# Military Data Space: Challenges, Opportunities, and Use Cases

Paulo H. L. Rettore, Philipp Zißner, Mohammed Alkhowaiter, Cliff Zou, and Peter Sevenich

***Abstract*—Combining big data and Artificial Intelligence (AI) has revolutionized industry and research by enabling accurate predictions and informed decision-making. These advancements have also found their place in the military domain, with initiatives aiming to integrate data sources and sensors from different domains, providing a shared situational awareness. In urban military operations, timely and context-aware information is crucial for achieving precision and success. Data fusion, which combines information from diverse sources, is vital in achieving this goal. Furthermore, civilian data provides crucial context information and can significantly impact mission planning. This article proposes the Military Data Space (MDS) concept to explore how big data can support military decision-making by combining civilian and military data. Use cases are presented, highlighting the benefits of data fusion and image authentication in enhancing data quality and trustworthiness. Furthermore, the challenges of data security, privacy, integrity, acquisition, fusion, networking, and leverage AI approaches are discussed while emphasizing the opportunities to build the next generation of military applications.**

***Keywords* - Big Military Data, AI, Data Fusion, Data Integrity.**

## I. Introduction

The rise of big data has transformed how organizations store, manage, and analyze vast amounts of data. Moreover, the availability of large datasets and the development of more powerful hardware have paved the way for the era of Artificial Intelligence (AI). Despite the limitations, these topics have also found applicability in the military domain. One example is the U.S military's use of Multi-domain Operations (MDO), later expanded to Joint All-Domain Command and Control (JADC2), and the Common Operational Picture (COP) concepts integrating various data sources and sensors across multiple domains (land, sea, air, space, and cyberspace) enabling faster and more knowledgeable decision-making, providing a shared situational awareness across all levels of an organization, from tactical to strategic. Moreover, the NATO community has discussed and tested the Data Lake concept through the NATO Core Data Framework (NCDF) to share reliable information with the coalition partners across domains at the proper time/form.

Utilizing advanced algorithms and computing power, AI can process vast data sets, revealing complex patterns often imperceptible to humans. This empowers defense operations to augment field experiences, facilitate tasks, make data-driven decisions, harmonize data from diverse sources, and bolster preparedness against threats and disasters. By collating data from varied sources, Command and Control (C2) can gain insights into urban landscapes, facilitating context-aware decision-making [1], [2] through data fusion techniques [3], [4]. Modern cities deploy sensor networks, leveraging big data that can support urban military strategies. Additionally, social media platforms serve as a valuable source of text, images, and videos, enriching situational awareness, but also introducing challenges like data integrity. In ”other than war” operations, including countering corrupt governments, narcotraffic, and humanitarian missions, the paramount role of big data, data fusion, data integrity, and AI in mission success becomes evident within the contemporary global landscape.

This article delves into using big data to bolster military decision-making and the associated challenges. It covers aspects relatively underexplored in the field in a non-dense and easy reading. In this context, the study introduces the concept of Military Data Space (MDS), a novel approach that incorporates the Intra-Military Data (IMD) and Extra-Military Data (EMD) to ignite the discussion and the development of military solutions. Then, it illustrates the benefits of big data through use cases focusing on data fusion and image integrity mechanisms. Finally, it discusses the challenges and opportunities of using big data, concentrating on four main aspects that must be considered to support strategic military decisions: i) data fusion, ii) security/privacy and integrity, iii) AI, and iv) networking as the means to access the big data.

The discussion of data dissemination from the network perspective is relevant and widely covered by the literature. Therefore, this study aims to ignite the discussion on the big data point of view and the possibilities of using it to benefit the military systems. Moreover, we emphasize the significance of addressing the challenges associated with integrating IMD and EMD. This integration is crucial for building cohesive big data, ultimately enhancing military decision-making capabilities. In summary, this article's contributions are as follows:

- Introducing a novel concept for integrating military and civilian data: the Military Data Space (MDS) framework.
- Identifying the key challenges and opportunities inherent in big data through the advent of the MDS framework.
- Two illustrative use cases highlighting the advantages of data fusion and integrity in supporting strategic decision-making.

The article is structured as follows. In Section II, the concept of MDS is introduced. Section III reviews recent literature

Paulo H. L. Rettore, Philipp Zißner, and Peter Sevenich are with the Communications Systems Department, Fraunhofer FKIE, Zanderstr. 5 53177, Bonn, Germany. E-mails: {paulo.rettore.lopes, philipp.zissner, peter.sevenich}@fkie.fraunhofer.de. Mohammed Alkhowaiter and Cliff Zou are with the College of Engineering and Computer Science, University of Central Florida, Orlando, United States. E-mails: {mok11@knights, czou@cs}.ucf.edu Paulo H. L. Rettore, Philipp Zißner, and Mohammed Alkhowaiter equally contributed to this manuscript Manuscript submitted July 05, 2023.

on big data in military and civilian scenarios. Section IV presents two use cases illustrating how big data can support military decision-making. The challenges and opportunities of the military data space are discussed in Section V. Finally, Section VI concludes the article by summarizing the key aspects discussed in this study.

