MITRE

SERIES
Number 11

INTELLIGENCE AFTER NEXT

BREAKING NEWS: DISCOVERING AND SHARING CURRENT INTELLIGENCE AT THE SPEED OF MODERN WARFARE

by Joseph Convery, Hassan Terry, William Wang, & Genevieve Whiddon

Forewarned is Forearmed

Delivering time-sensitive current intelligence to an operational commander, with sufficient warning of an impending threat, has become increasingly difficult as the pace of modern warfare shortens the time in which commanders must decide their course of action.

To remain operationally relevant in crisis and on the battlefield, the Defense Intelligence Enterprise must rapidly recognize the changes, trends, and anomalies that often point toward an impending enemy action – an outcome that is increasingly difficult as advanced collection capabilities offer progressively more raw intelligence for analysts to evaluate.

To address this challenge, MITRE chartered a study known as ‘Breaking News’. This multi-part project evaluates how technology can be applied to both speed the discovery of critical time sensitive intelligence, and then share that information faster with the decision maker.

- The initial study within the ‘Breaking News’ series focused on one specific technology known as an ‘insight engine’. The question posed by the study was “can an insight engine help an analyst recognize important data faster, and subsequently enable that intelligence to be shared more rapidly with the decision maker?”
- The second part of this effort was an initial examination of how a decision maker might visualize current intelligence data more effectively. As new methodologies to speed current intelligence analysis mature, the Defense Intelligence Enterprise must also examine new ways of rapidly sharing that information with military commanders to enable them to make time-sensitive decisions both faster and more accurately.

Finding a Needle in the Field of Haystacks…

As the speed of modern warfare has increased, so has the amount of data gathered by today’s collection capabilities. The result leaves already hard-pressed intelligence analysts to find that critical piece of information that could identify an emerging threat – the figurative “needle in the haystack” of intelligence data. While not every haystack has a needle, intelligence analysts must be able to evaluate each haystack of information to be sure they have not missed what could be a lucrative discovery

“IF YOU HAVE MORE HAYSTACKS, YOU DO NOT NECESSARILY GET MORE NEEDLES.” - MARK LOWENTHAL, FORMER ASSISTANT DIRECTOR OF CENTRAL INTELLIGENCE FOR ANALYSIS AND PRODUCTION

… And A Symphony of Noise

The majority of raw data in most collection and analysis scenarios is often considered “noise,” as in the analogy of signal-to-noise ratio for radio communications. Finding the “signal” or critical piece of data hidden in the abundance of background noise (irrelevant data) has always presented an analytic challenge.
The analyst must devote great care to finding the signal within the noise that dominates many large, potentially lucrative data sets. A unique element of information, or a subtle trend in a series of data points within a data set, may potentially lead to a discovery of significance. To distinguish valuable data from irrelevant or deliberately misleading information, analysts must employ new capabilities to discover and bring essential data to the foreground. Data must be conditioned for the rapid extraction of valuable signals and visualized in a way that

can influence an analyst's interpretation of its value. Once discovered, time-sensitive information must be then communicated to the commander quickly and with enough lead time to allow the commander to take appropriate action. The past process that took hours to deliver current intelligence reporting, normally in the form of a static presentation to leadership, no longer meets this need.

An Insight Engine

'Breaking News' experimented with an evolution in information search technology known as an insight engine.

An insight engine augments existing search technology with artificial intelligence (AI) to deliver insights from the full range of data available to an analyst. Insight engines differ from search engines in that they offer the capacity to understand the meaning of that data rather than just capturing content that matches a query. Insight engines offer the additional advantages of enabling users to combine information from different sources, show correlations in their data, and visually represent that data to highlight trends and anomalies.

Some insight engines use machine learning (ML) to extract and bundle specific elements of information from large volumes of complex structured and unstructured data. These engines can learn to categorize information and provide a comprehensive picture to individual analysts based on their prior interests. They can then proactively search databases to deliver information when needed rather than requiring analysts to initiate a search as part of their daily routine. Embedded Natural Language Processing (NLP) technology enables insight engines to understand unstructured data (e.g., text, video, imagery) and include that information in the comprehensive picture of the data provided to the analyst.

An insight engine also can employ what is known as a semantically enhanced logical data warehouse. This software stores semantic information that enables the insight engine to identify synonyms associated with an element of the search query and provide a more comprehensive set of results based on synonymous references to the data of interest.

