Privacy Concerns with Big Data Analytics: US DoD/Army Landscape

Mariusz A. Balaban
NTC
Fort Irwin, CA
mariusz.a.balaban.civ@mail.mil

ABSTRACT

Big data analytics is a relatively new field but matured enough to provide innovatory solutions to automate mining and extracting data from the next generation Mission Command (MC) and Live, Virtual and Constructive (LVC) simulation systems. It provides powerful technologies and methods to quickly analyze huge amounts of data but also introduces potential harms to individuals whose personal data is collected, stored, analyzed, used for decision-making, and disclosed. Desire to keep big datasets indefinitely, reuse it for different projects, combine it with additional data, and automate decisions based on data presents privacy and security challenges. Moreover, big data may increase system opacity because of many streams of data created by multiple stakeholders, complicated algorithms processing the data, multiple storage locations, and multiple data consumers with different data aggregation needs. This increased complexity can lead to data leaking, breaches, spillages, and re-identification of individuals. On the other hand, it is critical to take full advantage of big data by processing special data categories on individuals and a variety of data types. Without using combined personal attributes, quasi-identifiers, and sensitive attributes combined with insensitive attributes data utility decreases or may even render analysis useless. This paper discusses the importance and benefits of using big data especially focusing on the US Department of the Army (DA). It presents privacy laws, policies, and regulations relevant to the DoD/DA and investigates their incompatibilities with big data principles. Moreover, it identifies privacy-preserving components relevant to big data, allowing for a balanced approach that benefits the DoD/DA while preserving the privacy of individuals.

ABOUT THE AUTHORS

Mariusz A. Balaban, Ph.D., is a Battle Command and Simulations Officer at National Training Center, Fort Irwin. He received a Ph.D. in modeling and simulation (M&S) from Old Dominion University. His topics of interest include representation of complex systems-of-systems, advancing M&S methodology, conceptual modeling, input modeling, output analysis, selection and use of multiple M&S methods, interoperability, distributed simulation architectures, and collaborative M&S. He is also interested in big data, IT, and IoT, especially in the context of M&S enterprise allowing to achieve real-time analysis and control of a complex system of systems. He is an author and co-author of many papers in the area of M&S. Dr. Balaban is a member of SCS, ACM SIG**,** and GK International.

Privacy Concerns with Big Data Analytics: US DoD/Army Landscape

Mariusz A. Balaban
NTC
Fort Irwin, CA
mariusz.a.balaban.civ@mail.mil

1. INTRODUCTION

Big data analytics is a relatively new field. The initial version of Hadoop, which can be associated with the inception of big data, was released only in 2007 while its first stable release was in 2011. In contrast, privacy considerations existed in philosophical discussions for centuries and started to appear as laws by the end of the 19th century. Big data technologies decrease the processing time of large volumes of data facilitating faster insights than traditional analytical solutions. They offer unparalleled capabilities for generating insights from the vast volume of batch data and more recently real-time data; mixing different types of inputs such as text, audio, and video. Data science provides descriptive, predictive, discovery, and prescriptive methods, which amplified by scalable big data technologies enable new insights and decision support capabilities that military commanders could not have previously leveraged. The big data ecosystem provides powerful technologies and methods to quickly analyze huge amounts of data but also introduce potential harms to individuals whose personal data are collected, stored, analyzed, used for decision-making, and disclosed. Personally Identifiable Information (PII) is information which can be used to distinguish an individual's identity, such as their name, biometric records, social security number, or to trace individuals when linked with other personal or identifying information such as date of birth, place of birth, and mother’s maiden name. Pictures and audio recordings could also be used to identify individuals using facial and vocal recognition technologies. DoD must balance policy to mitigate privacy and security risks related to big data while providing its greatest benefits for the military.

