2nd Lt. Jesse Underwood, Alpha Company, 1st Battalion, 17th Infantry Regiment, engages enemy forces 19 February 2010 during Operation Moshtarak in Badula Qulp, Afghanistan. The International Security Assistance Force operation was an offensive mission conducted in areas of Afghanistan prevalent in drug trafficking and Taliban insurgency. (Photo by Tech. Sgt. Efren Lopez, U.S. Air Force)

Tactical Data Science

Col. Harry D. Tunnell IV, PhD, U.S. Army, Retired

Shortly after I returned from a combat deployment to Afghanistan in 2010, I commented to an interviewer that "the institution [U.S. Army Training and Doctrine Command] is putting out a good rifleman. What they are not doing is putting out a good rifleman for the digital age."[1] Despite a decade's passage, there appears to have been little progress. It seems that the military education and training system continues to emphasize industrial age information management practices in tactical units. This approach ultimately hampers military operations at the tactical level of war in today's information age.

The lack of progress is unfortunate because special skills and competencies are required for digital transformations.[2] This article proposes a set of skills and competencies for digital transformation at the tactical level of war that can easily be taught by the institutional Army and be sustained with small-unit training. The result will be a data science discipline customized for personnel in tactical units.

The U.S. Army needs a specialized data science discipline to help leaders transform raw data captured by myriad systems in small units and other sources into useful tactical information. A lack of knowledge in small units about digital data (hereinafter data) and the tools to capture, manage, and analyze data locally inhibits a battle staff's ability to gain tactical insights from data. Soldiers need more than mere competency with tactical information systems such as FBCB2 (Force XXI Battle Command, Brigade and Below, a command and control system in command posts [CPs] and combat vehicles). They need to understand the capture, management, and analysis of data, which I call tactical data science.

Using Data in Decision-Making

Data-driven decision-making is the practice of making decisions based more upon the analysis of data than intuition.[3] Today, writing about using data to inform decision-making is almost cliché. Yet the reality is that in the twenty-first century, small-unit leaders still use physical artifacts and static electronic data rather than dynamic or streaming data for tactical planning and decision-making.[4] As a practical matter, this means that leaders rely on artifacts informed by someone else's intuition rather than using data to inform their own coup d'oeil.

In today's era of big data (data that requires high-performance processing and large computational infrastructure, and is characterized by its volume, velocity, and variety at scale), there are tools designed for insight generation from data.[5] What is missing for tactical leaders are soldiers trained to use the tools to exploit today's modern data-rich tactical environment. Unfortunately, even when a battle staff has the right tools (and they

Col. Harry D. Tunnell IV, PhD, U.S. Army, retired, is an information technology director at Eli Lilly and Company and an adjunct lecturer in the Department of Human-Centered Computing at Indiana University–Purdue University Indianapolis. Tunnell is a West Point graduate who commanded 1st Battalion (Airborne), 508th Infantry in Iraq and 5th Stryker Brigade Combat Team, 2nd Infantry Division in Afghanistan. He is a senior member of the Institute of Electrical and Electronics Engineers.

do), soldiers do not have the training to use them well, and leaders do not have the background to ensure that the battle staff is focused on the right data to solve the tactical problem at hand.

Leaders need to ask questions with data in mind. They must frame their questions to the battle staff in ways that can be tested with data. Let us use a hypothetical intelligence briefing (based upon a 5th Stryker Brigade Combat Team, 2nd Infantry Division, CP practice in Afghanistan) as an example. The briefing is conducted using dynamic data stored in databases that can be queried during the briefing:

1. The commander identifies a topic that he or she wants more detail about and begins to ask questions.
2. The commander and staff brainstorm and shape the commander's questions into four hypothesis statements.
3. The intelligence section queries databases and creates visualizations. Enough data is available to test the first two hypothesis statements and decisions are made based upon these data.
4. After seeing the visualizations, the commander decides that the third hypothesis statement is not relevant. It is discarded, and the staff clearly understands that they do not need to follow up on the hypothesis.
5. The commander and staff revise the fourth hypothesis statement. Databases are queried, but acceptable data is not available.
6. The commander and staff craft priority intelligence requirements (PIR) and information requirements (IR) that are designed to gather the right data to test the final hypothesis.

The example highlights that commanders are an integral part of crafting hypothesis statements and making visualization decisions. To participate effectively, commanders must understand data and the tools and processes used to interact with data. The PIR and IR in the example are forms of research questions used to frame future data collection.

Vignettes

In this section, I provide a few real-world examples of generating insights from data to inform tactical decisions. The vignettes are all from combat and are filtered through the lens of my postmilitary work as an information technology professional and informatics educator.

Thinking out loud. During briefings, the practice of "thinking out loud" caused unnecessary work for the brigade staff. During a briefing, I might ask a series of questions about something I thought was interesting. In a nondigital environment, a staff officer would do a quick search of internal references (e.g., staff duty journal entries, situation reports, intelli-

data and modern information systems slowed down processes within the CP. My questions were constructed for an "old school" analysis. I soon realized that I was the delay and had caused the staff to spend days and weeks on data preparation whenever I asked a question because they were using data science techniques to interact with data. I had created many inef-


Capt. Matthew Quiggle, commander of Alpha Troop, 8th Squadron, 1st Cavalry Regiment, 5th Stryker Brigade Combat Team, examines structures in a village from the top of a Stryker 24 December 2009 before dismounting to cordon and search the area in Afghanistan. (Photo by Sgt. Chris Florence, U.S. Army)

gence presentations, doctrinal manuals) and provide feedback. I would then decide if additional investigation was merited. I had seen many commanders use a similar approach throughout my career.

I did not comprehend that digital systems and data required analytical processes that were much different from those for an analysis using traditional physical documents and electronic data. Data scientists spend an extraordinary amount of time preparing data; by some estimates, about 80 percent of their time is spent on such tasks.[6] My lack of understanding about

ficiencies for a staff working hard to respond to their commander, and I needed to correct the situation.