A CORE ASSUMPTION UNDERLYING THIS STUDY IS THAT INTERACTIVE REPRESENTATIONS OF DATA CAN AMPLIFY A HUMAN'S NATURAL ABILITY TO DETECT PATTERNS, ESTABLISH LINKS, AND DRAW INFERENCES FROM THAT DATA

The advantages of employing insight engines within the sometimes-overwhelming data environment confronted by intelligence analysts, operating in a time-constrained situation, could be especially significant.

For its experiment with an insight engine, the 'Breaking News' project team used an unclassified, previously vetted data set to enable the broadest sharing of the effort's results. Data used in a previous Department of Defense (DoD) illicit arms sales exercise scenario offered the team's experiment a large open-source intelligence (OSINT) data set fused from a variety of different data sources. This approach allowed MITRE to experiment with unclassified sources, yet still openly evaluate the performance of the insight engine against a basic goal of 'can an insight engine save time by allowing an analyst to recognize the value of specific data elements in a larger data set, faster?"

Understanding the Noise

The "Breaking News" team first sought to understand the existing or natural data structures that help an intelligence analyst organize data, and then develop a line of questioning against that data structure to recognize important information. The central question the team posed was “How does that human analyst understand what is of value within the constant noise, and how can a machine-based process support the natural way humans query their data structure?”

To begin a human-machine partnership, the study team began with a process to structure analytic questions in a human way and recognize gaps that a machine-based process can help answer. One such approach to drawing out those gaps is to apply a “building block” structure to the analytic questions and the operational concerns they seek to answer.

The scenario used for this effort was broken down into data building blocks that capture examples of both basic and complex analytic questions. Addressing each building block facilitated a more robust understanding of the overall threat while enabling recognition of new information and additional intelligence gaps that may warrant further attention as the analyst seeks to answer an operational need.

The ‘Breaking News’ team sought to employ this building block approach to see where specific automated display techniques can be used to answer unknown information requirements, filling previously empty building blocks of understanding.

INSIGHT ENGINES DIFFER FROM SEARCH ENGINES IN THAT THEY OFFER THE CAPACITY TO UNDERSTAND THE MEANING OF DATA RATHER THAN JUST CAPTURING CONTENT THAT MATCHES A QUERY. THEY ENABLE USERS TO COMBINE INFORMATION FROM DIFFERENT SOURCES, SHOW CORRELATIONS IN DATA, AND VISUALLY REPRESENT THAT DATA TO HIGHLIGHT TRENDS AND ANOMALIES

Experimenting with Elasticsearch

The ‘Breaking News’ team used a data technology known as Elasticsearch to demonstrate this capacity. Elasticsearch is an emerging competitor within the new market of insight engines. The team chose the Elasticsearch insight engine based on its ability to rapidly query unstructured text and amplify human insight with a range of visualization capabilities. Also unique was its use of nontraditional data storage, offering increased flexibility over traditional or relational databases. This tool has the advantage of providing data schemas that are dynamic and that forego the time-consuming and careful process of defining relational schemas.

Elasticsearch can display data in a wide variety of ways and allows analysts to exploit both ontological (relationships between concepts and categories) and taxonomical (a general classification system) properties of the data.

In its experiment with Elasticsearch, the ‘Breaking News’ team sought to perform a basic evaluation of the insight engine’s ability to quickly draw an analyst’s attention to key trends or specific elements of high-value data from within the test data set. This would provide some indication if this technology could enhance insight into the daily operational questions necessary to fulfill the intelligence analytic mission.

The Elasticsearch insight engine demonstrated it could help a user gain insight relevant to their information needs in various ways, and through a variety of data visualization techniques. The team posed a series of queries against the test data set using the Elasticsearch engine to see if the engine could discover and parse that data into our building block data structure, and then expose that data to analysts via a visualization option that would quickly highlight/draw their attention to areas where the data may reflect a trend or specific item of interest.

The visualization component of the Elasticsearch capability was paramount in achieving the full value of the insight engine. The placement of every data element, to include text on the display, has some significance in how the needed information was viewed by the analyst, in comparison to associated data or unrelated noise from the data set. With so much of a human's cognition associated with sight, how data is visualized is paramount to understanding new information.

One example of analytic interest that the team used within this experiment focused on air defense systems. The initial query of arms traded within the test data set resulted in a word cloud image (Figure 1) that displayed the weight of different major weapons systems that were referenced in the data set. Word clouds such as this example enable analysts to deduce within seconds which elements of interest are most relevant among tens of thousands of records collected within this arms transfer-based data set.