The lack of persistent on-demand access to performance data in support of LVC training, T&E, analysis, and experimentation inhibits advantage of new big data technologies and analytics even though military organizations could provide large datasets. Because of the fast growth and wide availability of big data technologies and analytics one should assume they are being leveraged by opponents. The inability to quickly adapt big data could create negative overmatch, hence, DoD needs to make big data dominance its high priority item. The data strategy should highlight the need for new effective data architectures with supporting infrastructure(s) enabling collection, availability, and analysis of the vast amount of individual and aggregated data from training events and real operations. This is required to enable more objective insights for acquisition, improve training, and decision-making related to multiple levels of conflict: political, strategic, operational and tactical.

Privacy regulations guide military organizations with privacy obligations but because of some of its contradicting principles with big data principles, one must find an acceptable common denominator by examining both perspectives. The remainder of this paper is organized as follows. Section 2 highlights a few examples of big data applications that can benefit the military. Section 3 briefly introduces relevant privacy laws, policies, and regulations with their limitations in the context of big data analytics. Section 4 describes the selected privacy-preserving components. Finally, Section 5 offers recommendations for handling privacy in military big data analytics and highlights applicable research gaps.

2. NEED FOR MILITARY USE OF BIG DATA ANALYTICS

Jensen and Ramachandran (2018) discussed a data-driven machine learning (ML) approach to modeling constraints used for a satellite planning exercise. They pointed out the importance of obtaining adequate data to cover decision-making factors for the required training tasks and parallels between operational and training data. Craven, Oden, Landers, Shah, and Shah (2018) developed ML model, demonstrating the ability to interpret the behaviors of experienced fighter pilots during flying missions on a realistic desktop flight simulator, useful to commanders in a mission debrief. ML model was trained using simulated mission data, which allowed classifying different parts of the mission into their respective phases. Deep Learning (DL) emerges as a core component of many innovatory solutions for the military. DL depends on big data sets, which are required to train Deep Neural Networks efficiently (Woodard & Enloe, 2018). Clinger et al. (2018) introduced the concept of Intelligent After-Action Review (IAAR) where training systems will automatically create individually targeted AAR products. The authors pointed out that the availability of suitable live exercise data as the largest hindrance as ML crave data to minimize model errors.

LVC training enhancements can be subjectively assessed based on measures of training quality, quantity, and adaptability (Shanley, Crowley, Lewis, Leuschner, & Masi, 2007). A critical piece of objective evaluation of defense acquisition systems is to capture the readiness of forces before and after new capabilities are implemented to show the actual difference (Balaban, 2015). Improvements in efficiency and effectiveness of training offered by adaptive training solutions, enabling more objective assessment of military readiness, require collecting a huge amount of data for a large number of performance and effectiveness measures (Freeman, Nicholson, Squire, & Bolton, 2014).

Higher accessibility of training by the end-users is required to improve readiness. This requires to close the gap in "…how applications, data, and services are hosted to meet future warfighter needs and exploit the evolving, data-driven, interconnected eco-system, despite resource constraints" (Marrou, Glenn, & Nielsen, 2018, p. 1). Durlach (2018) described the capabilities of the Army's planned Synthetic Training Environment (STE) focusing on the training management tools. STE aims to fulfill the vision of accessibility from Point of Need (PoN), e.g., home stations, armories, training centers, and during deployments. STE intends to enable multiple repetitions of training scenarios on various complex operational environments with specific locations, opponents, and multi-domain conditions. As part of STE, large datasets collected from constructive simulations, individual and collective virtual simulators, and live training will enable next generation of training feedback and analysis of training performance and effectiveness at different levels of aggregation, e.g., individual participant/equipment, small unit performance, and higher echelons. AI proficiency models supported by big data could provide near real-time feedback and intelligent tutoring for service members and civilians. Moreover, big data technologies enable more scalable and real-time data processing solutions which could be useful to query a large terrain or to develop real-time simulation engines.

