. According to the National Institute of Standards and Technology, big data are "extensive datasets … that require a scalable architecture for efficient storage, manipulation, and analysis."[7] Big data challenges are often characterized by various types of "Vs," such as volume, variety, velocity, and veracity. Big data, as the name applies, involves a mass volume of information. We live in a densely data-driven, technological world, where we produce approximately 2.3 quintillion bytes of data each day (2.3 trillion gigabytes),[8] and the volume continues to grow as more people move into the digital age, including those from less developed regions of the world. Besides volume, the variety of data comes in different

7. National Institute of Standards and Technology (NIST), *NIST Big Data Interoperability Framework: Volume 1, Definitions (NIST Special Publication 1500-1, September 2015)*, 5, http://dx.doi.org/10.6028/NIST.SP.1500-1.

8. IBM Analytics, "Big Data Analytics" video, accessed 9 November 2017, https://www.ibm.com/analytics/us/en/big-data/.

forms like email messages, videos, photos, tweets, online transactions, and audio recordings. About 95 percent of the data collected are unstructured (e.g., photographs, videos, radio broadcasts, web pages, and social media data), meaning they exist in a form not easily manageable.[9] Analysis of unstructured data for extraction of knowledge is very challenging and requires specialized data analytics tools. Data mining—a form of machine learning—extracts, processes, and analyzes massive volumes of data into useful user information.[10] The velocity of big data describes the speed of the data transmission, as well as the speed of data processing and analysis. With improved computer data exchange technology and increased network connectivity, data transfers at incredible speed, but processing and analyzing the data for use remains challenging. Automated versus manual analysis is necessary because of the vast scale of data. The fourth big data characteristic is veracity, the reliability of the data. Employing machine learning improves the reliability of data by enabling computers to learn from big data, recognize bad data, and bolster the processing and analysis of data into useful information.

Collected and analyzed data reveals patterns and trends, and increases understanding of social dynamics potentially impacting events and activities.

Despite the many definitions and often overuse of big data terminology, the exploitation of big data for useful or actionable intelligence is the critical factor. The *AF Future Operating Concept, A View of the Air Force in 2035* describes how big data "helps make collection from all sources discoverable, and improves human ability to assess, explain, and anticipate adversary actions while providing improved mechanisms for intelligence delivery."[11] Collected and analyzed data reveals patterns and trends, and increases understanding of social dynamics potentially impacting events and activities. Viktor Mayer-Schönberger and Kenneth Cukier, authors of *Big Data: A Revolution That Will Transform How We Live, Work, And Think*, explain that "big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more."[12] Businesses rely on big data on a regular basis to forecast retail sales and data-mine online shopping trends leading to nearly instantaneous pop-up ads on personal computers or Facebook pages. Besides commercial retail, other sectors taking advantage of the data-driven world include finance, healthcare, agriculture, travel, and telecommunications. Just as the commercial industry benefits from big data, SOF can exploit data from various open sources in support of pre-conflict activities. The following section highlights the big data needs and challenges of SOF.

9. Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, And Think* (New York: Houghton Mifflin Harcourt, 2013), 47.

10. Ethem Alpaydin, *Machine Learning: The New AI* (Cambridge: MIT Press, 2016), 14.

11. Honorable Deborah Lee James and General Mark A. Welsh, *Air Force Future Operating Concept, A View of the Air Force in 2035* (September 2015), 9, footnote number 5.

12. Victor Mayer-Schönberger and Kenneth Cukier, *Big Data* (New York: Houghton Mifflin Harcourt, 2013), 6.

Big Data Needs and Challenges for SOF Pre-Conflict Activities

Big Data Requirements

Compared to the civilian commercial sector, the U.S. military lags in big data exploitation. Robert O. Work, 32nd Deputy Secretary of Defense, explains "big data, advanced computing, miniaturization, robotics, artificial intelligence, and nanotechnology, among others, and all these things are being driven by the commercial sector."[13] Will Roper, the former Director of the Pentagon's Strategic Capabilities Office, also argues "the Pentagon's lack of recognition that data is going to be one of the primary tools, and fuels, and weapons, in future warfare. We don't treat data the same way that a company like Google, or Apple, or Amazon do."[14] He emphasizes the importance of DOD saving and storing all data for machine learning to help operators use data as a strategic resource.[15] Additionally, the Defense Science Board (DSB), a group of retired military, government, and industry experts, provided scientific and technological advice to the U.S. DOD regarding information technology. The board emphasized in several reports the need for a whole-of-government, population-centric principle beyond legacy intelligence sources, exploitation of data from social media and Internet of Things, data science, and data analytics to better operate in unconventional military operations, to include stability efforts, COIN, and constrained operations.[16] Furthermore, the *Marine Corps Intelligence, Surveillance, and Reconnaissance Enterprise Plan* stresses the importance of capitalizing on open source information, social media, and innovative ways to increase knowledge of complex, volatile, uncertain environments.[17] Fortunately, the DOD has recognized the importance and need for big data exploitation. As highlighted in the Joint Chiefs of Staff's *Capstone Concept for Joint Operations: Joint Force 2030*, the Joint Force must improve the "synthesis of big data and the fusion of traditional and non-traditional intelligence."[18] Big data requirements exist in the DOD, however, a U.S. Secretary of Defense and Chairman of the Joint Chiefs of Staff-led effort is required to capitalize on it.