After realizing that I was the bottleneck, I asked the staff to tutor me. I learned about the different information systems, databases, data formats, and data retrieval practices common in the CP. This helped me to anticipate what would be required from a technical perspective to answer my questions. The result was that I asked questions differently, and I began to give the staff parameters for data collection and analysis. This led to our practice of data derived PIRs and IRs.
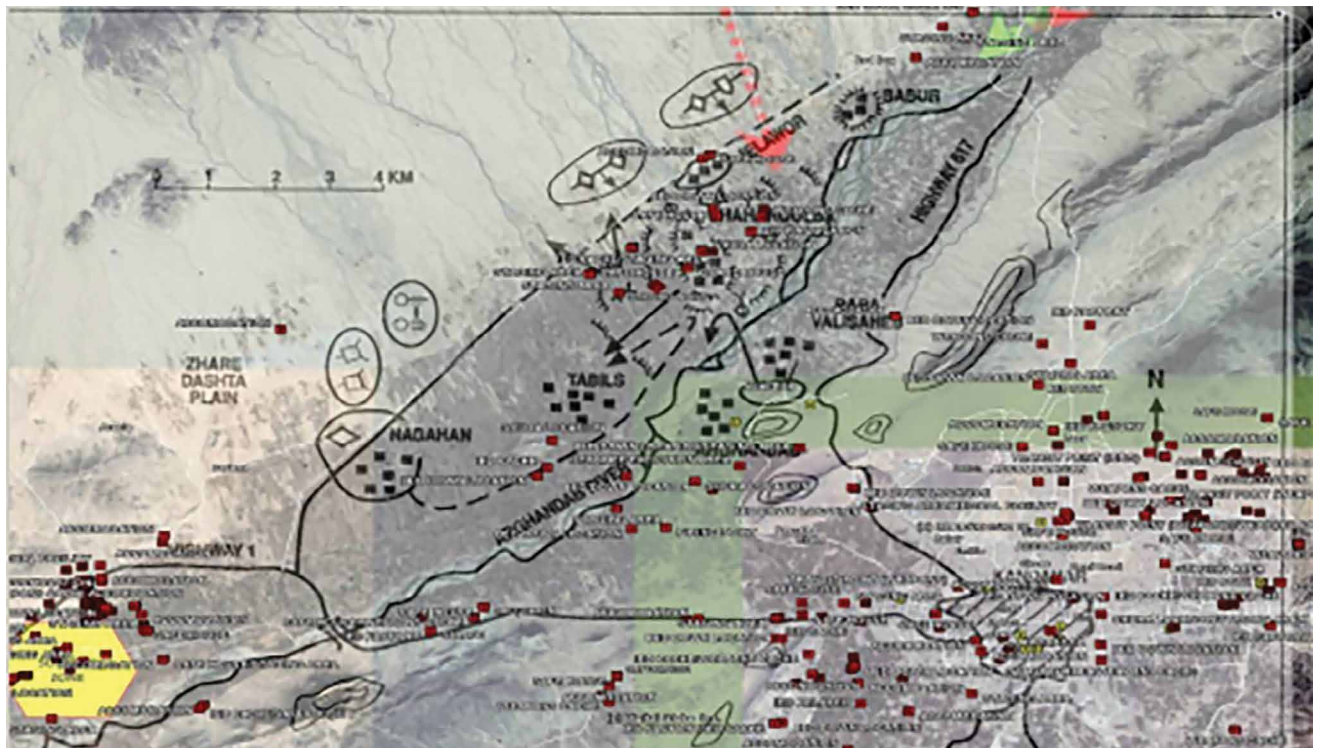
Figure 1. Screen Capture of a Brigade Digital Overlay

Digital overlays. Figure 1 depicts the enemy situation in the brigade's area of operation combined with a historical drawing of a mujahideen defense in the same area during the Soviet-Afghan War (1979–1989). Intelligence analysts came up with the idea to use historical artifacts to explore previously established patterns. The pattern that the drawing in figure 1 describes is that the mujahideen did not conduct military operations in the northeastern quadrant (bounded by Highway 617) because they used it as a living area for their families.[7]

The analyst's approach was brilliant. By incorporating historical drawings, we could watch for indicators of a previously established pattern (or the absence of indicators). These indicators allowed us to apply maneuver resources more efficiently. What figure 1 showed us was that enemy activity was unlikely to manifest in the northeastern quadrant.[8] When our own tactical data supported the pattern (red icons are enemy contact), we began to make decisions based upon the pattern.

We were interested in correlations and not causality. Consequently, we never investigated why the Taliban did not fight in this area. Once the historical pattern manifested, we surmised the area was secure enough for development resources (e.g., U.S. Agency for International Development). This allowed us to shift the main effort in this area to civil-military operations. We then refocused maneuver forces in other areas so that we could continue our hunt for enemy tactical formations and complete their destruction.

A battle captain's initiative. Unmanned aircraft system (UAS) streaming video would typically be displayed on a large plasma screen near other tactical displays at the front of the CP. While this provided excellent observation of a very localized area, it lacked tactical context because the surrounding environment could not be viewed.

Capt. Shaun Young, a battle captain, integrated observation and position location data from the UAS airborne assets in Regional Command–South (RC-S) into one view that was then combined with other tactical information. He integrated Land Warrior (a personal fighting system for infantry soldiers that includes a tactical computer that shows position location

information for those wearing the system) and FBCB2 information. His visualization also included position location information from fighter aircraft that was integrated into our FBCB2 network.[9]

Leaders could now see all UAS locations in RC-S and click an aircraft icon to see the camera feed. Consequently, a leader could see a UAS feed in context with other tactical information. (If a leader decided to observe a specific feed in detail, he or she could still

were able to review patterns set by *any* brigade unit traversing the area.[10] If a commander noticed a pattern or choke point, he or she could simply draw a box around the area on a digital overlay, send it to his or her subordinates, and direct them to stay out of the area for an extended period (e.g., ninety days) to disrupt an emerging pattern for the enemy.

Before implementing honesty traces, Stryker vehicle drivers were approximately 66 percent of the brigade's

> "Before implementing honesty traces, Stryker vehicle drivers were approximately 66 percent of the brigade's killed in action. After honesty traces, no Stryker drivers were killed.

view the large single-feed screen.) The solution was not perfect as there was approximately nineteen seconds of latency; however, this was acceptable because of the benefit of the additional context.

Honesty trace. During our deployment, I read an article in *Stars and Stripes* about a technique that U.S. Marine Corps units in Helmand Province were using to avoid getting blown up by improvised explosive devices. Dismounted patrols tracked their actual patrol routes using commercial handheld GPS devices. When a patrol returned to base, the data from their route was retrieved and posted on a physical map. This was known as an honesty trace. Once the Marines knew the patterns they were setting, they could design patrol routes to frustrate enemy improvised explosive device teams. The method was manual and local to individual small units. I asked the brigade staff to learn more about the approach.