## II. Military Data Space

The concept of Military Data Space (MDS) is proposed based on the ideas discussed in [5]. It provides a data-driven perspective of the military scenario and facilitates decision-making based on diverse data sources. MDS consists of two primary categories: Intra-Military Data (IMD) and Extra-Military Data (EMD), as illustrated in Fig. 1. The majority of the current military literature is dedicated only to the IMD proposing and evaluating systems (e.g., middlewares, protocols). However, with the non-precedent growth of Information and Communications Technology (ICT), civilian systems have become an essential source of data and infrastructure (networking) that cannot be neglected anymore. Therefore, MDS aims to support the discussion on how EMD can aid military decisions considering the challenges such as data privacy/security, integrity, acquisition, fusion, networking, and leveraging artificial intelligence.

### A. Intra-Military Data

IMD corresponds to data provided and consumed by the military, classified into two main layers: the infrastructure with real/virtual sensors (from spatial/aerial/ground/nautical unities) and the information layer, including operational, intelligence, and logistics data.

The infrastructure includes data collected by sensors (e.g., radar, sonar, cameras) and other electronic systems that can detect and track objects in the air, on land, or in water; vehicular sensors that provide the status of the military units and surrounding; and wearable/smart and Internet of Things (IoT) devices that support the infantry in the field with GPS position, maps, health measurements, live cameras (high resolution, infrared), etc. These data can be used to monitor and identify potential threats, assist in targeting enemy forces, and monitor the infantry conditions.

Besides the raw data from real/virtual sensors, IMD includes the information layer, which fuses data collected through various sources, from operational to intelligence, to create a more reliable and wide operational view as aimed by JADC2 and COP systems. The intelligence information can help military forces understand the capabilities and intentions of enemy forces, identify potential threats, and plan operations. The logistics data provides information on supplies, equipment, and personnel, such as transportation schedules, inventory levels, and maintenance records. These data are critical for ensuring that military forces have the resources to carry out a mission effectively.

### B. Extra-Military Data

The EMD corresponds to the subset of data provided by real/virtual sensors, individually or fused, that may describe the environment around the military operation. In that way, two main layers of data that may be used to support military operations can be defined: the infrastructure (e.g., transportation systems, weather, authorities) and the information (e.g., social media, news, government reports). These layers produce vast and highly variable information, from users' feelings and photos of real-time events (e.g., accidents, corruption, and terrorism) to traffic/weather conditions and people/drivers' behavior in urban environments.

The growth of ICT in urban areas has led to the emergence of Smart Cities, which address urbanization challenges through enhanced mobility, safety, and health solutions. Smart city infrastructure incorporates sensors capturing valuable data on vehicles, traffic, weather, and driver behavior. The proliferation of sensors and IoT devices also generate large volumes of data, enabling the development of intelligent systems leveraging cloud-based communication technologies and AI-based applications. Empowered by big data, data fusion emerged, integrating data from multiple providers to enhance quality and coverage and reduce massive data traffic. Fusing data from transportation, weather, cameras, health systems, and so on has the potential to support not only civilian applications but also strategic military operations by providing contextual data. When sensor infrastructure is limited, data from media sources like social media and government reports may enable the understanding of local behavior and identify factors impacting criminality, corruption, and narcotraffic.

Social media data is valuable for supporting emergency and disaster-related information, complementing other sensor data by capturing unique information (e.g., the location of groups requiring rescue or the presence of hidden individuals). Stationary sensors on buildings and surveillance cameras aid in human tracking for precise location identification. Combined with other data sources, social media data facilitates enemy detection and tactical planning. Transportation-related sensor data, particularly traffic surveillance cameras, play a significant role in emergency response and military logistics. It enables the detection of congestion and blockages resulting from incidents, allowing for improved route planning and traffic management during military operations. Integrating all collected information enhances situational awareness and facilitates effective planning and management of operations in urban environments.