Military organizations can use big data to bolster mission effectiveness, especially in future conflicts dominated by contested norms and persistent disorder, as described in *Joint Operating*

13. "An Interview with Robert O. Work," *Joint Force Quarterly* 84 (2017), 6, http://ndupress.ndu.edu/JFQ/Joint-Force-Quarterly-84/Article/1038783/an-interview-with-robert-o-work/.

14. Will Roper and Nicholas Thompson, *The Future of Warfare-SXSW 2017* (June 2017), video/interview, 25:20 to 26:45 min, retrieved from https://www.youtube.com/watch?v=GLh_ApVVBU4.

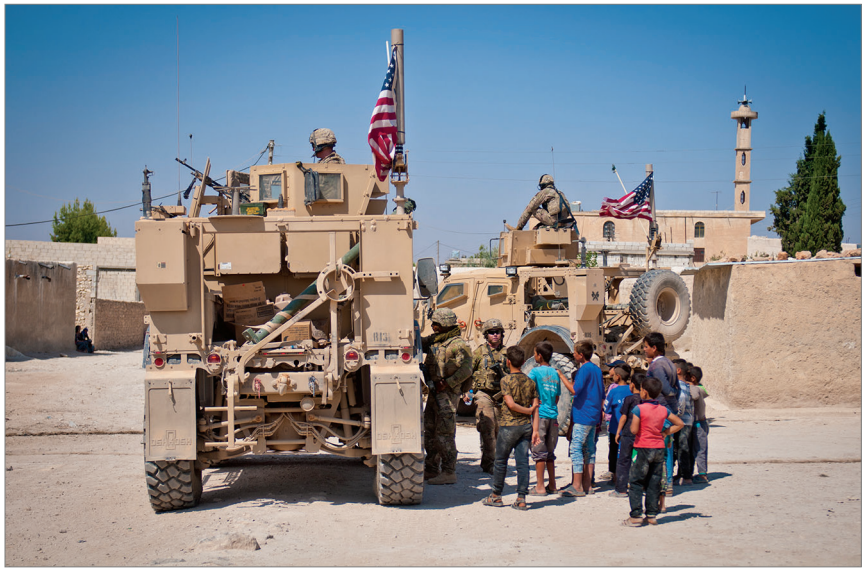
15. Roper and Thompson, *The Future of Warfare.*

16. Defense Science Board, *Seven Defense Priorities for the New Administration* (December 2016), http://www.acq.osd.mil/dsb/priorities.htm; Defense Science Board, *Counterinsurgency (COIN) Intelligence, Surveillance, and Reconnaissance (ISR) Operations* (February 2011), http://www.acq.osd.mil/dsb/reports.htm; Defense Science Board, *Summer Study on Capabilities for Constrained Military Operations* (December 2016), http://www.acq.osd.mil/dsb/reports/2010s/DSBSS16_CMO.pdf.

17. Headquarters Marine Corps, *Marine Corps Intelligence, Surveillance, and Reconnaissance Enterprise Plan, 2015-2020* (September 2014), 7.

18. Joint Chiefs of Staff, *Capstone Concept for Joint Operations: Joint Force 2030* (Draft June 2016), 10.

Environment 2035.[19] Exploitation of big data is particularly applicable to SOF, where there are high expectations for SOF personnel to be skilled at operating in the human domain with the complexities associated with understanding the social, cultural, physical, informational, and psychological elements.[20] This proves exceptionally challenging due to a diverse, complex world filled with many cultures, customs, and languages, even within the same region. However, information from big data may allow SOF to function more effectively in the human domain by enhancing SOF situational awareness. Additionally, SOF often operate in small units requiring small-footprint efforts; thus, the need for actionable intelligence from big data to ensure mission success is even more critical.



A group of children gather around a U.S. military convoy patrolling their village outside Manbij, Syria, on 14 July 2018. The U.S. and Turkey began conducting independent, coordinated patrols to help reinforce the safety and security in the region after the elimination of the Islamic State of Iraq and Syria. U.S. ARMY PHOTO BY STAFF SGT. TIMOTHY R. KOSTER/DVIDS

For instance, in the case of UW, where SOF operate in a denied environment and it is difficult to gather intelligence on the local population, big data from open sources offers population-centric insights. Live radio broadcasts and public social media websites or platforms are easily monitored and lead to valuable open-source intelligence (OSINT). Examples include knowledge about citizens' grievances and perceptions, and situational awareness on social, economic, political, geographical, and humanitarian issues. An increase in negative sentiments or an upward trend in protests may trigger a decision for human intelligence (HUMINT), geospatial intelligence (GEOINT), and other intelligence collection means.[21] As explained in JP 3.0, *Joint Operations*, "Where there is little U.S. Government or U.S. military presence, OSINT may be the best immediately available information to prepare U.S. forces to operate in a foreign country."[22] Big data from open sources also help verify or deconflict other intelligence sources to assist with decision-making and implementing actions.[23]