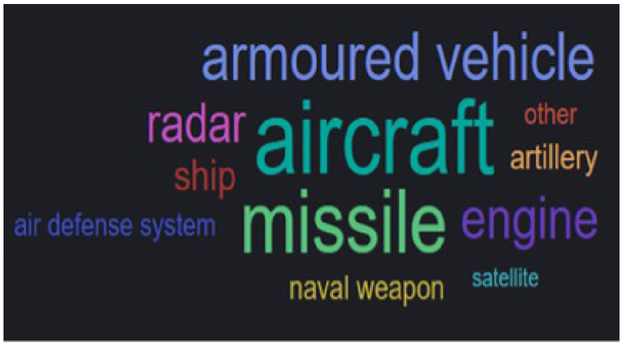


Figure 1 - Gaining insight via word cloud visualizations

Additional visualizations added country of origin, purchasing country or group, location, and other data elements to quickly begin to refine the picture gained from the data set. Combined with the analyst's human expertise, they were quickly able to tell the full story regarding the trade in air defense systems during the period of reporting covered by the team's test data set.

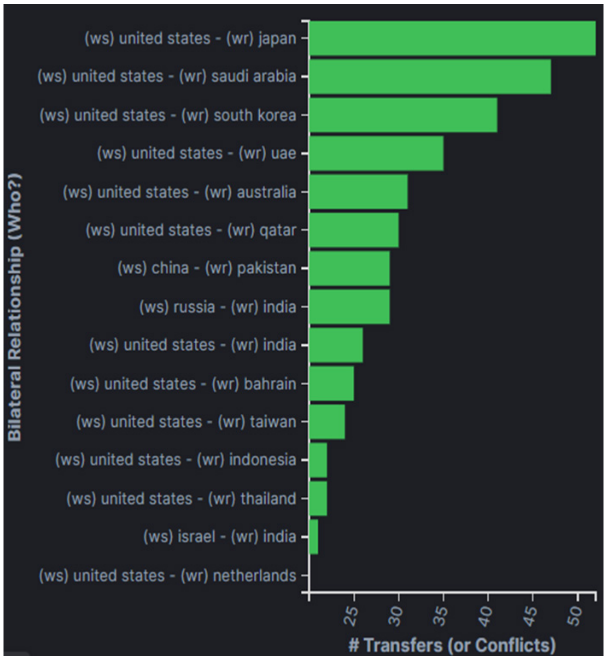


Figure 2 - Arms supplier and recipient relationship quantified by number of transfers

Each display and the combination of displays used are selected by the analyst. Repeated interaction with the insight engine and data set enables analysts to better appraise, estimate, and judge adversarial behavior by coupling expert insight with data.

Simplifying Complexity

Complexity grows as analysts attempt to satisfy multiple filtering conditions at one time The insight engine-facilitated analysis of a Russian air defense system, known as the S-400, is a good example of human-machine teaming.

An insight engine that constantly looks for related data across a wide range of information feeds and then displays the assembled data in an easy-to-review way, would undoubtedly assist the watch stander to rapidly recognize its importance among a constant stream of intelligence reporting.