When I followed up with the staff, I was told that they thought they could create digital honesty traces. The operations research and systems analysis (ORSA) personnel in the intelligence section had taken the lead and developed a solution. The procedure that emerged was that the intelligence section took all FBCB2 data in the brigade and created features to show honesty traces for every brigade company-size unit.

The effort resulted in a digital overlay that was updated frequently and was sent out on a routine basis to every unit in the brigade. This would ensure that when one unit was operating in another unit's area, they

killed in action. After honesty traces, no Stryker drivers were killed. While other tactical innovations likely contributed to the decline in casualties, there was a strong correlation with digital honesty traces and the lack of successful attacks on Stryker vehicles.

Predictive intelligence. Once the brigade learned to use FBCB2 data in visualizations, there were other innovations. FBCB2 time series vehicle data was combined with other time stamped data to create interactive visualizations. For example, information about enemy communication and FBCB2 movement data were integrated into a single animation displaying the intersection of friendly movement patterns and suspicious communication patterns.[11]

As the animation of friendly movement occurred, communication activities were displayed. The animation was critically reviewed to assess whether the communicator was in a position (based upon terrain analysis using other tools) to see the column. This assessment resulted in a predictive intelligence product to identify possible Taliban observation posts and plan countermeasures accordingly.

Killer data. Maj. Derek McClain, the brigade intelligence officer, asked to do a brigade data collection concept of operations (CONOP, a type of tactical plan in RC-S that had to be approved before large-scale operations could be conducted). The idea was to conduct tactical operations and focus intelligence assets on the area of an ongoing operation to collect data about the enemy.[12] A detailed digital overlay of the enemy

situation could then be created. The initial data capture was so successful that a second brigade CONOP for data collection was conducted.[13]

The payoff for the brigade came toward the end of the deployment. The Taliban was massing a large force in the Task Force (TF) Buffalo area to attack a Stryker platoon vehicle patrol base. As intelligence

methodology and their method can be used to inform a tactical data science practice.

Data science methods. Data scientists begin with a research question or hypothesis, which leads to finding relevant data. Once they have the data, it is preprocessed (e.g., data cleaning) using a variety of techniques to make the raw data suitable for analysis.

> " Industrial-age thinking assumes that there is a lack of data while information-age thinking is the opposite. The challenge today is not the absence of data, it is the lack of knowledge about how to acquire, manage, and analyze data. "

on the situation began to develop, I ordered a cavalry troop to reinforce TF Buffalo.

As part of the reinforcement, the brigade intelligence section began to forward intelligence products derived from the two CONOPs to TF Buffalo. As TF Buffalo consumed the data, I received a call from the commander, Lt. Col. Jonathan Neumann. He believed that the intelligence was precise enough for an offensive operation and requested permission to attack. I approved the plan, and the task force conducted a preemptive attack on the Taliban. Simultaneously, Maj. Michael Gephart, the brigade fusion chief, provided intelligence products to Australian and Dutch special operations forces operating in the area and coordinated for attacks on the Taliban flanks and rear.

The destruction of the enemy force was complete. Intelligence reporting indicated that the Taliban issued orders to not assemble in groups larger than three to five personnel for fear of renewed attacks.[14] All of this occurred while the brigade was in redeployment operations. Killer data combined with tactically savvy leaders and courageous soldiers ensured that redeployment continued unabated.

Tactical Data Science Framework

The vignettes are examples of using data to gain a maneuver advantage. But before similar successes with data can be efficiently adapted by other organizations to their tactical problems, a repeatable framework is necessary. Data scientists follow a

After preprocessing, the data is explored to understand its usefulness, and data scientists refine their research questions and hypothesis statements, develop ideas about variables, and decide how to transform or combine data to create features.

Machine learning models are built for classification and prediction. This is a resource intensive activity because data must be labeled for model training.[15] Once the model is trained and the analysis is completed, the results are disseminated. If the output is not going to be consumed by data scientists, it is usually shared as reports, summaries, and visualizations.

Tactical data science methods. Traditional data science methods are at the core of tactical data science. These methods are surrounded by a military context so they can be applied to solve tactical problems in combat. For example, data-derived PIRs and IRs feed into the formal tactical planning process. And CP personnel can assist with data labeling using modern tools such as Amazon SageMaker Ground Truth.

During planning, each staff element does its own estimate. This includes retrieving data from known sources and identifying previously unknown but existing sources with potentially relevant data. If the staff estimates find adequate data to answer the PIRs and IRs, a data capture plan is unnecessary. If they do not, then a plan for data capture is created (plans can include activities in the physical or digital space, e.g., capturing prisoners or identifying resources to convert interrogation notes into a specific digital format).

Even with such a framework, common digital skills are required. If one believes that data can lead to important insights, then every soldier and leader in the Army should have the basic skills to support tactical data science. (Not everyone will *be* a data scientist. The skills must be matched to the expected level of training and education for soldiers in a particular role.)

Tactical data science supports network-centric operations. Consequently, a fully implemented tactical data science practice occurs at the brigade level and above because battalion and below units do not have the resources for network-centric operations.[16] Even with this observation, a core set of digital skills is necessary throughout the Army because data is managed at all echelons (e.g., digital photos taken on patrol, squad patrol reports, CP staff journals).

Tactical Data Science Skills in Training and Education

It is unlikely that tactical leaders will have a deep understanding of big data from their military education and training. Developing such knowledge requires specialized skills and years of education. Fortunately, there are transferable skills learned by using small data (small data has characteristics like big data, but datasets are small enough to be held in memory on a local machine). Learning how to use small data can be trained in the current military education and training system.

Education is used to develop critical thinking skills, which can be advanced using the military university system. Training focuses on repetitive tasks. Tasks, conditions, and standards should be written for tactical data science activities and be evaluated during field training. Education and training will ensure that soldiers understand data management and can perform duties such as data preprocessing.

Having a battle staff that can perform exploratory analysis with small datasets is important. It lowers the burden on what will be a small tactical data science team, and it lessens the need to request intelligence products from another headquarters. With the right skills, battle staff personnel will be

Table 1. Summary of Tactical Data Science Skills That Should Be Taught to Enlisted Personnel

Initial entry	Basic battlefield data management (images, storyboards, etc.). Static data such as patrol reports and presentations should be managed with metadata. Labeling data for machine learning.
Battle staff noncommissioned officer and similar courses	Manage data in command posts (CP). Basic knowledge of managing small information systems (e.g., platoon or company CP). Labeling data for machine learning.
Advanced noncommissioned officer courses	Advanced data management skills in CPs. Supervising data preparation teams. Competence with managing intermediate information systems (e.g., battalion CP or brigade or higher staff section).
Sergeants Major Academy	Manage brigade and higher CPs as complete information systems.