19. Kevin D. Scott, *Joint Operating Environment (JOE) 2035: The Joint Force in a Contested and Disordered World* (Washington, D.C.: Joint Chiefs of Staff, July 2016), ii, http://www.dtic.mil/dtic/tr/fulltext/u2/1014117.pdf.
20. United States Special Operations Command, *Operating in the Human Domain*, Version 1.0, (August 2015), 1, 12–17.
21. *Joint Operations*, JP 3.0, (Washington, D.C.: The Joint Staff, January 2017), VII-7.
22. *Joint Operations*, JP 3.0.
23. *Joint Operations*, JP 3.0, VIII-27.

On the other hand, excessive data may exasperate the problem and stall decision-making. With the help of machine learning, the vast amount of big data collected is sorted leaving behind unused, irrelevant data. The selected data is further analyzed, processed, and translated into actionable intelligence along with other intelligence sources for quicker dissemination to decision makers. Moreover, big data analytics offer insights on targeted areas that SOF should not focus on, therefore saving valuable time and resources. Exploitation of big data from various open sources may also reveal trends and patterns for predictive analysis and indicators of potential conflicts.

Operational and Technical Challenges, Limitations, and Issues

Like the commercial sector, the DOD faces challenges associated with big data's attributes of volume, variety, velocity, and veracity. Fundamental questions include: Where should SOF collect data from and how is it collected? Is it reliable? In Zeynep Tufekci's article "Big Questions for Social Media Big Data: Representativeness, Validity, and Other Methodological Pitfalls," interpretation of data can be challenging since both sociocultural interactions and human behaviors are variable and complex and not easily understood by algorithms.[24] The local population's sentiments easily fluctuate based on their current environment and condition. In addition, the over-reliance on data from one type of social media source to study social activity may not produce an accurate assessment of an overall population's social issues. For example, Pew Research Center found younger internet users tend to use Twitter more than older adults.[25] Additionally, not all members from developing nations have access to cellular phones, internet, or information infrastructures. In both cases, the data collected is from a small group able to afford and access technologies, which may produce biased conclusions toward age or a more affluent group and not represent the entire population. However, this should not lead to the automatic dismissal of social media data mining because statistics show the amount of data produced in developing countries continues to grow exponentially. According to the Cisco Visual Networking Index, the world added over 400 million mobile devices in 2016, with the Middle East and Africa having the highest mobile traffic growth rate of 96 percent.[26]

With the help of machine learning, the vast amount of big data collected is sorted leaving behind unused, irrelevant data.

Misuse of information technology can increase violence and conflict, as demonstrated by the 1994 Rwandan genocide, where violent actors used radios to broadcast and spread racial propaganda.[27] Today, the widespread use of social media not only serves as another venue for misuse,

24. Zeynep Tufekci, "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls," *ICWSM 14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (2014), 505, 513.

25. Shannon Greenwood, Andrew Perrin, and Maeve Duggan, "Social Media Update 2016," Pew Research Center website, 11 November 2016, 4, http://www.pewinternet.org/2016/11/11/social-media-update-2016/.

26. Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper," (March 2017), 1, 4, https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

27. Kristin Bergtora Sandvik, Maria Gabrielsen Jumbert, John Karlsrud, and Mareile Kaufmann, "Humanitarian Technology: A Critical Research Agenda," *International Review of the Red Cross* 96, no. 893 (2014): 228.

but it also serves to influence the news cycle (e.g. viral videos) by quickly spreading information that may be inaccurate or malicious. Data is often manipulated or misinterpreted for self-interests. For example, Twitter bombs—defined as "flooding the micro-blogging site Twitter with similar hashtags, keywords, and links using multiple accounts, with the objective of attracting more viewers to a website, product, service, or idea"[28]—amplify fake news, and target ethnic groups, even against U.S. military personnel and units. Thus, when trying to exploit big data for pre-conflict intelligence or operations, SOF must consider the data source and determine if there is any bias. Kristin Bergtora Sandvik, author of "Humanitarian Technology: a Critical Research Agenda," argues that information technology removes the humanity principle because of the remoteness from the scene of conflict.[29] In other words, emotional factors like empathy, when observing a conflict first-hand, may not be correctly sensed by technology.

Even though technology is meant to provide neutrality and impartiality, collection and interpretation of data may not be neutral since humans possess possible bias.[30] Others argue that one of the benefits of big data analytics is the removal of emotional bias to assist in decision-making. However, in certain environments, leaders require instinct and intuition, and an understanding of population sentiments that are not necessarily revealed by big data analysis. One of the biggest concerns dealing with data is safeguarding sensitive, personal information (e.g., location, gender, and financial data), and privacy. Force protection issues also arise, as the case of cellular phone metadata including call logs and location data. Using data analytics, adversaries have the ability to link personal, deployed, and temporary duty locations of service members.