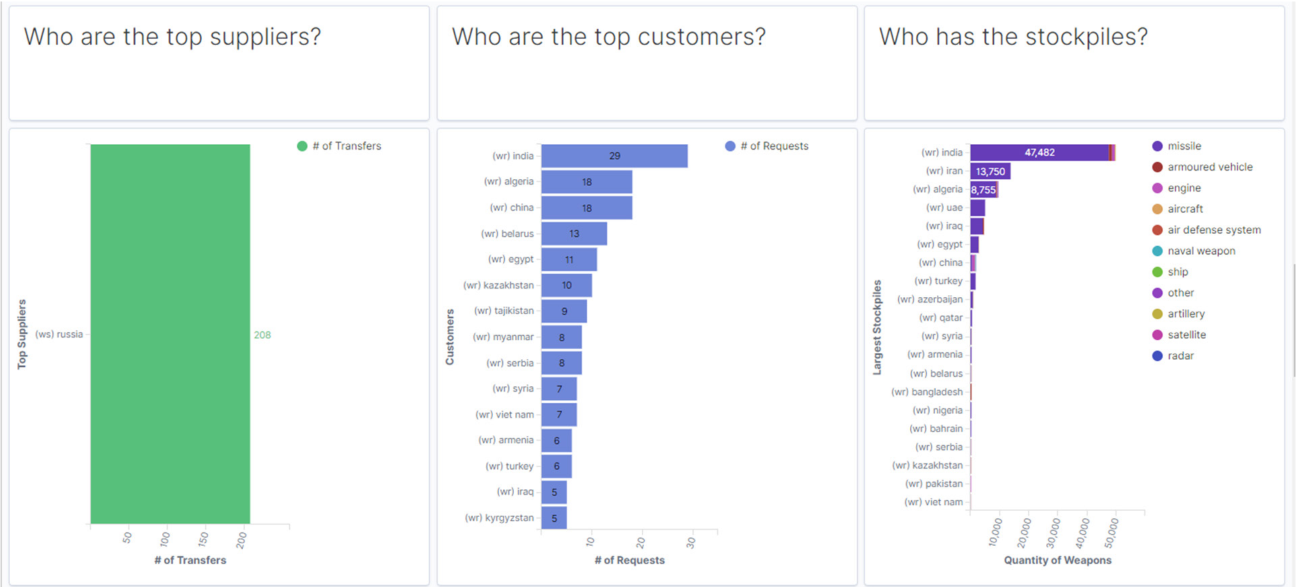
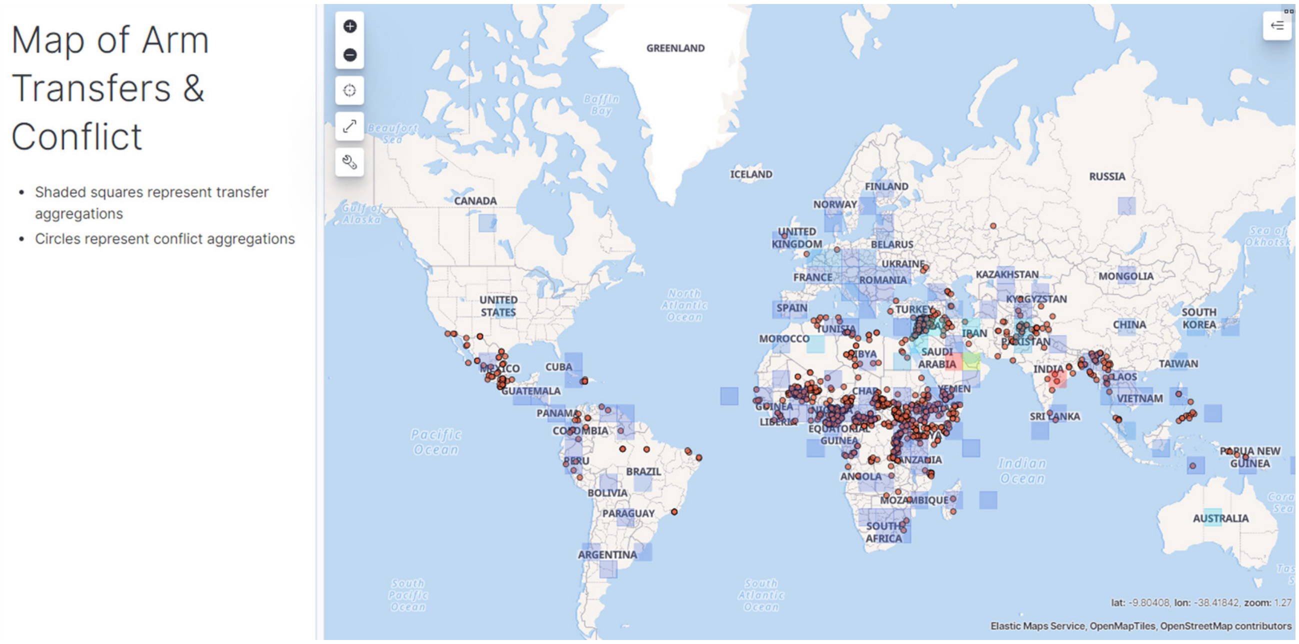


Figure 3 - Recipient and quantity of arms transfers supplied by Russia



Map of Arm Transfers & Conflict

- Shaded squares represent transfer aggregations
- Circles represent conflict aggregations

Figure 4 - Elasticsearch map display of arms trading events overlain on current zones of conflict

A query of S-400 data returned, within seconds, a coherent display of critical weapon intelligence found among 33,552 records and six data sources, including sources that had been conditioned by human language technology algorithms. The resulting display narrowed tens of thousands of records down to 77 and corroborated the desired weapon insight across two data sources.