(Table by author)

able to explore the data themselves; they will be able to answer many PIR and IR locally.

Training for Tactical Data Science

Industrial-age thinking assumes that there is a lack of data while information-age thinking is the opposite.[17] The challenge today is not the absence of data, it is the lack of knowledge about how to acquire, manage, and analyze data. This is a reason that core skills at all echelons are important. Table 1 and table 2 (on page 130) depict tactical data science skills that should be taught to enlisted soldiers and officers, respectively, at each level of professional military education.[18]

In addition to programs devoted to new skills, there are opportunities to modify currently existing

skills. For example, ORSA personnel have skills that can transfer to a data scientist role. In fact, ORSA personnel often serve as proxy data scientists. Specialists in simulations may also have crossover skills. Giving ORSA personnel and other specialists opportunities to transition to a data scientist role can quickly advance a tactical data science capability.

Tactical Data Science Example

The following example demonstrates the difference between how commanders consume information today and how they could possibly use it within a tactical data science practice. The scenario is that a brigade commander is reviewing reports of what has happened in his or her area of responsibility. Currently, such reports are typically captured on Department of the Army (DA) Form 1594, "Daily Staff Journal or Duty Officer's Log," by the brigade battle captain and members of his or her shift.[19] The same general version of the form has been in use since the early 1960s.

The current way to manage shift data. During a shift, several DA Forms 1594 are completed. The forms are used to record the date and time of an incident, a description of an incident and the action taken in response, and the initials of the person making the journal entry. Completed forms are often placed in a three-ring binder so that they can be reviewed by the commander at his or her workstation. Additionally, each staff section of a brigade battle staff keeps records using DA Form

1594. These records are maintained by an individual staff section and are not combined with the records from the battle captain's shift. Finally, subordinate units down to company level manage information using DA Form 1594. This methodology is common.

In the current practice, information for the commander is often in a single physical location and in a format that precludes widespread dissemination or integration with other data. Even when electronic versions of the form are used, the entries are typed and stored in a shared folder on a drive, or the form is

Table 2. Summary of Tactical Data Science Skills That Should Be Taught to Officers

Precommissioning sources	A basic understanding of statistical analysis and data science tools. Competence with basic programming skills.
Basic course	Basic battlefield data management (images, storyboards, etc.). Static data such as patrol reports and presentations should be managed with metadata. Supervising data preparation teams.
Career course	Competence in advanced search and retrieval techniques. Ability to mine metadata using statistical analysis and data science tools. Competence with geographic information systems (GIS).
Combined Arms Services Staff School	This school should be revived. It should teach officers to manage staff sections and command posts as information systems. Graduates of this course should be able to perform data preprocessing, conduct tactical analysis with business analytics tools, and use GIS. Basic programming skills refresher. Students should be introduced to research questions and hypotheses.
Command and General Staff College	Ability to build bespoke information systems. Programming skills refresher. Competence with project management tools (consider tactical planning a project). Skilled use of GIS and the integration of GIS with other systems.
War College	Advanced education about research questions and hypothesis statements. Application of the Data-Information-Knowledge-Wisdom hierarchy and network-centric warfare theory.

(Table by author)

LINE_2_REPORTING_UNIT	LINE_4_ACTIVITY	LINE_4a_ACTIVITY_TYPE	LINE_5a_LOCATION_DETAIL	LINE_6_DETECTED_UNIT	LINE_7_DATE	LINE_7a_TIME	LINE_9_KIA	LINE_10_WIA	INITIAL
4-23 IN	Private Citizens & Property	Bombing/Explosion	Nahri Saraj district	Taliban	12/30/2010	23:28	10	Unknown	AF
BSTB/5/2 SBCT	Transportation	Bombing/Explosion	Gereshk	Taliban	12/30/2010	02:45	14	4	BE
1-17 IN	Government (General)	Armed Assault	Taloqan	Taliban	12/19/2010	09:00	0	1	FD
A/1-9 IN	Business	Unknown	Mazari Sharif district	Taliban	12/18/2010	13:41	1	0	EE
3-17 FA	Police	Bombing/Explosion	Kandahar	Taliban	12/15/2010	14:26	3	9	FC
C/1-9 IN	Police	Hostage Taking (Kidnapping)		Taliban	12/13/2010	22:41	0	0	DC
HHC/1-508 IN	Government (General)	Hostage Taking (Kidnapping)	Qochin	Taliban	12/8/2010	04:40	4	0	DD
BSTB/5/2 SBCT	Government (General)	Hostage Taking (Kidnapping)	Shibirghan district	Taliban	12/7/2010	15:15	0	0	DD
B/1-9 IN	Government (General)	Hostage Taking (Kidnapping)	Darkha	Taliban	11/29/2010	18:43	1	0	BE
2-1 IN	Business	Armed Assault	Tezin	Taliban	11/29/2010	00:27	0	1	AC
3-17 FA	Educational Institution	Bombing/Explosion	Lashkar Gah	Taliban	11/29/2010	05:24	0	5	AC
3-17 FA	Government (General)	Assassination	Puli Alam	Taliban	11/28/2010	22:18	3	1	EE
2-1 IN	Business	Bombing/Explosion	Jalalabad	Taliban	11/26/2010	16:33	1	0	FC
1-17 IN	Transportation	Armed Assault	Bala Baluk district	Taliban	11/23/2010	05:55	0	4	DF
BSTB/5/2 SBCT	Private Citizens & Property	Bombing/Explosion	Zurmat district	Taliban	11/22/2010	01:17	4	2	BF
HHC/1-508 IN	Police	Assassination	Musa Khel	Taliban	11/21/2010	00:10	1	2	BD
HHC/1-508 IN	Government (General)	Hostage Taking (Kidnapping)	Borka	Taliban	11/20/2010	14:01	3	3	FD
3-17 FA	Religious Figures/Institutions	Facility/Infrastructure Attack	Gizab district	Taliban	11/19/2010	08:25	0	0	ED
C/1-9 IN	Airports and Aircraft	Unknown	Jalalabad	Taliban	11/13/2010	09:53	0	0	ED
HHC/1-508 IN	Business	Hostage Taking (Kidnapping)	Khaki Safed district	Taliban	11/13/2010	12:55	0	0	FC
HHC/1-508 IN	Government (General)	Bombing/Explosion	Shwak district	Taliban	11/9/2010	03:36	1	0	AC
4-23 IN	Government (General)	Bombing/Explosion	Sayid Karam district	Taliban	11/3/2010	10:45	1	1	AF
3-17 FA	Police	Hostage Taking (Kidnapping)	Pashad	Taliban	11/3/2010	07:17	2	0	ED
3-17 FA	Military	Armed Assault	Barmal district	Taliban	10/30/2010	21:56	80	5	CE
A/1-9 IN	Government (General)	Assassination	Farah	Taliban	10/26/2010	09:32	3	2	CD
1-17 IN	Government (Diplomatic)	Bombing/Explosion	Herat	Taliban	10/23/2010	10:36	4	2	DF
B/1-9 IN	Government (General)	Assassination	Dur Baba district	Taliban	10/22/2010	06:30	3	2	BE
2-1 IN	Educational Institution	Bombing/Explosion	Puli Alam	Taliban	10/21/2010	20:02	0	1	AD
3-17 FA	Business	Armed Assault	Gereshk	Taliban	10/17/2010	17:55	8	3	EC
1-17 IN	Business	Hostage Taking (Kidnapping)	Anar Dara	Taliban	10/17/2010	08:06	0	0	BE
C/1-9 IN	Business	Bombing/Explosion	Sari Pul district	Taliban	10/14/2010	00:55	0	0	BE
1-17 IN	Airports and Aircraft	Bombing/Explosion	Shindand	Taliban	10/12/2010	12:01	0	0	BF
HHC/1-508 IN	Government (General)	Bombing/Explosion	Taloqan	Taliban	10/8/2010	00:11	13	13	CC
3-17 FA	Government (General)	Armed Assault	Kandahar	Taliban	10/4/2010	03:54	1	0	FE
BSTB/5/2 SBCT	Government (General)	Armed Assault	Kandahar	Taliban	10/4/2010	05:17	1	0	DD
3-17 FA	Private Citizens & Property	Bombing/Explosion	Andar	Taliban	10/4/2010	13:32	2	15	BC