Several initiatives have emerged in response to recent events, such as Russia's war and the challenges posed by anti-democratic extremists in countries like the U.S. and Brazil. One example is the ACLED (Armed Conflict Location & Event Data) project, which offers real-time global data on political violence and protest events. Another noteworthy initiative is DATTALION, an extensive open-source photo and video footage database capturing Russia's war against Ukraine. The primary objective of this database is to counter the Russian government's dissemination of misinformation. The United Nations Development Programme (UNDP) utilizes Machine Learning (ML) algorithms and big data to detect war-damaged infrastructure in eastern Ukraine. A semantic damage detector (https://tinyurl.com/semdam) employing satellite imagery and ground-based photos trains the algorithms to iden-

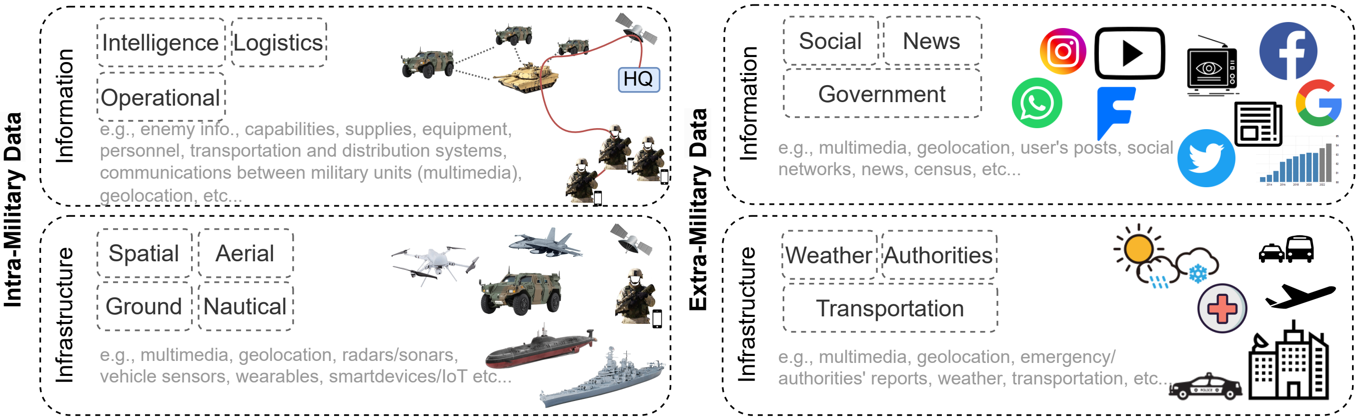

Intra-Military Data
Information
Intelligence
Logistics
Operational
HQ
e.g., enemy info., capabilities, supplies, equipment, personnel, transportation and distribution systems, communications between military units (multimedia), geolocation, etc...
Infrastructure
Spatial
Aerial
Ground
Nautical
e.g., multimedia, geolocation, radars/sonars, vehicle sensors, wearables, smartdevices/IoT etc...
Extra-Military Data
Information
Social
News
Government
e.g., multimedia, geolocation, user's posts, social networks, news, census, etc...
Infrastructure
Weather
Authorities
Transportation
e.g., multimedia, geolocation, emergency/ authorities' reports, weather, transportation, etc...

Fig. 1. Military Data Space.

tify potential damage in buildings, roads, and bridges, assisting local authorities and humanitarian organizations in prioritizing the actions. These initiatives contribute significantly to the MDS, specifically to the EMD, providing valuable resources for analysis and research.

## III. Big Data in the Military Domain

This section examines the application of big data in the military domain, focusing on both *intra*- (IMD) and *extra*-data (EMD) perspectives to outline the significance of big data in military operations and explore recent solutions that leverage its potential.

### A. Intra-Military Data

Some challenges of big data in the military field are presented in the literature and subject of discussion by the NATO community, such as operational security, hardening against vulnerabilities, and data reliability [1], [2], [6], and NATO IST-160, and IST-173. Incorporating autonomous isolation with little connection to the outside world (e.g., EMD) can limit the free flow of big data, demanding creative ways to utilize it while maintaining the autonomy and protection of the systems. In this direction, COP and JADC2 have guided researchers and industries toward using and fusing data from different military entities supporting strategic decisions.

Kun et al. [1] propose a detailed technical plan for constructing a big data platform in military enterprises, establishing multi-level data channels, and enabling comprehensive data management and control. The platform facilitates data collection, organization, processing, and analysis, transforming data into knowledge to enhance decision/service support, innovation, quality control and risk management. Xu et al. [6] emphasize the importance of data science in achieving information superiority in contemporary warfare. Their systematic review reveals a significant focus on data science risks in the social science literature, which can influence political and military policymakers. However, scientific literature lacks attention to risks at operational and strategic levels compared to the tactical level, indicating a research gap. This gap may arise from the lack of connectivity between IMD and EMD, which may support operational and strategic decisions.

### B. Extra-Military Data

*1) Data Fusion:* Big data plays a crucial role in heterogeneous data fusion, aiming to combine multiple records into a consistent representation, improving data quality and reducing communication overhead. However, challenges arise due to data semantics and spatiotemporal coverage. In military applications, heterogeneous data fusion is valuable for designing information systems that enhance information superiority and awareness in complex urban warfare or counter-terrorism scenarios. Robust systems are essential to handle sensitive data (e.g., personal data or strategic mission/governmental plan). Data fusion mitigates information overload, enhances accuracy, and leverages knowledge to support strategic operations and situation assessment [3].