Insight engine data can be structured to store and then intuitively indicate the most salient attributes for an analyst, via a display specifically designed by the analyst to expose data in a way that the analyst is most comfortable seeing. While the project team tailored its test display to the illicit arms transfer scenario, its underlying structure is purposely generalized enough to hold vastly different scenario data. If new or updated datasets enter the pipeline, they can be mapped to this display structure.

Visual display structuring requires a human to tell a machine how to map the data and then calls for a machine to automatically place the data into appropriate fields. In the case of the 'Breaking News' scenario, the task of mapping and then populating a display data index inside a search engine was automated via Python Logic – a rules-based approach to programming that sees computation as automatic reasoning over a database of knowledge made up of facts. Applying specific programming rules to the facts in the data set allowed us a rapid way of populating the data index, saving significant time.

Fueling an Insight Engine

It is important to note that as with any newly introduced technology, there will be both start up and sustainment costs in human capital and funding. Unforeseen and unprogrammed startup and sustainment costs have been repeated culprits in an organization's failure to embrace potentially helpful technologies. Not the least among these are the human costs to already limited analytic resources.

A significant challenge to the success of employing automated tools in support of analysis is conditioning the raw intelligence data necessary to feed that tool, such as in the creation of a data schema. Watch analysts will likely be unable on their own to configure the data resources to feed an insight engine when this technology is first introduced into their organization

AN ORGANIZATIONAL OR ENTERPRISE-LEVEL DECISION TO INVEST IN THE APPLICATION OF INSIGHT ENGINE TECHNOLOGIES WOULD NOT ONLY SAVE ANALYST HOURS BUT ALSO OFFER THE PROMISE OF REDUCING INSTANCES IN WHICH ANALYSTS MISSED CRITICAL INDICATORS BECAUSE OF DATA OVERLOAD

Standing up a technology-enabled data pipeline requires effort to access data feeds, create data conditioning algorithms, enable data access by the insight engine, build standing data queries to interrogate the data at specific intervals, develop the range of data visualizations and displays preferred by the analyst, and of course to interact with the data to increase analytic understanding. This represents no trivial effort and comes at a cost beyond the means of some watch centers.

Few watch centers will have the organic data science or IT support required to help establish an insight engine tailored to their specific interest. However, an initial investment, as well as sustainment considerations, may offer sufficient long-term cost savings in analyst time, to make the initial outlay worthwhile.

An organizational or enterprise-level decision to invest in the application of insight engine technologies to improve the current intelligence capabilities of each Combatant Command (CCMD) would not only save analyst hours but also offer the promise of reducing instances in which analysts missed critical indicators because of data overload.

A basic architecture that can be modified for specific intelligence needs at each CCMD's watch center could offer savings across the enterprise, by limiting the effort and investment at each location to only what a CCMD requires to tailor insight engine requirements and queries to specific intelligence requirements.

Data technologies as well as the data itself in an organization should be viewed as 'living and breathing' commodities that incur both start-up and on-going sustainment costs. As AI/ML applications increase in military workflows, underneath-the-hood access (ability to review the underlying code) that enables human checks and balances to the data, the algorithms, and the technology, is critical. While organizations have always planned for operations and maintenance costs associated with hardware, for analytic leadership the need to plan for sustainment of the data assets or the algorithms themselves may be unanticipated costs.

Initial Results

The initial 'Breaking News' experiment demonstrated that an insight engine could help an analyst to discover information more rapidly from a large data set. While this result on a small-scale application was promising, additional effort will be required to fully evaluate its potential. A test, at scale, in an operational intelligence environment will be required to fully evaluate its worth as an element of an enterprise level solution.

In a future effort, MITRE will seek to partner with a CCMD to create an operational test bed to evaluate the use of an insight engine, at scale, tailored to ingest the core data feeds in use by an analytic element within an intelligence watch setting, applying the necessary data conditioning algorithms to capture the information required by that watch team. A test at this level will allow the effort to fully characterize the initial investments necessary to bring this approach to bear in an operational setting, as well as quantifying any resulting time savings from the effort as compared to the past analytic workflow/tradecraft used by that team.