(Figure by author)

Figure 2. The Department of the Army Form 1594 Prototype

printed and placed in a three-ring binder. Regardless of what happens to the form, this is static data that lacks context. The commander receives information that is fragmented because of these data silos. He or she only sees what the battle captain's shift has recorded and placed at his or her workstation.

A modern way to manage shift data. In a digital environment, the organization, storage, and use of data should be much different. Rather than using physical or electronic forms, the battle staff could enter data into a relational database management system (RDBMS). To support this, standardized reporting formats should be updated to enhance management and search of data in an RDBMS. A platform that takes advantage of machine learning could then be used to interact with the data.

Using an RDBMS provides structured data that is more useful for analysis with modern tools. Furthermore, data is available from all staff sections and subordinate units since they can use the same RDBMS. (And vice versa, the brigade battle captain shift data is available to others.) Finally, if the commander, after an initial personal exploration, decides

that he or she wants a deeper analysis, then he or she can task the tactical data science team.

To highlight how such a process should work, a prototype RDBMS DA Form 1594 (DA 1594 Prototype) was created. The prototype is a combination of selected attributes from the DA Form 1594 and U.S. Army Spot Report format as well as attributes for additional tactical context.[20] The prototype is a Microsoft Access RDBMS, which is commonly available in CPs due to the dominance of the Microsoft Office suite of products for office productivity tasks throughout the Army.

The data for the prototype represents reports in a fictional brigade CP. The data was created for illustrative purposes and *does not represent the real activities of any Army unit*. The dataset combined fictional headings, fictional reporting by units, and fictional attack times with 551 records of real terrorist attacks from the Global Terrorism Database.[21] The DA 1594 Prototype (see figure 2, page 131) demonstrates how a modern staff tool should look.

An advantage of using an RDBMS rather than a three-ring binder is that leaders can interact with the

A soldier attached to 2nd Battalion, 1st Infantry Regiment, 5th Stryker Brigade Combat Team, 2nd Infantry Division, uses a Handheld Interagency Identification Detection Equipment (HIIDE) device 16 January 2010 during a cooperative medical engagement in the Maiwand District, Hutal, Afghanistan. The HIIDE system is a multimodal biometric system that collects and compares fingerprints, iris images, and facial photos against an internally downloaded biometric watch list. (Photo by Staff Sgt. Dayton Mitchell, U.S. Air Force)

data differently. For example, one can apply filters. This allows users to explore data in ways that are impossible to do with printed forms in three-ring binders or forms stored as static electronic data.

Another advantage of an RDBMS is that the data is easily accessed with business intelligence (BI) tools. The advantage of BI tools is that they are designed for laypeople and often include an embedded machine learning capability, which is a powerful technology for gaining insights from data. This approach is novel for small-unit CP data, which is typically not analyzed at all. Microsoft Power BI Desktop is the BI tool used for this example.

There are many options when using BI tools. For example, one can visualize data as geospatial data. By using the map functionality, patterns about the proximity of attacks to one another are discernible. Making such connections from individual staff duty journal pages in a three-ring binder or forms stored as static electronic data is impossible for most people.

Using a BI tool to visualize position location data can be done in seconds, while transcribing the same data

from a written form to a physical or digital map takes much longer and is rife with opportunities for transcription errors. Another option with a BI tool is the reports functionality. This is useful for summarizing data in charts, graphs, and other formats. Figure 3 (on page 133) is an example of the different reports that can be created from data stored in the DA 1594 Prototype.

