EMBRACING BIG DATA Analytical Techniques

by Colonel Jeremy Hartung and Major Eric Nolan

Introduction

Several recent articles in the *Military Intelligence Professional Bulletin* have argued for the importance of incorporating big data analysis into the military intelligence (MI) toolkit.[1] The consensus is that increasing amounts of information and sources of data are overwhelming both technical and cognitive capabilities within our intelligence sections. Proposed solutions to the gap in big data analysis tend to focus on support from industry experts or the purchase of new analytic software. While both of these are important aspects of improving the Army's big data capability, these solutions ignore innovative analytical techniques that MI professionals could employ today. Furthermore, methodologies like the Multi-INT Spatial Temporal (MIST) toolsuite, developed as part of the Fusion Analysis Development Effort (FADE) program, provide a sufficient platform to apply the principles of big data analysis without waiting for the perfect solution in a future system.

The 2016 textbook *Activity-Based Intelligence: Principles and Applications* provides a foundational analytic approach to big data analysis for MI professionals.[2] Activity-based intelligence, also known as ABI, can be applied at the brigade combat team (BCT) level immediately to leverage underutilized sources of data and fill intelligence gaps for the commander. This article will describe ABI's four main ideas, or pillars, with examples of their immediate utility at the BCT level.

What is Activity-Based Intelligence?

"ABI is an analysis methodology that rapidly integrates data from multiple sources to discover relevant patterns, determine and identify change, and characterize those patterns to drive collection and create decision advantage….ABI practitioners have advanced the concept of large-scale data filtering of events, entities, and transactions to develop understanding through spatial and temporal correlation across multiple data sets."[3]

The Development of Activity-Based Intelligence

ABI was born out of the Special Forces intelligence community during the height of the wars in Iraq and Afghanistan.[4] Targeting individuals in the modern interconnected world allowed the Special Forces to leverage more sources of data than ever before. However, as they did this, they became overwhelmed and developed ABI as an approach to work with big data. As the name implies, ABI starts with observed activity. All activity must happen at a defined time and place. If analysts can gain access to data sets of events, including locations and times, they can make spatiotemporal correlations between events that would otherwise seem to be unrelated or unimportant. This key insight led to the development of ABI and its four pillars:

- Georeference to discover.
- Integration before exploitation.
- Sequence neutrality.
- Data neutrality.[5]

A Summary of the Four Pillars[6]

- **Georeference to Discover:** Focusing on spatially and temporally correlating multi-intelligence (INT) data to discover key entities and events.
- **Integration before Exploitation:** Correlating data as early as possible, rather than relying on vetted, finished products (from single INT data), because seemingly insignificant events in a single INT may be important when integrated across multiple INTs.
- **Sequence Neutrality:** Understanding that we have the answers in the data collected at any time to many questions we do not yet know to ask.
- **Data Neutrality:** The premise that all data may be relevant regardless of the source from which it was obtained.

Georeference to Discover

Georeference to discover refers to the ability to add location and time information to data sets, allowing for geospatial and temporal correlations. The resulting correlations are "discovered" as a result of structuring the data to allow a quick comparison of the locations and times of these events. Using methodologies like the FADE program's MIST toolsuite, users can pull in many disparate data sets and perform geospatial discovery. The key is pulling in the right data and ensuring it is properly georeferenced.

The BCT offers many good examples of how to effectively employ georeference to discover. In fact, the BCT has tremendous advantages over higher echelons for analyzing georeferenced data because it has a narrow geographic focus. However, not all the data available at the BCT is readily georeferenced in databases. The georeferenced data is rarely in a format that allows machines to digest it easily and enable a geospatial discovery environment. This is due to a complex combination of problems with system interoperability and a systemic failure to train using mission command systems in accordance with their design. For example, most BCT tactical operations centers do not send reports using the United States message text format (USMTF) between mission command systems. Instead, they depend heavily on text chat programs like TransVerse, mIRC chat, or Joint Battle Command-Platform texts that machines do not automatically scrape for georeferenced data to include in databases. Soldiers have to read the chat message text line by line and relay relevant information to the rest of the BCT headquarters. However, some BCT S-2 sections have been using applications like Rip-It or ChatSurfer to scrape georeferences from text services.[7] These programs can read thousands of lines of text and provide an overlay with all the locations and times of events mentioned in the message traffic.

> "An ABI analyst correlating activities and resolving objects will enable real-time tipping and cueing of sensors, thereby driving collection, again, in ways that cannot be done today.[10]"

Georeferencing the text applications of tactical operations centers will allow intelligence analysts to easily ingest reports from subordinate units into their larger geospatial discovery environment. Automating the process through ChatSurfer will allow faster discovery of previously unknown correlations between subordinate unit reports and all other intelligence reports, making the brigade significantly more responsive to the environment. Scraping text applications is just one example of how to apply georeference to discover. Enterprising analysts across the Army are likely to find many more uses once they are aware of the concept.