The Multi-Sensor Data Fusion (MSDF) approach is an example of providing fast and efficient target detection, tracking, and threat evaluation in tactical scenarios, as shown in [4]. Another area where data fusion shows promise is the incorporation of Location based Social Media (LBSM), which can enhance knowledge in various fields, including transportation with traffic characterization and incident detection [7]. More detailed transportation data can be obtained by leveraging LBSM systems, benefiting military logistics. The potential of LBSM systems can be harnessed in specific military contexts to augment data availability and enable context-aware operations.

*2) Data Security, Privacy, and Integrity:* Security and privacy are critical considerations in designing military systems that store and gather information in databases. Security aims to prevent unauthorized data modification, while privacy safeguards individuals' information [8]. However, gathering data from open sources, especially from regular users (EMD), poses risks to system security and user privacy, making them vulnerable to attacks and data breaches. IBM's "Cost of a Data Breach Report 2022" highlights a 2.6% increase in cyberattack costs compared to the previous year, with the global average data breach cost reaching 3.35 million USD. Additionally, the report reveals that 83% of organizations studied experienced multiple data breaches, underscoring the challenges of securing these systems.

Data integrity is critical to maintaining trust in the MDS [9].

Manipulated data can have significant consequences, affecting both civilian and military decision-making processes and undermining confidence in data sources. This challenge is exemplified by the proliferation of misinformation on social media platforms, often exploited for political influence, as observed in the ongoing conflict in Ukraine. In response to such issues, platforms like Twitter have revised their policies, labeling many tweets linked to Russian state-affiliated media and detecting billions of real-time tweet impressions related to the conflict [10].

Meanwhile, image authentication has emerged to address concerns about image integrity and origin verification. However, the rise of advanced image manipulation tools, including AI-powered software, has made image verification increasingly tricky. While various techniques such as watermarking, digital signatures, and Perceptual Hashing (pHash) have been introduced for image validation [11], each has advantages and limitations. Watermarking, for instance, offers authenticity and ownership protection but may compromise image quality and can be vulnerable to advanced processing techniques. In contrast, pHash provides flexibility for image operations and sensitivity to content changes, making it particularly well-suited for use on social media platforms. Considering these challenges and solutions in the context of data integrity and image authentication is crucial.

## IV. Use Cases

### A. Data Fusion

First, the spatiotemporal fusion of big data is motivated to support military decisions. Due to the lack of available IMD as discussed, the Multi-Data Fusion (MDF) framework [12] is instantiated to collect, prepare, and process EMD, fusing them to provide enriched information. To demonstrate the enrichment of spatiotemporal data, MDF acquired transportation system data due to the public availability of cloud-based systems sharing data. However, the framework is extensible to various further data types. The goal is to improve data quality, C2 systems, and military logistics and support the COP/JADC2 in urban areas, allowing the creation of novel approaches that use fused EMD with the available IMD from different domains. In the following, Fig. 2 describes the main functionalities of MDF. Furthermore, the benefits of fusing big data are discussed by analyzing numerical results.

For the data acquisition, Fig. 2 (1), a set of parameters (e.g., region, request frequency) and data sources are configured, for which MDF collects data in various formats, storing them in files. In the preparation phase (2), the input dataset is standardized by converting the different feature names and types into a uniform representation. This includes a variety of data mappings to generate uniform data types, e.g., mapping descriptive to numerical values or reducing data granularity. Moreover, map-matching is initiated to fuse all geo-located data, which may have different precisions, into the same road network. MDF pre-processes all collected data and acquires a Shapefile (SHP) from the collected area. Notice that, depending on the application goal and the available data types, the framework may apply different feature extraction methods such as Natural Language Processing (NLP) (sentiment analysis, keyword extraction, lemmatization, stemming, and automatic summarization) or image processing (image segmentation, edge detection, and object detection) to extract the information from non-structured data types. In the use cases, we did not use NLP algorithms as the data was text-free image and transportation-based. However, the proposed data fusion framework is versatile and can handle various data types, including text data, where NLP techniques can be applied.