Making sense of tactical data. What is noteworthy about the example to this point is that any commander or member of a battle staff would have the skills to interact with the data. Even more noteworthy is that this level of analysis could be done in a few minutes. (Trying to create such ad hoc reports using
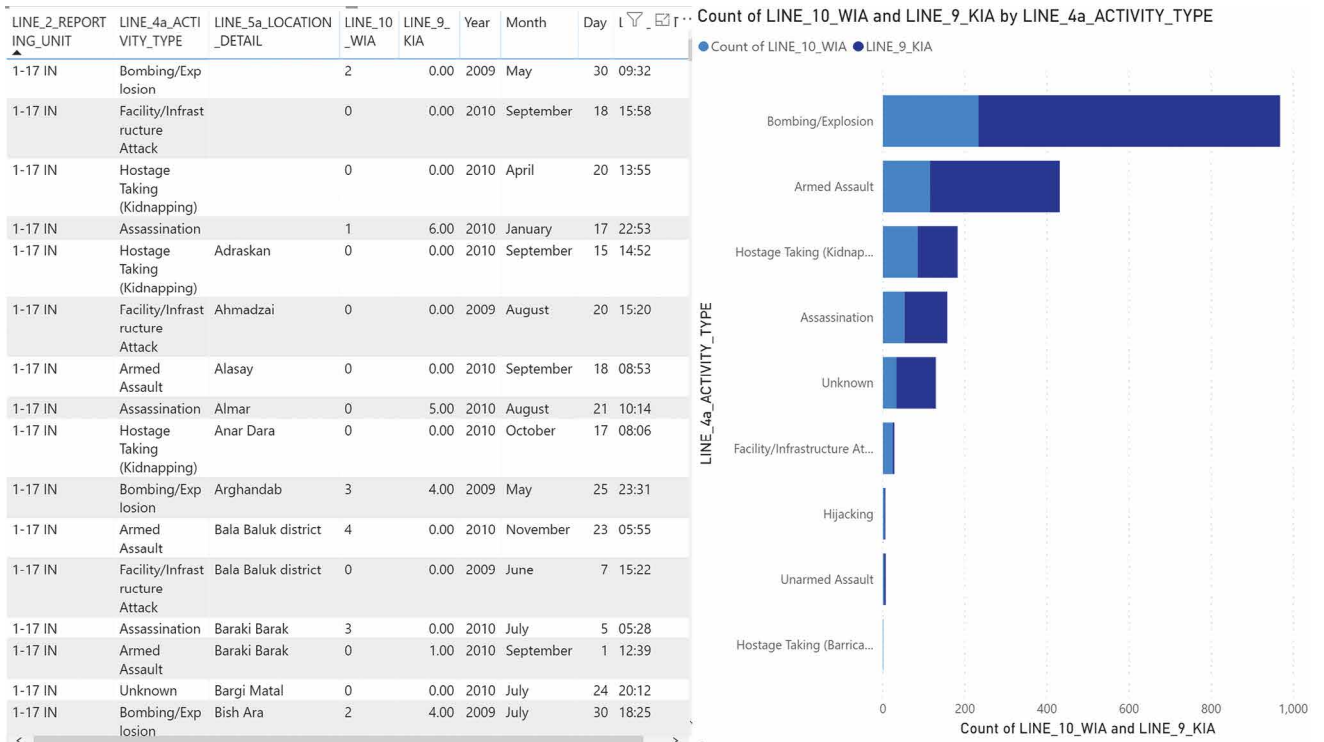
(Figure by author)

Figure 3. Patterns about Reporting Units Shown in a Table (*left*) and Killed and Wounded by Type of Attack Shown in a Bar Chart (*right*) Using Microsoft Power BI Desktop

traditional resources and methods takes hours or days.) Furthermore, another advantage of modern tools is that the analysis can be set up as standard reports and the data will be refreshed as the data changes.

Handoff to the tactical data science team. The bar chart in figure 3 shows that the bombing/explosion attack is the most successful enemy operation. Consequently, it is a good topic for deeper analysis. Once the commander has identified this, he engages the tactical data science team. The team uses different tools with more capability (that also require more education and training). Even though they may start with the same dataset, the tactical data science team uses it differently and for a different purpose. The research question for the tactical data science team is, "What characteristics of the bombing/explosion attack can be analyzed for potential countermeasures?"

This type of research question is salient because the data available from a battle captain shift is inherently limited due to the nature of recordkeeping in a CP. Consequently, even with the enhancement of an RDBMS, there are limitations to such data. Rather, the dataset is used by tactical data scientists to identify ideas for new data sources to mine for different insights.

IBM SPSS Statistics is a statistical analysis tool that many ORSA personnel use, and it is the application used for this part of the example. To focus the analysis, the success of the bombing/explosion attack was redefined as the total number of casualties rather than the attributes of count by type (killed in action and wounded in action) used in the DA 1594 Prototype. The SPSS transform function was used to combine the *LINE9_KIA* and *LINE10_WIA* variables and create the *Casualties* feature for the analysis.

The descriptive statistics of *Casualties* in table 3 (on page 134) provide interesting insights. First, there are three cases of missing data, which means that 98.72

percent of the reports include casualty data. One can infer from this statistic that units are trained on the reporting format, and they are following the reporting procedure correctly. Second, the mean (M = 8.21) and standard deviation (SD = 16.47) indicate quite a bit of dispersion around the mean. This can be an indicator of a lack of consistency or outliers in the data. This could imply a need to validate any conclusions from the analysis with multiple techniques. Third, the median (Mdn = 3) indicates that 50 percent of attacks result in three casualties or less. This suggests that there might be opportunities to isolate attack results with a low number of casualties and learn from them.

Visualizing the data as a histogram (as shown in figure 4, page 135) can help one understand the mean and median results. For example, not only does the median indicate that 50 percent of the attacks result in three casualties or less, but most of them also do not result in any casualties. If the tactical data science team can identify patterns for no-casualty attacks, they can share them throughout the force, and units can update their tactics.

The histogram can also be used to home in on high-casualty attacks. In this scenario, the commander decided that thirty casualties in a single incident is a high-casualty attack because the loss of a platoon can make a company ineffective. (A platoon is approximately thirty to forty people, and a company is approximately 80 to 170 people.) If the data science team identifies patterns for attacks with thirty or more casualties, it could be possible to design countermeasures to reduce enemy opportunities to attack.

Finally, by conducting a crosstabs analysis (see table 4, page 136), it is possible to explore the frequency of unit reporting. Three units (Company A, 1st Battalion, 9th Infantry Regiment; Headquarters

Table 3. Descriptive Statistics of Bombing/Explosion Attack Casualties

		Bombing/explosion	Casualties
N	**Valid**	233	230
	Missing	0	3
Mean		NA	8.21
Median		NA	3
Standard deviation		NA	16.47
Variance		NA	271.43
Range		NA	128
Minimum		NA	0
Maximum		NA	128

(Table by author)

and Headquarters Company, 1st Battalion, 508th Parachute Infantry Regiment; and 2nd Battalion, 1st Infantry Regiment) report 41.2 percent of the attacks. Analyzing operations and casualties in these unit areas is an opportunity to understand enemy and friendly tactics related to bombing/explosion attacks.