Using big data analytics combined with simulation models could automate the mining and extraction of relevant mission data as part of the Common Operational Environment (COE) and enhance ISR capabilities to improve Common Operational Picture (COP). COE tools could offer improved real-time data services about individual military personnel, including their performance trends, current status, and health info enabling more objective training and evaluation analysis and improved SA during missions. Soldier performance is influenced by morale, but the degree of this influence is difficult to assess without data or when based on a single point estimation (Balaban, Mastaglio, Sokolowski, & Ezell, 2014). More granular data from sensors, e.g., accelerometers, orientation sensors, optical sensors, acoustic/sound sensors, proximity sensors, velocity sensors, infrared sensors, thermal sensors, pulse sensors/ heart rate monitors, blood pressure sensors, sleep monitors, GPS sensors, humidity and gas sensors, electromagnetic sensors, force sensors, and light sensors can help to capture individuals' state, actions, and interactions. Real-time personal data feeds in the context of objective readiness assessment and as an input to behavioral models should be investigated. For instance, Huber, Winslow, Chiang, and Aziz (2018) created a stress classifier that allowed the monitoring of stress in real-time. Similarly, capturing the state of the environment during training (Freeman et al., 2014) and mission events are also needed. Data collection on individual service members and environment during multiple training phases and modes of LVC training and real missions could potentially enable better human behavior representation to include cognitive, physical and psychological performance effects of combat on unit performance. Combining MC planning data with simulation tools could enable simulated effect-based course of action (COA) analysis and facilitate more objective war-gaming training. For instance, Harder, Blais, and Balogh (2017) proposed a conceptual framework for an automated battle planning system using combat simulations. The real-time big data technologies could provide the scalability required for the back-end processing of real-time data and simulations. Moreover, batch data processing using massive LVC training data could be used to develop more objective MoP and MoE training, validate TTP concepts, and provide continuous improvements of M&S capabilities.

Danner and Djahandari (2008) discussed policy, management and technical challenges to information sharing during multi-national training events requiring cross-domain solutions (CDS). CDS allows US and coalition partners to train together in a distributed simulation environment. Big data sharing between coalition partners during training could improve the evaluation of readiness based on data collected on individual players, systems, and units during multi-national exercises, enabling more objective assessments within NATO. Similarly, better intelligence sharing between coalition partners is required to counter hybrid threats (HT), especially if some threats are automated (Balaban & Mielniczek, 2018). Big data technologies are enablers for the development of an automated HT management system, which could include data collection nodes, communication of threats between coalition partners, storage of threat reports, and data processing to provide an overview of how each case progresses. Moreover, simulation models combined with AI models could suggest responses to HTs considering all PMESII-PT operational variables of the battlespace (Balaban & Mielniczek, 2018).

3. PRIVACY IN THE ERA OF BIG DATA ANALYTICS: THE DOD/DA PERSPECTIVE

An important step when planning a big data system/project is to identify if the solution will handle personally identifiable information (PII). If the data includes PII, then privacy laws such as the Privacy Act of 1974 in the US and the General Data Protection Regulation (GDPR) in the European Union (UE) are applicable (EU, 2018; US_Congress, 1974). In case PII is required, one must mitigate risks of unauthorized collection, storage, processing, and dissemination. This section first presents relevant to the US DoD/DA privacy legislation, policies, and regulations and then briefly introduces GDPR. The last section discusses the unique privacy incompatibilities with big data analytics.

3.1 The US DoD/DA Privacy Landscape

Figure 1 shows elements of DoD privacy and information security landscape with a focus on the DA branch. It includes privacy laws, policies, and regulations and their overlap with those relevant to IT security as a prerequisite of privacy.