Integration before Exploitation

The second pillar of ABI is integration before exploitation. Integration in this context is *fusion*, defined in ADP 2-0, *Intelligence*, as "consolidating, combining, and correlating information together."[8] Fusion occurs in the traditional intelligence process during the produce step after collection tasking, information collection, and information processing.[9] Preliminary exploitation and processing prioritize and limit the amount of information passed to all-source analysts for fusion. In the traditional intelligence process, limiting the information passed to all-source analysts for fusion is a positive feature because traditional information management techniques could easily overwhelm these analysts. However, the ability to perform geospatial discovery in ABI depends on having access to all the available data, not just the bits that single-source reports provide to answer priority intelligence requirements. Performing ABI requires analysts to have access to all data before making judgments about the information's relevance or importance.

The clearest example of integration before exploitation in the BCT is intelligence data that the integrated broadcast service provides from overhead collection systems. Analysts in the BCT can currently access the integrated broadcast service layer in near real time through the Joint Tactical Terminal on the Tactical Ground Station or through the MIST toolsuite on the SECRET Internet Protocol Router Network (SIPRNET) and Joint Worldwide Intelligence Communications System. S-2 sections often assign this data analysis to signals intelligence analysts and geospatial intelligence analysts because they are better suited to perform the analysis and provide a finished intelligence report. However, new programs like FADE and the MIST toolsuite require only a few hours of training and practice for most all-source analysts to learn how to manipulate the data. Then they can compare multi-INT reporting in a specific geographic area to discover previously unknown correlations. The MIST toolsuite provides intuitive geospatial discovery environments that make integration before exploitation possible. Rather than the data overwhelming analysts, ABI allows analysts to process more and more data efficiently to fill intelligence gaps for the commander.

Sequence Neutrality

ABI's third pillar is sequence neutrality. Its basic premise is to recognize that establishing temporal causality is difficult when looking back at events that have already occurred. Performing discovery requires the analyst to be conscious of the logical fallacy—*post hoc, ergo propter hoc* (in other words, after this, therefore because of it).[11] In layman's terms, this is a logical fallacy in which one might believe an event caused another event simply because it happened first. This fallacy serves as a warning to ABI analysts that they cannot be satisfied with simply establishing causality. When analysts believe they have discovered what caused an event, they may stop looking for other potential causes or indicators of it. Instead, they should seek all indicators of an event. More sources and types of data will provide more clarity on the various indicators of an event. The benefits of this analytical work might result in identifying new indicators of enemy activity that could be included in future priority intelligence requirements for collection tasking. It could contribute to building enemy doctrinal templates by adding nuance to the activities the enemy performs in certain circumstances. In current operations, it could result in targeting information that contains more detail.

Sequence neutrality has implications for the types of data that units need to access and store. The ABI analyst conducting geospatial discovery does not necessarily know what data will lead to actionable intelligence. Similarly, sequence neutrality suggests the ABI analyst needs access to data from a broad period of time. These requirements mean that the ABI analyst will depend on data from outside the BCT. Unfortunately, accessing external data depends on establishing robust primary, alternate, contingency, and emergency (PACE) plans. The BCT must overcome concerns about operating in denied, intermittent, and limited bandwidth environments using redundant PACE plans that have different transport layers. A BCT S-2 section today can gain access to the integrated broadcast service layer using three different types of transport: satellite communications over the Joint Tactical Terminal, tactical SIPRNET over the Warfighter Information Network-Tactical, and Tactical Data Network-1 through the TROJAN. This PACE plan will only improve with the future introduction of the Tactical Intelligence Targeting Access Node (TITAN) system. The crucial lesson from this pillar is that units need to recognize the absolute necessity to access data from echelons above brigade and then to prioritize, resource, and train their PACE plan to be successful in modern large-scale ground combat operations.

Data Neutrality

The fourth pillar of ABI is data neutrality. This pillar serves as a reminder that the best intelligence is not always highly classified. The best intelligence is simply that which provides timely and relevant support to the commander's decision making. Top secret information is not better than secret information. The sources and methods used to obtain the information are simply more sensitive, and we must protect them more carefully. With this in mind, data neutrality requires ABI analysts to fully understand the capabilities and limitations of MI systems, the systems of the other warfighting functions in the BCT, and even the systems in adjacent or supporting units to make sure they do not overlook valuable data. The modern battlefield is littered with sensors providing data for all sorts of varying purposes. Identifying opportunities to pull more data into the ABI geospatial discovery environment is a critical part of planning for any operation.

The BCT S-2's running estimate will often list the available organic collection platforms. It may also list the requested echelons above brigade support platforms. These running estimates must also include the Q-36 and Q-50 counter-battery radar systems—not just to note their presence in the BCT or to attempt tasking them as collection platforms but as a reminder to incorporate their data into the geospatial discovery environment. Supporting aircraft are another source of

data. Most have air defense threat warning systems that can report the locations of enemy air defense assets. Comparing aircraft threat warning data to data from echelons above brigade and overhead systems' sensors could result in enough target fidelity to enable the destruction of high-payoff targets. Pulling these sources of data into the geospatial discovery environment might be as simple as automating the text-scraping applications previously discussed or using USMTF messages between mission command systems to deliver machine-readable location data. Either way, the S-2 section must plan how it will receive the data and ensure the architecture supports the proposed plan. The S-2 section should seek out any data source that can provide the location and time of an event and pull it into the geospatial discovery environment.