This part of the example has demonstrated how to explore the data. This is an early step in any analysis. Among the next steps are to use advanced techniques such as machine learning to conduct an analysis or if the data is insufficient, to identify additional data sources. A search for more data could result in extremely large

Mean = 8.21
Standard deviation = 16.475
N = 230

—— Normal
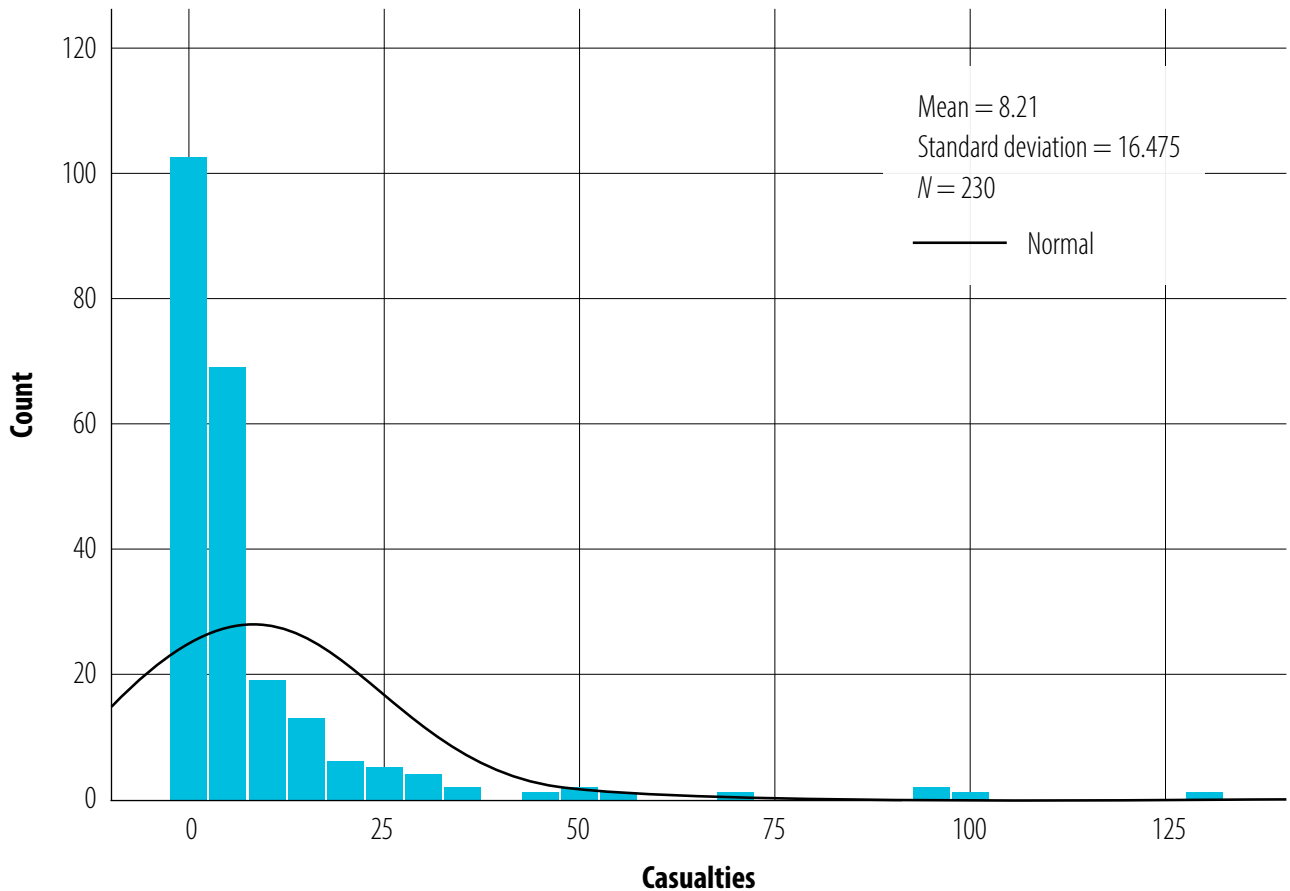
Figure 4. Histogram for Bombing/Explosion Attack Casualties

(Figure by author)

datasets for a small tactical data science team to prepare. Fortunately, since all soldiers in the CP have a core set of skills, the battle captain shift can be used to help with basic data preprocessing tasks and data labeling.

Discussion

A tactical data science practice within CPs allows units to take advantage of locally generated raw data and other sources of raw data. The example demonstrates that meaningful insights are possible with the data managed in small units. This affords leaders better opportunities than the ones they currently have when they must rely on external resources that they do not control. This is not to imply that only locally sourced or managed data is useful. Rather, it shows how commanders can directly interact with data and use it to inform their own decisions and guidance to the battle staff. Furthermore, it prevents the battle staff from being left at the mercy of the external agencies that generate operations and intelligence products that may or may not be timely or solve the local tactical problem.

The Army has one of the most extensive university-level education systems in the United States (based upon the combination of colleges, universities, and scholarships). There is no excuse for such educational potential to be wasted teaching old processes. To gain the most benefit from data, people doing an analysis should be as knowledgeable about the people, process, and technology required for a true digital transformation as they are about fighting.

Ideas about modern technology, data, and fighting should be integrated and complementary. This is in contrast to the construct proposed by some military

Table 4. Crosstabs Analysis Summarizing Unit Reporting of the Bombing/Explosion Type of Attack

	Frequency	Percent	Cumulative percent
1-17 IN	27	11.6	11.6
2-1 IN	31	13.3	24.9
3-17 FA	26	11.2	36.1
4-23 IN	20	8.6	44.6
A/1-9 IN	33	14.2	58.8
B/1-9 IN	15	6.4	65.2
BSTB/5/2 SBCT	27	11.6	76.8
C/1-9 IN	22	9.4	86.3
HHC/1-508 IN	32	13.7	100.0
Total	**233**	**100.0**	

(Table by author)

intellectuals that they are distinct, competitive, and undesirable at the tactical level of war. Military students can, and should, be required to take classes that will teach them how to capture, manage, and analyze killer data as small-unit leaders.

Today, data from the vast collection of DA Forms 1594 are not useful for analysis because the content of these forms is simply not available. The forms are maintained as handwritten or printed papers or electronic static documents on storage devices. The United States has been at war in Afghanistan since 2001. Every deployed Army unit kept records using the DA Form 1594. This means that nearly two decades of organizable, searchable, and maintainable small-unit data have been kept from deploying units.