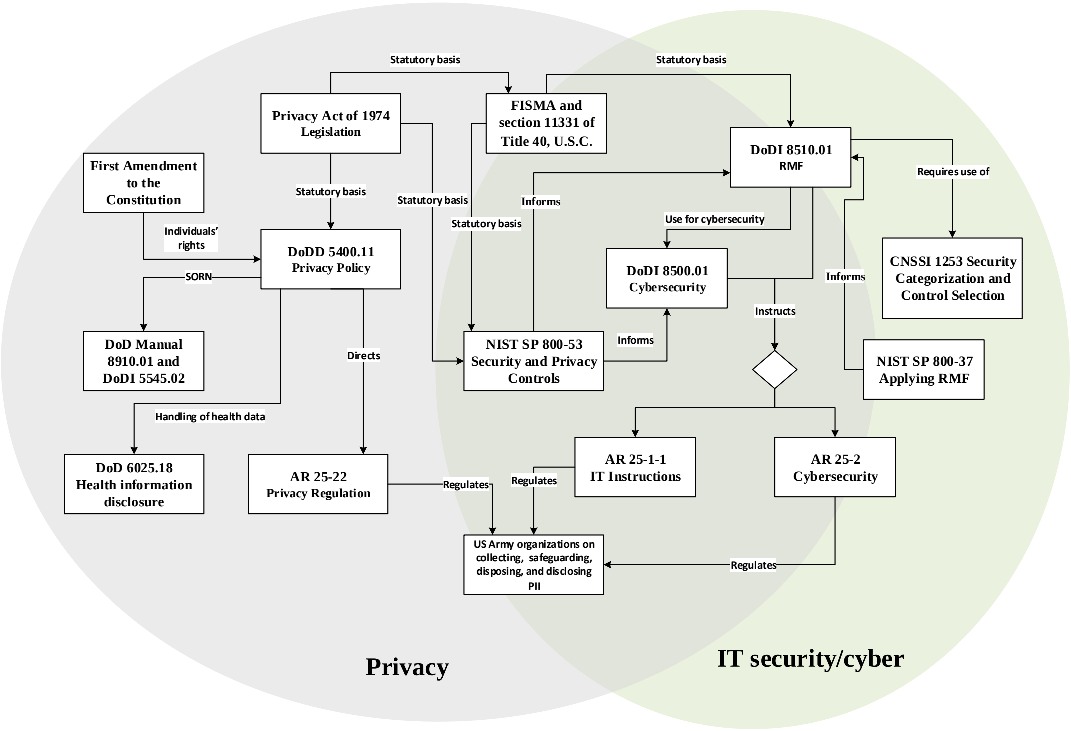

Figure 1. The intersection of the DoD/DA privacy and IT security/cyber landscape

DoDD 5400.11 (2014a) sets forth a policy to balance DoD's need to collect, use, maintain, or disseminate PII with the privacy rights of individuals following the principles captured in The Privacy Act of 1974 (US_Congress, 1974). As guided by DoD Manual 8910.01 (2014b) and DoDI 5545.02 (2008), before PII can be collected, a System of Record Notices (SORN) report should be submitted to Congress and the Office of Management and Budget (OMB) for approval and published in the Federal Register (FR). SORN should specify the limited use of PII. It is maintained by Records Management and Declassification Agency (RMDA) providing info to the public about all PII collected and maintained by DA organizations (DA, 2019b). The retention and disposal period of PII military records varies, but in the case of the DA PII, they are often kept for several years after members separate. The DA records are regulated by AR 25-400-2 (2007). As stated in AR 25-22 (2016), Privacy Impact Assessment (PIA) is mandated by the E-Gov Act of 2002, section 208 (2002), also known as the Federal Information Security Management Act

(FISMA), for all new and significantly altered information systems and data collections. PIA describes administrative, technical, and physical safeguards to assist in ensuring the integrity and security of PII.

NIST 800-53 (2013) provides security and privacy controls for federal information systems. Privacy controls are divided here into eight main groups: 1) authority to collect and purpose, 2) accountability, audit, and risk management, 3) data quality and integrity, 4) data minimization and retention, 5) individual retention and redress, 6) security, 7) transparency and 8) use limitation. It promotes automating privacy controls and advocates building privacy controls into system design and development. NIST 800-53 informs both RMF in DoDI 8510.01 (2014) and cybersecurity in DoDI 8500.01 (2014c). Army Regulation (AR) 25-22 (DA, 2016) is directed by DoDD 5400.11 (2014a) and provides policies and procedures that govern collecting, safeguarding, and disclosing PII within the DA. It lists the main privacy principles that should be adopted by the DA and operationalizes controls proposed in NIST 800-53 (2013). For instance, transparency and accountability to the individuals regarding their PII are promoted with the notice and consent. Organizations within the DA should state a specific purpose for using PII and, accordingly, collect only the necessary data. Individuals should be provided an appropriate mechanism to access and make corrections of their PII. DA organizations should secure PII against loss, unauthorized access and use, destruction, modification, and unauthorized disclosure.