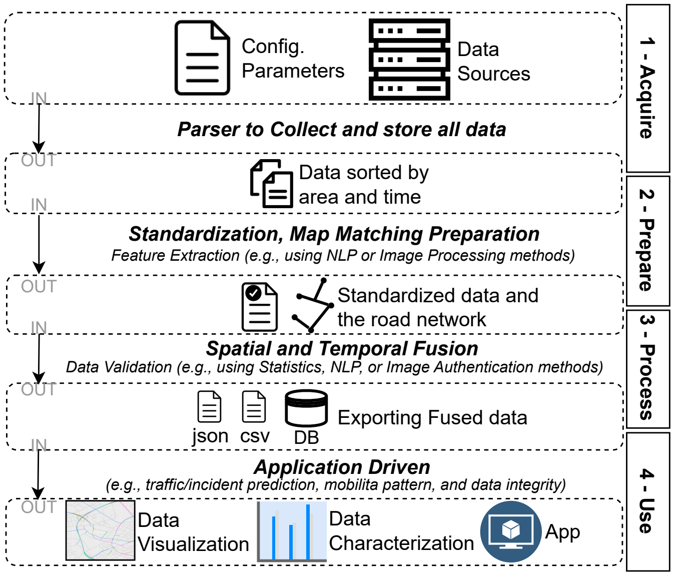

Config. Parameters
Data Sources
1 - Acquire
IN
Parser to Collect and store all data
OUT
Data sorted by area and time
IN
Standardization, Map Matching Preparation
Feature Extraction (e.g., using NLP or Image Processing methods)
2 - Prepare
OUT
Standardized data and the road network
IN
Spatial and Temporal Fusion
Data Validation (e.g., using Statistics, NLP, or Image Authentication methods)
3 - Process
OUT
json
csv
DB
Exporting Fused data
IN
Application Driven
(e.g., traffic/incident prediction, mobilita pattern, and data integrity)
4 - Use
OUT
Data Visualization
Data Characterization
App

Fig. 2. Workflow of the data fusion framework.

The third phase implements temporal/spatial data fusion and data export. To ensure data integrity, non-trusted or biased information is filtered in advance, e.g., using methods to validate the information based on its occurrences over different data sources or image authentication mechanisms as discussed in Section IV-B. Temporal data fusion is achieved by grouping the data within an arbitrary time window (e.g., minutely, hourly, daily). To perform the spatial fusion, MDF leverages map-matching, aligning GPS points under a defined degree of accuracy based on an underlying road network. This is mandatory given the varying precision of GPS reports in different data sources, resulting in all geo-located data being mapped into the same road network.

Finally, in Fig. 2 (4), the enriched data are exported in different formats, offering many possibilities for military and civilian fields. The outputs of MDF support a spatiotemporal analysis by creating different types of statistics and visualizations, characterizing the available information under various spatial and temporal aspects.

*1) Results:* To show the benefits of data fusion, Table I summarizes the results of the MDF framework in a real-world experiment. The experiment covered nine months, collecting four types of civilian transportation data (traffic levels, incidents, vehicular data, and weather conditions) in two different cities. The data fusion increased the data coverage by 173% in Cologne, covering 5081 roads compared to 1379 using only

TABLE I
COVERED ROADS BY THE DATA SOURCE.

| | **Bonn** | | | **Cologne** | | |
|---|---|---|---|---|---|---|
| Source | Total Roads | Unique Roads | Fusion Portion | Total Roads | Unique Roads | Fusion Portion |
| Traf. HERE | 684 | 339 | 21.0% | 2940 | 1379 | 27.1% |
| Traf. OD | 581 | 195 | 12.0% | 914 | 173 | 3.4% |
| Inc. HERE | 206 | 53 | 3.3% | 946 | 370 | 7.3% |
| Inc. BING | 597 | 256 | 15.8% | 1944 | 821 | 16.2% |
| Inc. OD | 52 | 31 | 1.9% | 193 | 86 | 1.7% |
| Envirocar | 433 | 178 | 11.0% | 905 | 245 | 4.8% |
| **Overlap** | **567** | **567** | **35.0%** | **2007** | **2007** | **39.5%** |
| Total | | 1619 | | | 5081 | |

For more numeric analysis, access: https://github.com/prettore/DataFITS

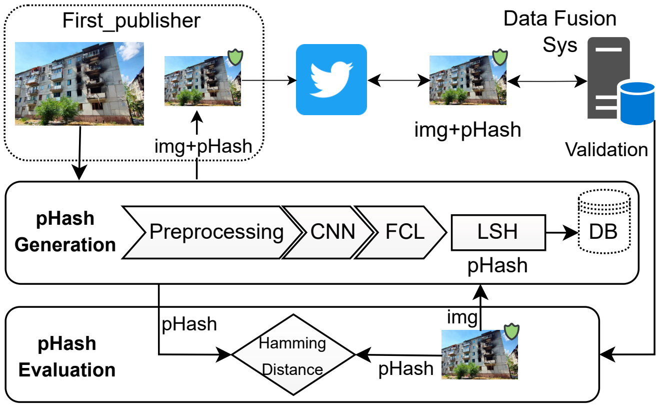


Fig. 3. Image-Fact-Checker (IFC).