Conclusion

Managing data in a CP is the first echelon of a tactical data science practice. The second is having trained data scientists as part of the modified table of organization and equipment. CP personnel should manage and explore routine data while data scientists transform and combine it for a deeper and more robust analysis. Data scientists also have skills to create data pipelines that automate processing and moving data.

Creating a tactical data science discipline provides commanders with an ability to use advanced techniques with data for tactical decision-making in combat. A great deal about the environment and enemy can be derived from data that are readily available from the tactical information systems common in small units. However, leaders are missing opportunities to use these data. Tactical data science corrects this and provides the Army with an opportunity to gain a maneuver advantage through the smart use of locally captured and managed data and raw data from other sources.

Finally, in looking toward the future, having tactical leaders who understand data science can alleviate challenges in emerging artificial intelligence programs, such as bias in machine learning models. For example, factors that contribute to model bias are selecting the wrong data to train the model and building models that do not reflect environmental realities as they are based upon incorrect assumptions. To mitigate this, some researchers are creating

audit systems to scrutinize predictive models before they are deployed.[22] Building a tactical data science capability ensures that combat leaders at the tactical level of war understand the basic principles of machine learning and are available to knowledgeably help with the development or governance of artificial intelligence programs. ■

Notes

1. William Reeder, "COL Harry Tunnell Brigade Commander 5/2 SBCT," in *S.L.A. Marshall Combat Leader Interview Series* (Joint Base Lewis-McChord, WA: Battle Command Training Center [BCTC], 2010).

2. Stephen J. Andriole, "Skills and Competencies for Digital Transformation," *IT Professional* 20, no. 6 (2018): 78–81.

3. Foster Provost and Tom Fawcett, "Data Science and Its Relationship to Big Data and Data-Driven Decision Making," *Big Data* 1, no. 1 (2013): 51–59.

4. Dynamic data is data that is updated, streaming data is a constant flow of data, and static data is unchanging. Data can be digital and electronic. We are interested in digital data, which has elements that can be extracted and shared across diverse systems. Some digital artifacts are also electronic data, such as a PDF locked for editing, which is ultimately static data. Metadata is available for all forms of data (dynamic, streaming, and electronic). It can be managed and analyzed using the techniques described in this article.

5. Daniel E. O'Leary, "Artificial Intelligence and Big Data," *IEEE [Institute of Electrical and Electronics Engineers] Intelligent Systems* 28, no. 2 (2013): 96–99; Salvador Garcia et al., "Big Data Pre-Processing: Methods and Prospects," *Big Data Analytics* 1, no. 9 (2016), https://doi.org/10.1186/s41044-016-0014-0.

6. Gil Press, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," *Forbes* (website), 23 March 2016, accessed 2 April 2020, https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#41614c676f63. See Gracia et al., "Big Data Pre-Processing," for more detail about common big data preprocessing tasks such as removing noisy data, data transformation, data cleaning, and normalization.

7. Ali Ahmad Jalali and Lester W. Grau, "Vignette 9: Battle for Chaharqulba Village," in *The Other Side of the Mountain: Mujahideen Tactics in the Soviet-Afghan War* (Quantico, VA: U.S. Marine Corps Studies and Analysis Division, 1995), 311–16.

8. 5th Stryker Brigade Combat Team, 2nd Infantry Division, "Governance Then and Now" (Joint Base Lewis-McChord, WA: Governance, Reconstruction, and Development Fusion Cell, 2010), slide 1.

9. Joseph Turnham, et al., "Digital Air/Ground Integration in Afghanistan: The Future of Combat Is Here!," *Fires Bulletin* (March-April 2012): 57–62.

10. 5th Stryker Brigade Combat Team, 2nd Infantry Division, Operations Research/Systems Analysis, *Honesty Traces: Draw Something ... Not Attention*, U.S. Army Brochure (Kandahar, Afghanistan: Task Force Stryker, 2009).

11. Harry D. Tunnell IV, "CMH-OEF-16-009 Col Harry D Tunnell," in *Chief of Staff of the Army: Operation Enduring Freedom Study Group*, ed. J. Stark (Washington, DC: U.S. Army Center of Military History, 2015).

12. Harry D. Tunnell IV, "Crisis Management and Combat Operations in Afghanistan," in *Crisis Management: A Leadership Perspective*, ed. Jerry D. VanVactor (New York: Nova Science, 2015), 195–209.

13. Ibid.

14. Ibid.; BCTC, *Operation Blowfish: Decision Making Exercise*, DVD (Joint Base Lewis-McChord, WA: BCTC, 2011).

15. Supervised, unsupervised, and reinforcement learning are popular machine learning techniques. Supervised learning requires data to be labeled. Unsupervised learning does not require labeled data. Reinforcement learning has an agent interact with the environment and learn based upon rewards. Supervised learning is arguably the most common machine learning approach.

16. Harry D. Tunnell IV, "Task Force Stryker Network-Centric Operations in Afghanistan," Defense and Technology Papers 84 (Washington, DC: National Defense University Press, 2011).

17. Harry D. Tunnell IV, "The U.S. Army and Network-Centric Warfare: A Thematic Analysis of the Literature," in *2015 IEEE Military Communications Conference: Proceedings of a Meeting Held 26-28 October 2015* (Tampa, FL: IEEE 2015), 904–9.

18. Harry D. Tunnell IV, "Network-Centric Warfare and the Data-Information-Knowledge-Wisdom Hierarchy," *Military Review* 92, no. 3 (May-June 2014): 43–50.

19. Training Circular 3-22.6, *Guard Duty* (Washington, DC: U.S. Government Publishing Office, 13 December 2019, Change 1), E-1–E-3.

20. Field Manual 6-99.2, *U.S. Army Report and Message Formats* (Washington, DC: U.S. Government Printing Office, 2007 [obsolete]), 219–219.1.

21. Harry D. Tunnell IV, "Simulated Combat Reports Dataset," IEEE Dataport, last updated 11 March 2020, accessed 2 April 2020, https://ieee-dataport.org/documents/simulated-combat-reports-dataset.

22. Eliza Strickland, "Racial Bias Found in Algorithms that Determine Health Care for Millions of Patients: Researchers Argue for Audit Systems to Catch Cases of Algorithmic Bias," *The Human OS Blog* (blog), IEEE Spectrum, 24 October 2019, accessed 2 April 2020, https://spectrum.ieee.org/the-human-os/biomedical/ethics/racial-bias-found-in-algorithms-that-determine-health-care-for-millions-of-patients.