3.2 Privacy in the EU

GDPR is the privacy legislation that provides a new legal basis supporting privacy preservation within the EU since May of 2018. It applies to organizations with operations located within or outside of the EU, if an organization is involved in processing PII that belong to EU citizens. The important reason US organizations should be aware of GDPR is the risk of being fined for non-compliance, which can result in fines up to €20 million or up to 4% of the total worldwide annual turnover, whichever is greater. According to GDPR, personal data must be protected, including names, addresses and images of individuals and special categories including data revealing, political opinions and religious and philosophical beliefs, biometric data, health data, genetic data, data concerning sexual orientation, data revealing trade union membership, and data revealing racial or ethnic origin. In GDPR parlance, individuals making decisions about the purpose and manner of data processing are called data controllers, while those involved in data processing are called data processors. For instance, a defense contractor that has employees operating in the EU would be a data controller, while an HR subcontractor processing compensations would be called a data processor. Data controllers should have the overall oversight over the processing of PII but both controllers and processors share responsibility for compliance with GDPR. Please refer to 99 Articles of GDPR for details (EU, 2016b).

The EU-US Privacy Shield applies to any EU data subjects whose PII has been transferred from EU to the US by organizations that self-certified their adherence to the privacy principles agreed upon with the Department of Commerce (Weiss & Archick, 2016). Main principles of the EU-US Privacy Shield include purpose, consent, data limitation and integrity, requirement for organizations to make their privacy policies public, choice principle, security principle, data subject access principle, recourse principle, accountability for onward transfer principle, and enforcement and liability principle. Adherence to the principles is limited to the extent necessary to meet national security, public interest or law enforcement requirements (EU, 2016a, pp. 14-15).

3.3 Privacy Incompatibilities in Big Data Analytics

This section offers a discussion on the incompatibilities of privacy laws and regulations with big data analytics. Privacy principles such as lawfulness, consent, purpose limitation, data minimization, transparency and openness, individual rights, IT security, accountability, and data protection by design were broadly applicable before the advent of big data (Danezis et al., 2015). Zarsky (2016) listed incompatibilities between GDRP and big data, i.e., purpose limitation, data minimization, special categories, and automated decisions. It appears that several of the incompatibilities listed above are also applicable in the context of the Privacy Act of 1974 and DoD/DA policies and regulations, e.g., (DA, 2016).

US federal laws and military policies and regulations list the principle of purpose limitation (DA, 2016; DoD, 2014a; NIST, 2013; US_Congress, 1974). On the other hand, big data attitude is quite the opposite, i.e., to collect data without clearly identified purpose and explore it for insights. First, the purpose limitation does not clearly state how to determine if the stated purpose is specific enough so more objective criteria to evaluate this requirement should be introduced. Second, projects involving big data often evolve by the virtue of realized opportunities to generate new insights. NIST (2013) privacy control IP-1 requires an organization to obtain consent from data subjects before initial collection and each time when considering the reuse of PII. Every time the purpose of processing data changes, the individuals should be informed about this fact to renew their consent. When data are merged or acquired, data protection obligations follow the data, and one should ensure that data is used only according to the original privacy consent. The risks of re-identification are behind the logic that new authorizations to process data would be required if an organization would like to combine data with other datasets. For instance, if military unit data with PII was collected by different organizations, e.g., at multiple training locations and training phases, each organization interested to use the data not originally specified would have to obtain a separate consent. That means that each purpose of every organization should be identified before the data was collected, which is highly unlikely in the case of big data. This, without some automatic solution or clear regulation to exempt similar cases, would slow down big data projects and likely involve too much burden on service members regarding the consent requirement. Data subjects should be able to access and check which organizations keep and use their PII and find out specifically what and when was disclosed. According to AR 25-22 individuals should be able to review and correct PII collected on them. If the big data system does not provide efficient ways to process these requests, the system may bottleneck resulting in data errors degrading quality.