Besides the challenges and limitations of collecting data, it is critical to efficiently and effectively process the vast amount of data into useful information. More specifically, data must be mined to extract useful information through computational methods and visualized in an easy to understand form such as histograms and scatter plots for data scientists and analysts to interpret.[31] Otherwise, overwhelming data or difficult to comprehend reports will be useless, or more significantly, unactionable intelligence. Machine learning, data mining, statistical analysis, and algorithm development help overcome the big data issues of volume, variety, velocity, and veracity. As explained by General Thomas, USSOCOM Commander, "machine learning will help us mitigate thousands of man-hours spent on sorting through vast sums of data," as in the case of processing and analyzing open source information for SOF.[32]

It is essential to understand that big data is not actionable intelligence without analysis, and there are still trust issues in these systems. In Colonel Robert Dixon's article "Bringing Big Data to War in

28. Mary C. Joyce, *Digital Activism Decoded: The New Mechanics of Change* (New York: International Debate Education Association, 2010), 221.

29. Sandvik, "Humanitarian Technology," 241.

30. Sandvik, "Humanitarian Technology," 241-242.

31. Examples of computational methods and data visualization found in Florin Gorunescu, *Data Mining*, Intelligent Systems Reference Library, vol. 12, (Berlin: Springer, 2011), 3.

32. General Raymond A. Thomas, III, *Statement before the Armed Services Committee*, 4 May 2017, 12-13, https://www.armed-services.senate.gov/imo/media/doc/Thomas_05-04-17.pdf.

Mega-cities," the leader's judgment and decision on how to use big data remain critical because "data analytics typically reveals correlation, but does not speak definitively about causation."[33] To elaborate further, data analytics may indicate a correlation between a group's social and political activities with an increase in terrorist activities in the region, yet the targeted group may not be the cause of that terrorist event. SOF data scientists and leaders must understand and be aware of the limitations of big data analytics. As explained by David Drake, technical director of the Communications and Information Directorate at the National Air and Space Intelligence Center, "one big challenge is that the data itself can have specific context and/or semantic meaning in relation to other data that may not have been accounted for during the creation of the big data analytical algorithms."[34] Even though there is significant research on big data like intelligent data mining and cognitive computing,[35] the human factor still plays an essential role in big data analysis.

SOF data scientists and leaders must understand and be aware of the limitations of big data analytics.

Once the data collection, processing, and analysis steps are complete, there remains the challenge of integrating that information with intelligence from other sources (e.g., HUMINT) for decision-making and action. As explained by the USAF Deputy Chief of Staff, Intelligence, Surveillance, and Reconnaissance in the article "Revolutionizing AF Intelligence Analysis," the variety of big data from traditional and nontraditional sources, sensors from new technologies, and open source information contribute to challenges across the range of military operations.[36] However, leaders and users must remember that although big data may enhance situational awareness, it should not be the sole rationale for decision-making. Anthony Townsend, author of *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*, warns that although data sheds light on the visibility of a problem, leaders need to heed caution about insufficient data and not rush toward actions and decision.[37] Decision makers need to verify information through other sources (e.g., satellite imagery, unmanned aerial systems, or in-person observation) because the big data component is just one piece of the intelligence collection puzzle. To illustrate, the United States Marine Corps used the Economic and Political Intelligence Cell in Al-Anbar, Iraq, to enhance their knowledge about political, economic, and tribal relationships—not relying on one component for intelligence.[38]

33. Colonel Robert Dixon, "Bringing Big Data to War in Megacities," *War on the Rocks*, 19 January 2016, 7, https://warontherocks.com/2016/01/bringing-big-data-to-operations-in-mega-cities/.

34. Quoted in John Edwards, "Using Big Data for Real Insight," C4ISR & Networks 14, no. 4 (2014), 16.

35. IEEE, "2016 IEEE International Conference on Big Data," conference website for examples of recent Big Data research, accessed September 25, 2017, http://cci.drexel.edu/bigdata/bigdata2016/index.html.

36. Deputy Chief of Staff, Intelligence, Surveillance and Reconnaissance (DSR/ISR), *Revolutionizing AF Intelligence Analysis*, White Paper (January 2014), 4, http://www.defenseinnovationmarketplace.mil/resources/20140211_IntelligenceAnalysisWhitePaper_PA.pdf.

37. Anthony M. Townsend, *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia* (New York: W.W. Norton & Company, Inc., 2013), 191.

38. Angela Rabasa, et al., *From Insurgency to Stability, Volume II: Insights from Selected Case Studies* (Santa Monica: RAND Corporation, 2011), 175–176, 183.