*Traffic HERE* data source, reaching an increase of 137% in Bonn. Moreover, the potential of information enrichment is given through overlapping road segments, reaching 39.5%, which provides a detailed description of the event from multiple sources.

### B. Data Integrity

The previous work [11] introduced an image authentication system using Twitter and Facebook to ensure image integrity. It employed a Convolutional Neural Network (CNN) and Fully Connected Layers (FCC) for feature extraction, Locality Sensitive Hashing (LSH) for hash construction, and a contrastive loss to maximize differences between original and manipulated images. The model's output is a fixed-length vector representation of 1024 bits for each image.

To address the importance of maintaining image integrity in urban military operations and civilian systems, Image-Fact-Checker (IFC) is proposed as shown in Fig. 3. It detects fake images, ensures data trust, and serves as an authentication system led by authorities to combat misinformation. The system generates a verified version of the photo with a logo or icon, indicating its validation through the IFC system. Additionally, the IFC provides a Perceptual Hashing (pHash) string representation of the image, which can be included in descriptions or shared on other websites. The data fusion system is a possible end-user to the IFC, verifying crawled images before applying spatiotemporal fusion and generating enriched data.

The concept of having an automated system that provides instant and authentic information is relatively new, making it challenging to assess its effectiveness through comparisons. However, implementing an image authentication system is now crucial due to the rise of AI generative models creating convincing fake images. Adding this system as an authentication layer can help prevent or reduce the spread of misinformation, especially in light of evolving Internet regulations that penalize platforms lacking anti-misinformation measures. One effective approach is to connect the IFC system with government institutions. The IFC approach is versatile and scalable, fostering increased awareness and trust among individuals.

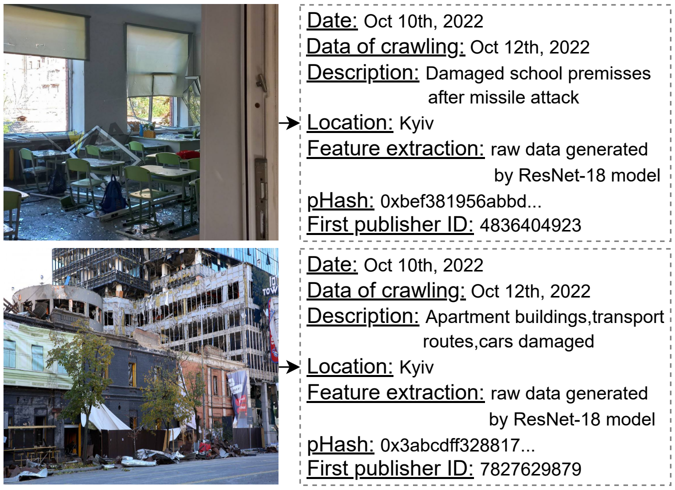
*1) Results:* Using the IFC system improves data trustworthiness and detects image manipulation. In conflict situations, like the Ukraine-Russia conflict , civilians impacted by Russian attacks share images on social media, but questions about their authenticity arise. IFC can verify these images using the DATTALION dataset, allowing for quick distribution to relevant organizations such as rescue teams, the United Nations, or NATO. This speeds up responses to attacks and provides credible evidence against Russia. In civilian scenarios like transportation, accessing real-time and verified information from regular users can enhance decision-making for route updates and emergency responses.

In Fig. 4 (left), two unverified images are gathered from regular social media users through DATTALION. These images represent a small fraction of a larger dataset. Users often hesitate to trust these sources, making it challenging to use them effectively. However, when these images are processed through the IFC mechanism, their reliability increases because any further manipulation becomes easily detectable. After applying IFC, each image receives a pHash, and relevant information such as image description, extracted features, location, event date, crawling date, publisher ID, as depicted in Fig. 4 (right). For future queries, these processed images are stored in the IFC database. This database serves various purposes: duplication detection, integrity verification, and catering to specific end-user requirements.

## V. Challenges and Opportunities of the Military Data Space

### A. Data Fusion

The first challenge of data fusion is finding and acquiring available data in military and civilian fields. Due to privacy/security concerns, information may not be widely available or with limited access. Data is a subject of even more constraints in the military field (IMD), opening opportunities to explore the available civilian data (EMD) to support strategic informational decisions. The second noticeable challenge is the fusion of multiple data sources that could have different structures (structured, semi-structured, and unstructured data), standards, data types (e.g., text, image, video), measurement units, granularity, and spatiotemporal coverage. Therefore, it requires a deep view to prepare and process the different datasets into a fused one.