Application of Activity-Based Intelligence in BCTs

BCTs today could employ the four pillars of ABI to improve analytic outcomes and fill gaps for the commander. The key component to being prepared to perform ABI on the battlefield is being able to train in data-rich environments. The National Training Center (NTC) has been leading the effort to build a modern data-rich intelligence training environment, enabling just such a training opportunity. Leaders at the NTC have recognized that BCTs are preparing to fight against highly technical systems with distinct signatures, yet training environments do not typically incorporate these signatures. Rather, finished reporting is often pumped directly into databases, denying intelligence sections the ability to do their own data analysis.

NTC is approaching this problem by hand-scripting data into the scenario, giving rotational training units the opportunity to "eat the data raw" rather than relying on external processing, exploitation, and dissemination support. The NTC is also experimenting with the automated production of raw reporting in order to create the vast volume of reporting that would be realistic in any future technology-enabled environment. In fact, units unprepared to deal with the overwhelming amount of data available at the NTC will struggle to gain a situational understanding of the operational environment.

Conclusion

Industry is offering advanced technical solutions, but the MI community cannot wait for the fielding of new systems to start developing new doctrinal approaches to analyzing big data. ABI offers an analytic approach that is ready to fill the current gap. The examples in this article are just a few ways BCTs could immediately employ ABI to help leverage underutilized sources of data.

Armed with the four pillars of ABI, analysts across the Army will discover untapped sources of data they could quickly georeference and pull into geospatial discovery environments for improved analytical outcomes. Big data is a reality on the battlefield now, and the MI community should embrace ABI to keep pace with that reality.

Endnotes

[1] Jason Boslaugh and Zachary Kendrick, "The Application of Data Science in the Intelligence Warfighting Function," *Military Intelligence Professional Bulletin* 45, no. 4 (October–December 2019): 57–63; Iain J. Cruickshank, "On Data Science and Intelligence Analysis," *Military Intelligence Professional Bulletin* 45, no. 3 (July–September 2019): 29–32; and 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.

[2] Patrick Biltgen and Stephen Ryan, *Activity-Based Intelligence: Principles and Applications* (Norwood, MA: Artech House, 2016).

[3] Patrick Biltgen, Todd S. Bacastow, Thom Kaye, and Jeffrey M. Young, "Activity-Based Intelligence: Understanding Patterns-of-Life," United States Geospatial Intelligence Foundation, April 18, 2017, https://medium.com/the-state-and-future-of-geoint-2017-report/activity-based-intelligence-understanding-patterns-of-life-481c78b7d5ae.

[4] Biltgen and Ryan, *Activity-Based Intelligence*, 23–25.

[5] Ibid., 33–51.

[6] Lauren Zabierek, "Enabling OSINT in Activity-Based Intelligence (ABI)," Recorded Future, August 31, 2016, https://www.recordedfuture.com/activity-based-intelligence/.

[7] The 1-2 Stryker Brigade Combat Team (SBCT) S-2 section used Rip-It and ChatSurfer with some success during Joint Readiness Training Center rotation 21-02 with the 5th Security Force Assistance Brigade (SFAB). However, since the 5th SFAB was the rotational training unit, and the 1-2 SBCT was role-playing a host-nation force, the brigade had limited access to intelligence feeds that would have made creating a geospatial discovery environment possible.

[8] Department of the Army, Army Doctrine Publication 2-0, *Intelligence* (Washington, DC: U.S. Government Publishing Office, 31 July 2019), Glossary-3.

[9] Ibid., 3-2.

[10] David Gauthier, "Activity-Based Intelligence Definition for the Intelligence Community," National Geospatial-Intelligence Agency, 2013, quoted in Chandler P. Atwood, "Activity-Based Intelligence: Revolutionizing Military Intelligence Analysis," *Joint Force Quarterly* 77 (2nd Quarter, 2015): 26, https://ndupress.ndu.edu/Media/News/Article/581866/activity-based-intelligence-revolutionizing-military-intelligence-analysis/.

[11] Biltgen and Ryan, *Activity-Based Intelligence*, 47–49.

COL Jeremy Hartung is the Senior Intelligence Officer at the National Training Center. He previously served as the 7th Infantry Division G-2 and as the S-2 for the 173rd Airborne Infantry Brigade Combat Team. He is a graduate of the Army Intelligence Development Program–Intelligence, Surveillance, and Reconnaissance.

MAJ Eric Nolan is the Analysis and Control Element Chief for the 1st Multi-Domain Task Force at Joint Base Lewis–McChord, WA. He previously served as the S-2 for the 1-2 Stryker Brigade Combat Team and as the Deputy G-2 for the 7th Infantry Division. He holds a master of arts in strategic studies and international economics from Johns Hopkins University's School of Advanced International Studies.