Big Data Opportunities—Insights and Lessons Learned from the United Nations Program

Although there are many needs and challenges in data collection, processing, and analysis to form actionable intelligence as described above, the scope of this paper is limited to developing recommendations for improvement in these areas by analyzing the UN Global Pulse initiative as a case study. Since the UN has invested in big data analytics beginning in 2012, it is advantageous to examine the program's results and applicability for SOF pre-conflict operations. The UN Global Pulse program can provide SOF greater insights and lessons learned because big data often is derived from human inputs, such as thoughts and perceptions about socio-culture, educational, economic, and political issues reported on social media, radio broadcasts, the internet, and news reports. The DSB describes population-centric as a deep understanding of populations to include "historical, socio-cultural, economic, educational, and environmental aspects of the area of operations."[39] In turn, this enhances situational awareness, indications and warnings, and real-time feedback. For instance, an increase in social media traffic due to political disagreements may be an indicator of possible protests and uprisings. This type of population-centric information and sentiment analysis affords SOF real-time analysis and intelligence and increased knowledge of local concerns and attitudes to warn and possibly prevent conflict relapse.

In "Five Examples of Big Data Analytics and the Future of ISR," Dr. Jon Kimminau, Defense Intelligence Senior Leader, explains how big data relates to intelligence, surveillance, and reconnaissance (ISR): "intelligence discovery (e.g., collect and organize data in a uniform manner across all sources for exploring data to an analyst's workstation); intelligence assessment (e.g., produce near-real-time battle damage assessment); intelligence explanation (e.g., collaborate in a global enterprise where analysts analyze data, not segregated by type of source and regional assignments; intelligence anticipation (e.g., recognize potential situations quicker); intelligence delivery (e.g., produce a broader variety of relevant intelligence to commanders near real-time)."[40] The UN Global Pulse program has also explored multisource information, real-time assessment, and early warning response. Extending the Global Pulse Big Data initiatives for sustainable development goals, humanitarian actions, and conflict prevention, support SOF operations. For example, the UN report *Integrating Big Data into the Monitoring and Development Programmes* describes four areas for implementing their monitoring and evaluation (M&E) big data strategy: 1) descriptive and exploratory analysis, 2) predictive analytics, 3) detection, and 4) evaluation/prescription.[41] For SOF pre-conflict situations, M&E assists with early warning, delivers vital information about emergency incidents, ethnic clashes, or political uprisings, and helps monitor and manage regional stability

39. Defense Science Board, *Counterinsurgency (COIN) Intelligence, Surveillance, and Reconnaissance (ISR) Operations* (February 2011), vii, 27, http://www.acq.osd.mil/dsb/reports.htm.

40. Jon. A. Kimminau, "Five Examples of Big Data Analytics and the Future of ISR," *Joint Force Quarterly* 77 (2015), 30–31, http://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-77/jfq-77_30-31_Kimminau.pdf.

41. Michael Bamberger, edited by Tamara Karaica and Felicia Vacareulu, *Integrating Big Data into the Monitoring and Evaluation of Development Programs*, UN Global Pulse report, (2016), 22, 57–59.

before further escalation. Along with other intelligence sources, it reveals a broader perspective of regional and local conditions and aids in the decision-making whether SOF need to engage or not.

Another benefit of analyzing the UN Global Pulse program is that collected data is published in reports available on the UN Global Pulse website and those reports include a description of the open source tools used for analysis. However, there are instances where the data, tools, and analysis report are not publicly releasable, and it may be more difficult for SOF or any military organization to obtain. Because nongovernmental organizations (NGOs) and humanitarian agencies support and lead many of the UN big data programs, extra caution must be taken to not violate humanitarian space and the necessities of neutrality, impartiality, and independence.[42] Intelligence obtained from big data analytics can be of dual-use to support humanitarian and military missions, but many NGOs are hesitant or do not want to work directly with the U.S. military. Additional considerations include active data collection, such as surveys from local population and crowdsourcing, which may be more difficult for SOF to obtain, especially in situations where their presence is not welcome by the host. Besides, surveys may not always produce an accurate representation since respondents may not give honest answers or can interpret questions differently. In JP 2-01.3, *Joint Intelligence Preparation of the Operational Environment*, big data sources such as postal facilities, radio stations, and social media, are listed as information nodes useful to joint operations.[43] The following UN Global Pulse case studies examine online news and digital content, social media, radio, and multiple big data sources and associated data analytics to sort through the vast amount of data and produce actionable intelligence to assist SOF with increased situational awareness, early warning, predictive analysis, decision-making, and operating in the human domain.

Case Studies: Online News, Digital Content, and Economic Data

By analyzing trends and areas of concern, SOF gain greater situational awareness before conflict occurs. For instance, Russia saw a deterioration of its economic status starting in 2009, resulting in protests by the urban middle class in late 2011 to 2012.[44] Once a sign of instability is exposed, SOF can collect additional intelligence focused in that area for monitoring and possible preventive and mitigating actions. To demonstrate further, the UN Global Pulse conducted a study in 2016 analyzing publicly available global news media records supported by Google Ideas—Google's think tank, now known as Jigsaw—to determine the feasibility of utilizing big data for early warning and conflict in support of their humanitarian program.[45] The data mining study focused on Tunisian citizens' feelings from online news data, and recorded the average volume and tone of articles over time, the type of actors in-country, and events discussed.[46] Results indicated a significant, rapid, and

42. Robert Perito, *Guide for Participants in Peace Stability, and Relief Operations* (Washington D.C.: United States Institute of Peace Press, 2007), 118.