Data minimization and retention controls (DM-1, DM-2, and DM-3) advocate organizations to minimize the amount of stored PII data (NIST, 2013), which also aligns with the DA (2016) stance. To implement these controls one must identify minimum relevant personal data to be collected and determine how long the data will be retained for a specified purpose, which is highly problematic in big data because it contradicts the first principle of big data, i.e., ‘volume’. The value of big data unveils from accumulating large datasets and it is difficult to determine upfront when one would have enough data to achieve a successful outcome (Soria-Comas & Domingo-Ferrer, 2016). The benefit from permanent retention of all data cannot be anticipated to suggest a correct retention period, so when a new purpose becomes apparent all the useful data may be long gone. Big data technologies spur the desire for keeping data indefinitely, reusing it for different purposes, and combining it with additional data.

Big data analytics often involve higher opacity because of many data streams from different sources, complicated algorithms processing the data, multiple storage locations, and multiple data consumers with different data aggregation needs. In the complex big data system, things can more easily be overlooked leading to privacy data breaches or insufficient anonymization leading to re-identification of individuals. Big data technologies continue their fast evolution and the full impact on IT security should not be overlooked. Entry points like backdoors, trapdoors, and rogue certificates can bring big data system down, cause security breaches, undermine encryption, and grant access to system assets and communication links. Interception of the communication between big data applications could compromise data confidentiality and integrity. Unsecured API can enable injection attacks. Technical implementation of an efficient accountability system allowing for enforcing privacy compliance in big data can be challenging because of technical interoperability issues (D'Acquisto et al., 2015). Vulnerabilities to security emerge due to the interoperability and incompatibility limitations between infrastructures, platforms, devices, applications, and data formats. Information leakage in big data may happen due to the loss of cloud console control. Data source filtering and validation in the big data system pose challenges because they are collected from many sites and hardware devices, some of which may be of unknown trustworthiness. The replication of data in big data storage may introduce additional vulnerabilities. Outsourcing big data computations across different security-level tiers located across different continents, countries, and regions introduce new types of breach, leakage and degradation threats. The data in big data technologies use links required for parallel data processing. This additional information may introduce data leakages and breaches. Non-SQL databases, e.g., Cassandra, introduce different data modeling principles. Because it lacks capabilities to join tables users must create different tables suitable for different queries, and the same data is stored multiple times using different schemas.

4. PRIVACY-PRESERVING COMPONENTS

It is a good practice to use IT services from companies certified in big data privacy. It is important to adopt a Service Level Agreement (SLA) adequate for big data to include proper resource isolation, access control, purpose limitation, data subjects’ consent, and exit strategies. Mont and Thyne (2006) proposed a way to automate the enforcement of privacy within an enterprise in a systemic way. Privacy by Design (PBD) in big data (D'Acquisto et al., 2015) and NIST’s Big Data Interoperability Framework: Volume 4 (Chang, 2018) provide an overview of privacy-preserving components in big data. Figure 2 displays an extended view of privacy preservation components proposed based on adapted NIST Big Data Reference Architecture (Chang, 2018). This section discusses selected privacy preservation components applicable to big data systems such as anonymization, privacy notice, privacy consent, personal data stores, sticky policies, and enhanced IT security.