Date: Oct 10th, 2022
Data of crawling: Oct 12th, 2022
Description: Damaged school premisses after missile attack
Location: Kyiv
Feature extraction: raw data generated by ResNet-18 model
pHash: 0xbef381956abbd...
First publisher ID: 4836404923

Date: Oct 10th, 2022
Data of crawling: Oct 12th, 2022
Description: Apartment buildings,transport routes,cars damaged
Location: Kyiv
Feature extraction: raw data generated by ResNet-18 model
pHash: 0x3abcdff328817...
First publisher ID: 7827629879

Fig. 4. Extracted image details with IFC.

Working with images and text from social media requires further procedures such as feature extraction methods (e.g., NLP and image processing) to extract usable information. Despite the challenges of data fusion, the benefit of combining different data sources describing the same space and time from different perspectives (e.g., command, users, journalists, governments, sensors) can enhance the planning and strategy phase of military operations, supporting the COP and JADC2 systems.

### B. Data Security, Privacy, and Integrity

*1) Data Security and Privacy:* Protecting sensitive military information is crucial for national security. Requiring robust encryption, secure data storage, and access controls to mitigate risks. Proposed techniques include Public Key Infrastructure (PKI) security, protected cores, data encryption, firewalls, and intrusion detection. However, balancing data sharing, beneficial for information fusion, and security/privacy measures remains a challenging task for the military.

*2) Data Integrity:* Manipulated data risks civilian and military decision-making and deteriorates trust in data providers. The rapid spread and increased engagement of manipulated images, aided by advancements in AI models generating content, highlight the need for smart and integrated solutions. Images shared via social media have the power to quickly convey complex ideas that can support rescue operations, enabling immediate actions such as altering transportation routes in case of urban incidents/disasters. Images can also evoke emotional connections and enhance readers' understanding of news events. However, crises like the Ukraine war have amplified the dissemination of misinformation, necessitating the involvement of human fact-checkers such as snopes.com and norc.org to combat misinformation. Nevertheless, real-time human-based verification during wars or to combat corrupted governments can be time-consuming, creating opportunities for the design of automated systems to authenticate images and address misinformation.

### C. Networking

While the primary focus of this work lies in the data perspective and the relevance of ensuring the use of trusted data from diverse sources to support military operations, it is also important to acknowledge the significance of networking in delivering data and services efficiently. In network-centric military operations, wireless communication is vital, utilizing various technologies such as HF, VHF, UHF, SatCom, Wi-Fi, and LTE 4-5G. Some excel in long-range coverage but have limited bandwidth, high latency, and are prone to disruptions. Others prioritize reliability with a shorter range, greater bandwidth, and lower latency.

Network paradigms like Information-Centric Networking (ICN) and Software-defined Networking (SDN) are crucial for optimizing data dissemination and network orchestration [13], especially in scenarios with limited network resources. In military networks, particularly at the tactical edge, challenges like constrained resources and security concerns arise during data dissemination. To address these, the military may explore diverse infrastructures, including civilian networks, to acquire and fuse non-military data. 5G technology, exemplified by the European consortium 5G COMPAD, is being considered. However, it is challenging due to costly, hardware-specific communication systems with limited bandwidth and interoperability. This calls for tailored reference architectures to meet military communication needs.

The recent Ukraine-Russia conflict exposed communication network vulnerabilities when Russian attacks on Ukraine's infrastructure caused Internet outages. SpaceX's Starlink satellite Internet constellations offered a solution, demonstrating the value of utilizing civilian network infrastructure during wartime. Despite its promise to improve Internet reliability for data and emergency communication, this technology confronts challenges related to cybersecurity, coverage, reliability, and cost-effectiveness.

### D. Artificial Intelligence

Accessing military-owned big data for AI research poses challenges due to privacy, security, and restrictions imposed by military agencies to prevent misuse and limit the availability of the IMD. Moreover, the AI functionality can be compromised by adversarial attacks, which deceive AI models through changes causing misclassification. Techniques like Fast Gradient Sign Method (FGSM) and semantic attacks help identify and mitigate such attacks in computer vision and NLP, respectively. Yuan et al. [14] provide a comprehensive review of attacks, countermeasures, and application-based taxonomies.

To detect adversarial attacks, one effective approach is to use a secondary ML model with distinct characteristics from the primary AI model. This idea draws inspiration from the early stage of satellite communication. In those days, a secondary system, such as a telegraph, was employed to prevent man-in-the-middle or jamming attacks on satellite communication. With limited bandwidth, the secondary system only transmits summary data corresponding to the complete satellite data for detecting attacks and emergency communication. Similarly, when safeguarding against adversarial attacks in AI, traditional ML can serve as a secondary system to produce results that align with the primary CNN approach. Adversarial attacks rely on gradient techniques in computer vision deep learning models, whereas traditional ML methods use different approaches that are mostly immune to those attack manipulations.