43. *Joint Intelligence Preparation of the Operational Environment*, JP 2-01.3, (May 2014), III-42–43.

44. United States Army Special Operations Command, *"Little Green Men,"* 37.

45. United Nations Global Pulse, "Feasibility Study: Analysing Large-Scale News Media for Early Warning of Conflict," *Global Pulse Project Series*, no. 3, (2016), 52–53, https://www.unglobalpulse.org/big-data-development-case-studies.

46. United Nations Global Pulse, "Feasibility Study."

overwhelming increase in negative tone against Tunisia's military and police forces in January 2011, during the government transition and departure of the former Tunisian President.[47] Although the UN's in-country staff were aware of the tumultuous situation in 2011, the study concluded that data mining of news articles offered "useful insights about the overall conflict-related trends" in Tunisia, and demonstrated the potential to monitor conflict and assess risk, but recommended social media and other open source information for integrated understanding of conflict analysis.[48]

The UN program demonstrates that data mining of online news and digital content provides a simple, expedited way to analyze the vast amount of data for SOF missions to enhance situational awareness about actors (e.g., nation's military or government) and key political and social events. Furthermore, compilation and analysis of big data from various sources serve as predictive analysis for SOF by revealing abnormal trends and patterns, especially where in-country intelligence may not be available or is difficult to obtain. Monitoring of the news media shows perspective and indications on adversaries' actions, such as the case with Russia. An example of information warfare is the conflict with eastern Ukraine from 2013-2014, when Russia applied asymmetrical tactics by using television news, social media, radio propaganda, and deception, as key components.[49] Russia justified its intervention in Ukraine because of the "people's choice" through news media aimed at Western audiences.[50]

The UN program demonstrates that data mining of online news and digital content provides a simple, expedited way to analyze the vast amount of data for SOF missions to enhance situational awareness about actors (e.g., nation's military or government) and key political and social events.

SOF operate in the human domain, but the expectations to know all aspects of population dynamics are quite challenging. The importance of sociocultural analysis of "adversaries and other relevant actors that integrates concepts, knowledge, and understanding of societies, populations, and other groups of people, including their activities, relationships, and perspectives across time and space at varying scale" is highlighted in JP 3-24, *Counterinsurgency*, to better understand the operational environment.[51] Exploitation of big data may provide results similar to face-to-face HUMINT collection while reducing the risk and dangers associated with working in an insurgent-embedded environment.

SOF personnel face operational pressure to adapt and quickly respond in an emerging global environment characterized by the importance of global commons, a wide range of operating environments, and geopolitical instability. Knowledge of regional, national, or local-level economic interests and issues aid SOF with greater situational awareness on the global environment. The UN Global Pulse program is using big data analytics to comprehend macroeconomic issues better.

47. United Nations Global Pulse, "Feasibility Study."
48. United Nations Global Pulse, "Feasibility Study," 53.
49. United States Army Special Operations Command, *"Little Green Men,"* 1–3.
50. United States Army Special Operations Command, *"Little Green Men,"* 1–3, 48.
51. *Counterinsurgency*, JP 3-24, (November 2013), IV-3.

For example, the UN Global Pulse Jakarta and the UN Capital Development Fund in Bangkok are analyzing financial data in Cambodia as an early warning system for economic problems.[52] The second and third order effects of economic turbulence are potential increases in political and economic instability and conflicts. Alex Braithwaite in his article "Does Poverty Cause Conflict? Isolating the Causal Origins of the Conflict Trap," indicates the causal connection between poverty and conflict,[53] supporting the consideration of using economics and exploiting financial data as a possible pre-conflict indicator. In addition, the UN Global Pulse conducted a project in 2011 on monitoring food security issues by evaluating over 20,000 news articles from hundreds of sources, using automated text-mining network analysis tools.[54] Geolocation tagging of news articles also allowed researchers to map economic issues worldwide.[55] An online video presentation of the project illustrated the interactive web-based platform and easy to comprehend visual results.[56] This big data initiative supports using online media mining to track emerging trends to assist decision makers during economic issues and crisis.

These case studies reveal opportunities for SOF to exploit big data for greater situational awareness about a region, however, analysis of news articles and digital information may not necessarily deliver a comprehensive, accurate assessment of a nation's conflict. Additional open source information like social media postings yield a broader perspective.

Case Studies: Social Media

Using UW as an example, monitoring social media accounts with the appropriate analysis tools assists SOF by providing more relevant knowledge of adversaries' intentions, which can then be used to counter those threats. According to the USASOC report *"Little Green Men,"* professors Igor Panarin and Alexandr Dugin explain two schools of thought on geopolitics and information warfare in Russia.[57] Panarin favors media manipulation and propaganda tactics to influence the population, and similarly, Dugin advocates for the use of Facebook, Twitter, and other social media tools for Russian information warfare against the West.[58] The employment of SOF in similar UN Global Pulse big data initiatives helps monitor local perceptions on socio-culture, educational, economic,

52. United Nations Global Pulse, "Projects, Using Financial Data to Understand Macroeconomic Issues in Cambodia," accessed 1 December 2017, official website, https://www.unglobalpulse.org/projects/using-financial-data-understand-macroeconomic-issues-cambodia.