New Ways of Seeing and Sharing

As analysts take advantage of new tools that enable them to discover information faster, they will also need to share that information faster. The speed at which analysts share emerging intelligence is as critical as the speed of data discovery. However, in today's environment, the time necessary to create and then share an intelligence product with the intended consumer often decreases the impact of emerging intelligence. This challenge is nowhere more apparent than in the current intelligence process, and the intelligence watch structures supporting its production.

The 24-hour watch cycle and commander's daily intelligence update are a ubiquitous tradition at all levels of command both within and beyond the DoD. Aside from the inherent delays in information sharing created by the process's timeline, the drawback of a 24-hour production cycle also includes the time necessary to create and review the presentation, and then the static nature of the product, often leaving the commander with information that is already hours old.

While the initial "Breaking News" effort focused heavily on applying technology to assist in data curation to improve intelligence production speed and fidelity, rapidly sharing information is the next critical leg of the race to ensure decision makers are armed with the best possible information to maintain decision advantage over their adversaries.

The Human Consumer

Within DoD, all operational decisions are based to some extent on an intelligence assessment of the threat. However, even in the best possible scenario in which intelligence has been discovered and shared in time to enable a commander to make a timely decision, some time lag created by the human decision process will remain.

The "human factor" in this decision-making process is an important aspect in considering how intelligence can be most effectively shared and understood. Various factors, including training, cognitive bias, experience, etc. affect and influence the human decision-making process. In wartime situations, an added factor is the limitation of time itself.

The less time an individual has to decide, the greater the influence of time on the decision process. Limiting the time for thoughtful review in crisis often forces decision makers to rely largely on their existing perception of the threat and the training they possess for that situation, often eliciting the 'muscle memory' response provided by their past training.

If we take these factors into consideration and apply them to the operational environment, we see why it is essential to deliver intelligence assessments, at the highest level of fidelity, in a way that allows for rapid, straightforward consumption and assimilation of information. The objective in visualizing intelligence should be to harness human sense-making abilities while limiting the impact of human biases on the speed of the decision process.

The Impact of Presentation

The way information is presented can impact the way a consumer perceives the data provided. For example, the same information presented by two different briefers may leave a consumer with two different opinions. It is also true that the same information communicated by two different mediums (graphic versus written), may also leave the intended consumer with a slightly different perception of that data. In this way, each method for sharing intelligence has some limitations. As we seek to modernize how we share intelligence we must look to those methodologies which strengthen the consumers ability to clearly understand the context and gravity of the information being presented.

To achieve this goal, the next generation of capabilities must embrace a practical standard representation for a consumer to easily ingest the information they are being presented, and for the range of consumers relying on that product to perceive the data in the same way.

Future visual representations of modern intelligence data will need to clearly communicate multiple attributes. These visualization designs must be consistent with currently accepted constructs and 'mental models' to increase learnability and utility and decrease user learning curves. They must also seek to balance providing access to all information with the need for simple interfaces that initially direct attention to the most relevant information, which is the enduring challenge of sharing too much versus too little information.

As speed remains a critical factor in sharing current intelligence, we must seek methodologies to address visualization requirements systematically. This could include creating a standard set of visualizations that can be applied to specific situational awareness needs.

The challenge of visualizing intelligence information in this way is not new. MIL-STD-2525 has been the DoD standard for sharing graphic operational information. It defines the way military forces must depict symbology in their operational graphics, including a set of rules for symbol construction and generation to be implemented in C2 systems to promote interoperability; and standard symbols and building blocks, including a frame, icon, modifier, and amplifier using color, graphic, and alphanumeric representations.

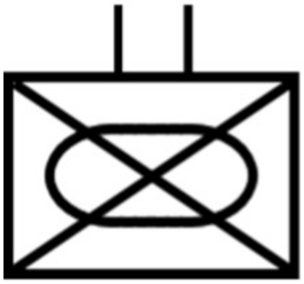


Figure 7 - MIL-STD-2525 generic mechanized infantry battalion symbol

In the context of military operations, the standardization of data remains foundational to how dispersed operational forces must view both friendly and enemy force disposition within a comprehensive picture of the battlefield. Today this is expressed in efforts to improve the ability to portray a common intelligence and operational picture.

The drawback of MIL-STD-2525 today is that it associates each symbol with a single geographic coordinate. Intelligence data may have multiple components of location information: source, origination, manifestation, and other information components for transient/mobile threats not easily represented within the current standard.