Another aspect that concerns using AI in the military is the need to share sensitive data for training the models. In this direction, Federated Learning (FL) has emerged as a technique for training ML models where the data is not exposed, ensuring data security and privacy [15]. While it cannot be regarded as a defense technique against adversarial attacks, this approach hides sensitive data and portions of models or parameters. This technique is valuable for emerging military applications built upon AI.

## VI. Conclusion

This article explores the use of big data in the military domain. The opportunities and challenges associated with integrating diverse data sources, ensuring data security, privacy, and integrity, as well as networking, and leveraging AI, have been examined. The concept of MDS is introduced to enrich and guide the discussion, emphasizing the potential of incorporating civilian data to enhance the quality and quantity of information for strategic decision-making in military operations. Furthermore, the article includes two practical use cases illustrating the benefits of data fusion and the importance of implementing image authentication mechanisms to maintain data integrity. These findings highlight the significance of big data in the military and emphasize the need for further research and exploration in the field.

## References

[1] W. Kun, L. Tong, and X. Xiaodan, "Application of big data technology in scientific research data management of military enterprises," *Procedia Computer Science*, vol. 147, pp. 556–561, 2019.

[2] J. Cui and S. Rao, "Us army big data military applications and reflections," in *Proceedings of the 2021 3rd International Conference on Big-Data Service and Intelligent Computation*, ser. BDSIC '21. New York, NY, USA: Association for Computing Machinery, 2022, p. 92–96.

[3] S. Jusoh and S. Almajali, "A systematic review on fusion techniques and approaches used in applications," *IEEE Access*, vol. 8, pp. 14 424–14 439, 2020.

[4] R. P and T. Mishra, "Target detection, tracking and threat evaluation in multi sensor system using machine learning," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 2020, pp. 837–842.

[5] Rettore, P. H. L., G. Maia, L. A. Villas, and A. A. F. Loureiro, "Vehicular data space: The data point of view," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2392–2418, thirdquarter 2019.

[6] L.-l. Xu and F. Jin, "Research on military intelligence value evaluation method based on big data analysis," in *Multimedia Technology and Enhanced Learning*, Y.-D. Zhang, S.-H. Wang, and S. Liu, Eds. Cham: Springer International Publishing, 2020, pp. 192–200.

[7] Rettore, Paulo H. L., B. P. Santos, R. Rigolin F. Lopes, G. Maia, L. A. Villas, and A. A. F. Loureiro, "Road data enrichment framework based on heterogeneous data fusion for its," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1751–1766, 2020.

[8] P. Jain, M. Gyanchandani, and N. Khare, "Big data privacy: a technological perspective and review," *Journal of Big Data*, vol. 3, pp. 1–25, 2016.

[9] G. Warner-Søderholm, A. Bertsch, and A. Søderholm, "Data on social media use related to age, gender and trust constructs of integrity, competence, concern, benevolence and identification," *Data in Brief*, vol. 18, pp. 696–699, 2018.

[10] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE transactions on multimedia*, vol. 19, no. 3, pp. 598–608, 2016.

[11] M. Alkhowaiter, K. Almubarak, M. Alyami, A. Alghamdi, and C. Zou, "Image authentication using self-supervised learning to detect manipulation over social network platforms," in *2022 IEEE Military Communications Conference (MILCOM)*, 2022, pp. 672–678.

[12] P. Zißner, P. H. L. Rettore, B. P. Santos, J. F. Loevenich, and R. R. F. Lopes, "Datafits: A heterogeneous data fusion framework for traffic and incident prediction," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–0, 2023.

[13] G. M. Leal, I. Zacarias, J. M. Stocchero, and E. P. d. Freitas, "Empowering command and control through a combination of information-centric networking and software defined networking," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 48–55, 2019.

[14] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2805–2824, 2019.

[15] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Generation Computer Systems*, vol. 115, pp. 619–640, 2021.

**Paulo H. L. Rettore** is a scientist at Fraunhofer FKIE in Bonn, Germany, with a Ph.D. in Computer Science. His research interests include MANET, Tactical Networks, SDN, IoT, ITS, 5G, and Smart Mobility.

**Philipp Zißner** is a scientist at Fraunhofer FKIE in Bonn, Germany, with research interests in Intelligent Transportation Systems, Smart Mobility, the Internet of Things, and Tactical Networks.

**Mohammed Alkhowaiter** is pursuing his Ph.D. at the University of Central Florida, USA, interested in Computer Security, Network Security, Digital Forensics, Image Authentication, and Machine Learning.

**Cliff Zou** is a Professor at the University of Central Florida, USA, specializing in cybersecurity and computer networking.

**Peter Sevenich** is the head of the Robust Heterogeneous Network group at Fraunhofer FKIE in Bonn, Germany, with expertise in military communications, IP routing, Service-Oriented Architecture, SDN, and SDR.