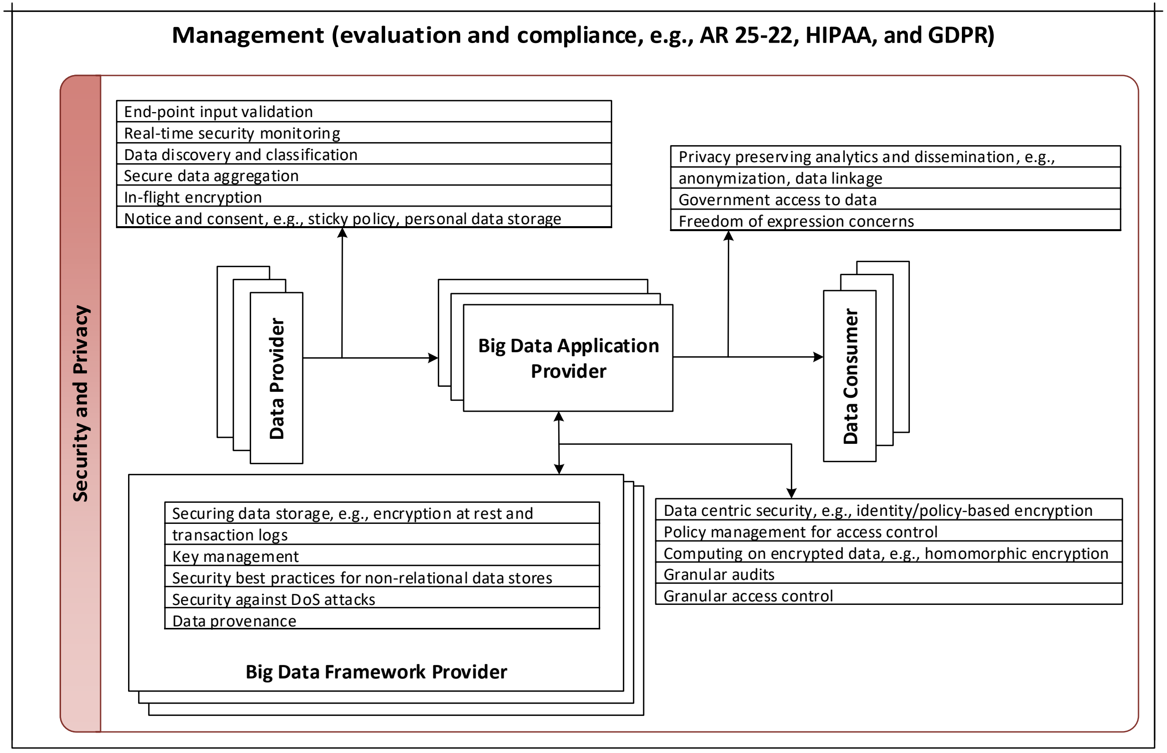
Management (evaluation and compliance, e.g., AR 25-22, HIPAA, and GDPR)
Security and Privacy
End-point input validation
Real-time security monitoring
Data discovery and classification
Secure data aggregation
In-flight encryption
Notice and consent, e.g., sticky policy, personal data storage
Privacy preserving analytics and dissemination, e.g.,
anonymization, data linkage
Government access to data
Freedom of expression concerns
Data Provider
Big Data Application Provider
Data Consumer
Securing data storage, e.g., encryption at rest and
transaction logs
Key management
Security best practices for non-relational data stores
Security against DoS attacks
Data provenance
Big Data Framework Provider
Data centric security, e.g., identity/policy-based encryption
Policy management for access control
Computing on encrypted data, e.g., homomorphic encryption
Granular audits
Granular access control

Figure 2. Privacy-preserving components adopted from (Chang, 2018)

4.1 Anonymization

The purpose of anonymization is to mitigate the risk of disclosure and loss of PII. Multiple methods can be used to anonymize data, for instance, data masking/removal, pseudo-anonymization, aggregation, and banding (Graham, 2012). All or selected PII can be removed. For instance, during the removal of selected identifier attributes, i.e., information that uniquely and directly identifies individual, e.g., name and address should be removed. On the other hand, quasi-identifiers, e.g., gender, age or date of birth would be kept. Deciding about the sensitivity of data would depend on the context, which can be subjective, hence, the risk assessment should be conducted.

Data aggregation is used to display data as totals, averages, and partitions; therefore, no individual data are distinguishable or shown. When an intruder may identify someone based on a unique combination of sensitive attributes specific data fields can be removed, marked missing, made less specific, swapped across the rows, or otherwise altered to reduce the information content. Probabilistic methods, e.g., Markov matrix, adding noise using probability distributions, or resampling data can be used to alter datasets. Depending on the method used altered data may or may not maintain statistical resemblance and correlations between variables as compared to the original data. If the details about the probabilistic method used were known by the intruders PII could be restored. A released dataset is said to be of K-anonymity if the quasi-identifier values about each person do not allow to distinguish an individual with confidence greater than 1/K (Li, Li, & Venkatasubramanian, 2007). The unsorted matching attack, temporal attack, and complementary release attack can be used to re-identify K-anonymous data (Sweeney, 2002). When sensitive values in a group exhibit homogeneity L-diversity approach can be used, ensuring that at least L distinct values for the sensitive attribute exist in each equivalence class by inserting fictitious data. L-diversity cannot prevent sensitive attribute disclosure against skewness and similarity attacks (Li et al., 2007). T-closeness treats each sensitive attribute distinctly by ensuring that "…the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold T" (Li et al., 2007).