53. Alex Braithwaite, Niheer Dasandi, David Hudson, "Does Poverty Cause Conflict? Isolating the Causal Origins of the Conflict Trap," *Conflict Management and Peace Science* 33 (2016): 60.

54. United Nations Global Pulse official website, "Monitoring Food Security Issues through News Media (2011)," accessed 2 December 2017, https://www.unglobalpulse.org/projects/news-awareness-and-emergent-information-monitoring-system-food-security%20. The team identified vital food security themes over a span of eight years, and results revealed that the news initially focused on humanitarian issues in 2008 during the global Great Recession, followed by interests in volatility in food prices in 2008, then social unrest in 2011.

55. United Nations Global Pulse official website, "Monitoring Food Security Issues."

56. Jean Philippe Cointet, *Mining News for Emerging Trends*, 2011, online video and slides, 11 min., https://www.unglobalpulse.org/projects/news-awareness-and-emergent-information-monitoring-system-food-security%20.

57. United States Army Special Operations Command, *"Little Green Men,"* 15–17.

58. United States Army Special Operations Command, *"Little Green Men,"* 15–17.

and political issues, and identify propaganda and violent incidents reported through social media and other open sources. Moreover, with advanced information technology increasing the speed and scope of communication dissemination, exploitation of data analytics quickly assist SOF by informing them of adversaries' propaganda and influence on the population during conflicts. Furthermore, big data analytics on social media platforms help verify incidents reported through other intelligence means. For example, the chemical weapon attack in Syria in 2013 shows the importance of using multiple data streams for verification. Social media reports by local Syrians, as well as videos and reports by international humanitarian organizations and hospitals, provided credibility of the attack.[59] Big data analytics aid in the decision-making process when SOF personnel require entry into a region by offering additional intelligence verification.

UN Global Pulse has utilized social media in several case studies including publicly available Twitter data to monitor population perceptions. SOF gain a greater understanding of local population negative sentiments and tensions against the government and awareness of any potential global instability. One of the benefits of machine learning is the ability to sort through the massive amount of social media data, thus preventing unnecessary consumption of SOF personnel time and resources. To illustrate, the UN's Mining Citizen Feedback Data for Enhanced Local Government Decision Making program used Twitter posts, text messages, and internet feedback from Indonesia's national complaint system called LAPOR[60,61] (Citizens' Aspirations and Complaints Online System). Utilization of social media geospatial information allowed researchers to determine if the public's feedback came from the region of interest.[62] Resulting graphs uncovered trends and patterns related to citizen's concerns about poverty, energy, food sufficiency, and other critical Indonesian government national priority areas.[63] Lessons learned include the integration of real-time data from a variety of sources for more accurate assessments.

An example of Global Pulse Big Data initiatives for sustainable development goals and humanitarian actions for SOF pre-conflict operations is Haze Gazer—a web-based analysis tool using data from multiple real-time social media, online photo sharing programs, video channels, satellites, population census, and a citizen complaint system for haze and fire disaster management in Indonesia.[64] Not only are the program's objectives to supply real-time information on fire and haze hazards and vulnerable populations, it also anticipates crisis events, public movement patterns, and

59. Chandler Atwood, "Activity-Based Intelligence," *Joint Force Quarterly* 77 (2015): 29.

60. United Nations Global Pulse, "Mining Citizen Feedback Data for Enhanced Local Government Decision-Making," *Global Pulse Project Series* no. 16, (2015), 1-2, http://www.unglobalpulse.org/sites/default/files/UNGP_ProjectSeries_Citizen_Feedback_2015.pdf.

61. Hans Nicholas Jong, "Govt Adopts LAPOR! Online Public-service Complaint System," *The Jakarta Post*, 20 December 2014, http://www.thejakartapost.com/news/2014/12/20/govt-adopts-lapor-online-public-service-complaint-system.html.

62. Jong, "Govt Adopts LAPOR!" The process involved removal of irrelevant information, conversion of unstructured data to structured information, and various analysis of data.

63. Jong, "Govt Adopts LAPOR," 2.

64. United Nations Global Pulse, "Haze Crisis Analysis and Visualization Tool," Tool Series, no. 2, (2016), 1-2, http://unglobalpulse.org/sites/default/files/2-pager%20Haze%20Gazer%20-%20Feb%202017_0.pdf.

in-situ behaviors to prepare the government for crisis response.[65] When SOF utilize a variety of tools, they allow for quicker analysis of regional instability and implementation of any preventive measures. Conversely and equally important, big data analytics for situational awareness of population dynamics reveal areas that SOF should not engage in because an uptick of negative sentiments may be the norm compared to past trends.