Based on the doctrinal acceptance of standard symbology, evaluating a symbology-based approach may be a good starting point for considering how to portray time sensitive intelligence information more effectively.

The 'Breaking News' team focused on how existing standard symbology could evolve to address the human factors challenge. The list below identifies the core areas that standard military symbology could evolve to support enhanced intelligence information sharing across a broad operational consumer group:

- 3D variation to symbology: This will help to address specific sizing associations that intelligence analysts can use in functional areas such as targeting and locational concerns.
- Updates to address newer equipment, resources, data: symbology should evolve to support cyber and space equipment; high-value individuals; variation of intelligence source data such as Publicly Available Information (PAI) reporting and OSINT. Such changes will aid analysts to quickly note items of interest and promote leadership's understanding when viewing the dynamic environment.
- Fidelity: Developing a standard component of the symbology that helps to determine the fidelity of the content. How confident is the analyst in the information he has posted as a symbol to the CIP?

The 'Breaking News' team identified these initial findings during the first stage of the human factors assessment, but significant additional work is required to provide both greater detail about the proposed changes and the implementation plan to execute this evolution.

Potential Next Steps

While limited in duration and scope, the initial 'Breaking News' effort demonstrated that advancing technology will offer a variety of solutions to the IC as it seeks to answer the current intelligence challenge and meet operational demands for timeliness. Key next steps as we continue to address this challenge may include:

- Test an insight engine within an operational watch setting with the metrics in place to fully evaluate its ability to speed the discovery and evaluation of information in support of the command's operational intelligence requirements. The test will be essential to evaluate if it can work at scale, and to fully understand the costs of establishing and sustaining this capability, as well as evaluating if this approach will offer utility as a future enterprise level application.
- Continue the effort to develop the data structures or data arrangements required for specific/priority Combatant Command use cases (e.g., problem sets).
- Perform a study to determine the required data normalization and standardization necessary to apply AI/ML technologies to intelligence in support of operational decisions.
- Carry out the next stage of human factors assessment to develop guidance and a reference implementation of how best to present data to the decision-maker for rapid consumption and maximum understanding.

Authors

Joseph Convery is the Chief Engineer for the Command Priorities Department of the MITRE Intelligence Center's Analysis Division. He is a former U.S. Army Intelligence officer, civilian analyst, collector, and leader with over 39 years of experience across the U.S. Intelligence Community. He currently provides oversight of all MITRE activities across the 11 Combatant Command Intelligence Directorates.

Hassan "H" Terry is a MITRE Lead Systems Engineer and Military Intelligence subject matter expert. He is a former U.S. Air Force all-source intelligence analyst and mission planner with more than 27 years of experience in the Intelligence Community. He currently serves as the U.S. European Command, Joint Intelligence Directorate's Project Lead, focusing on resolving intelligence capabilities gaps and designing the next generation of analytic capabilities. Hassan remains closely tied to J2 leadership supporting the Commands' most pressing crisis intelligence requirements.

Genevieve Whiddon is a MITRE Lead Data Analyst and Multi Discipline Systems Engineer with over 20 years of expertise developing new ways to look at data. Her efforts to apply modern data analytics have contributed to improving naval indications and warning and data analytics at the National Geospatial Intelligence Agency, cyber analytics at the Department of Justice, and across numerous federal government programs. Genevieve currently provides data analytics support across a range of projects within the Technical Analysis Department of the Analysis Division in MITRE's Intelligence Center.

William Wang is a Computer Science major at the University of Chicago, contributing his expertise to the Breaking News project during his Internship with MITRE during the summer of 2021. His experience with machine learning contributed greatly to this effort.

Intelligence After Next

MITRE strives to stimulate thought, dialogue, and action for national security leaders developing the plans, policy, and programs to guide the nation. This series of original papers is focused on the issues, policies, capabilities, and concerns of the Intelligence Community's analytical workforce as it prepares for the future. Our intent is to share our unique insights and perspectives surrounding a significant national security concern, a persistent or emerging threat, or to detail the integrated solutions and enabling technologies needed to ensure the success of the IC's analytical community in the post-COVID-19 world.

About MITRE

MITRE's mission-driven teams are dedicated to solving problems for a safer world. Through our public-private partnerships and federally funded R&D centers, we work across government and in partnership with industry to tackle challenges to the safety, stability, and well-being of our nation.

MITRE

AUGUST 2022