Case Studies: Radio Data

Another open source medium for the UN's exploitation of big data that could be beneficial to SOF are radio broadcasts. There are regions of the world, such as parts of Africa, where they have not updated communications and information technology infrastructure to the newest systems. Radios may be the primary source of communication, and this would serve as a source for gathering information about public sentiments, issues, and concerns. The UN is conducting a case study demonstrating the employment of radios and machine learning to analyze discussions on community radio stations in Uganda. Led by the UN Pulse Lab Kampala, the overall focus of the program exploits radio and data analytics to fill information gaps for decision-making to achieve their sustainable development goals (e.g., peace and justice, reduced inequalities, and economic objectives).[66] The program employs the Radio Content Analysis Tool to capture relevant radio discussions about poverty, education, diseases, financial and work concerns, and violence in Uganda, utilizing a speech recognition technology to transform the discussions into text, and delivers real-time analysis using a web interface.[67] The technology is not fully automated because it still requires a small evaluation team to capture false detections and categorize the relevant topics by categories, therefore a much smaller, more manageable structured dataset is produced. Analysts are also required to translate relevant data into English.

When SOF utilize a variety of tools, they allow for quicker analysis of regional instability and implementation of any preventive measures.

Challenges remain in the speech recognition technology due to different dialects, lack of clarity of speakers, poor reception, topics not discussed over the radio, and unknown demographics of contributors, which may lead to biases.[68] Although Google, Microsoft, and other companies are investing in speech recognition technologies, they have not captured all languages and dialects around the world, especially in Africa and Asia. Despite those challenges, these studies show potential usefulness of radio toward all-source intelligence. For SOF operating in nonurban, austere areas where radio broadcasts are the primary source of communication, utilizing big data analytics quickly remove irrelevant topics and help analysts focus on issues of concern. SOF personnel's proficiency of the local language is critical for many missions, but realistically it is challenging for them to be regional, national, or local experts in a world diversified with many cultures, languages,

65. United Nations Global Pulse, "Haze Crisis Analysis."
66. United Nations Global Pulse Kampala, *Using Machine Learning to Analyse Radio Content in Uganda*, United Nations Global Pulse report (September 2017), 3.
67. United Nations Global Pulse Kampala, *Using Machine Learning*, 7–11.
68. United Nations Global Pulse Kampala, *Using Machine Learning*, 22.

and dialects, even at the local level. With improved machine learning tools capable of translating vast amount of information in indigenous languages to English from radio broadcasts or other audio sources, SOF personnel gain quicker insights on cultural, social, and information elements to interpret the environment better. It also saves considerable time and effort not having to translate and sort through irrelevant information.

Conclusion and Recommendations

Technology advancements have changed the way we prepare for wars today and will have a considerable impact on how we fight in the future. Russia and other near-peer adversaries are already using open source social media, news, and radio broadcasts for their information warfare to influence populations and politics. One of the ways to gain more significant insight about an adversary's intention is through the exploitation of big data—the ability to take a vast amount of open source information, process, and analyze it for useful and possibly actionable intelligence. Because of the specialized core activities of SOF, which require operations in the complex human domain and in small-footprint efforts with limited resources, exploitation of big data using machine learning can significantly assist SOF in several ways. Exploitation of big data can provide enhanced situational awareness, produce real-time actionable intelligence, filter excessive, irrelevant data that can stall decision-making, deliver insights on where not to target, and offer predictive analysis and indicators of potential conflicts.

There are many operational and technical challenges to big data. This paper, based on open source, publicly available literature, examined the UN Global Pulse activities as case studies for lessons learned and opportunities for SOF operations—the UN program began employing big data in 2012. Although the UN's mission and SOF operations differ, both require an understanding of population, economics, and political issues. The UN program has shown that big data from news media, financial data, social media, and radio broadcasts provide greater insights and situational awareness, predictive analysis, and early warning detection systems through increased understanding of societies, populations, and events for SOF operations in the human domain. Big data analytics also save valuable resources by sorting through massive amounts of irrelevant information and delivering actionable intelligence to aid in decision-making. However, it is crucial not to rely on one data source, remembering that open source information is just one contributor to all-source intelligence.

Further recommendations include strengthening collaborations with other government agencies such as the Air Force Research Laboratory, the Defense Advanced Research Projects Agency, and commercial industries conducting research and development of data mining, machine learning, and artificial intelligence to help reduce big data volume, variety, velocity, and veracity challenges. The commercial sector is leading in the development of these information technologies, and the U.S. military should capitalize on them. Moreover, the integration of open source unclassified data with other intelligence sources, interoperability, data management, architecture, and platforms, policies, culture, and personnel, remain areas that require further examination. The *Air Force Future Concepts:*

A View of the Air Force in 2035 and the *Joint Operating Environment 2035* envision a force exploiting big data with machine learning and automation to increase the speed and accuracy of decision-making. The urgency and need to foster technological innovation to support U.S. military missions dramatically increases as adversaries narrow the technological gaps.

Joint Special Operations University
Center for Strategic Studies
7701 Tampa Point Blvd.
MacDill AFB, FL 33621

JOINT SPECIAL OPERATIONS
JSOU
UNIVERSITY
PRESS