Pseudo-anonymization uses coded reference or pseudonym attached to a record that allows for tracking individuals without disclosing their identities. This idea is extended under term data linkage, but other terms such as record matching, entity resolution, and merge-purge are also used with a synonymous meaning (Harron, 2016). Data is hidden under a reference code that requires a key to be linked to an individual allowing for tracking individuals as part of longitudinal studies. Record linkage is a process that involves combining data, from different sources, that belongs to the same individuals. Deterministic and probabilistic methods can be used to link the datasets. Depending on requirements, the linkage may be conducted in-house or by a trusted third party. Figure 3 presents a concept of the data linkage, where data generated by data providers A and B are used by a research site based on an anonymous key generated by a trusted third party (Graham, 2012). The trusted party does not receive personal attributes from data providers but receives identifiers and serial record IDs used to create match key mapping both datasets. The research site can merge personal attributes using this anonymous match key and record IDs. When using this method, it is important to evaluate the impact of data linkage quality on study results related to linkage errors.

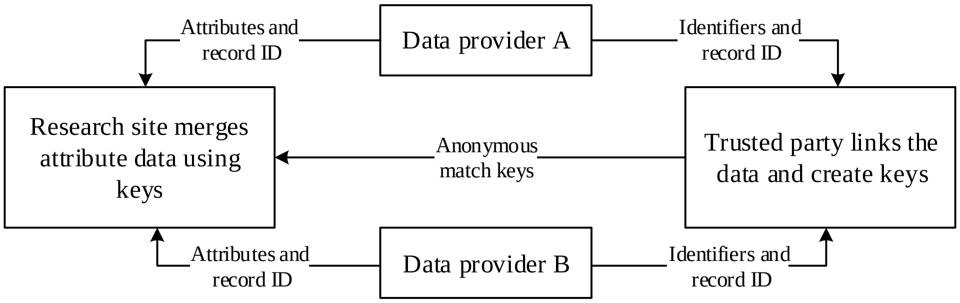
Attributes and record ID
Data provider A
Identifiers and record ID
Research site merges attribute data using keys
Anonymous match keys
Trusted party links the data and create keys
Attributes and record ID
Data provider B
Identifiers and record ID

Figure 3. Concept of data linkage by a trusted party

Encryption when transferring data between data providers, a research site and a trusted party is required. Encryption algorithms can be reversible or not reversible. In cases of non-reversible encryption, it is not possible to return to the original identifiers because non-reversibly encrypted identifiers disallow the research site to manually evaluate the quality of created linkages.

Differential privacy is a probabilistic anonymization method gaining momentum in its application to the big data privacy problems (Greenberger, 2016). User's queries are processed via the privacy guard aware of all earlier queries. It assumes that the attacker has the maximum background information about the subject. Based on the evaluated privacy risk privacy guard applies distortion to query results. Differential privacy involves injecting a controllable amount of noise into the answers to data queries at the individual level at the cost of data utility, which can be controlled depending on the desired balance between privacy and utility (Dwork, 2008). Differential privacy mechanisms can prevent users' privacy leakage but inevitably cause data utility loss. Moreover, this approach can be vulnerable when multiple data sets are correlated (Wu, Wu, Khan, Ni, & Dou, 2018). Jung, Park, and Park (2014) proposed a privacy-preserving algorithm that uses association rule mining in Hadoop's Map Reduce allowing to minimize privacy violation without utility degradation. In this approach, the risk of association rule leakage is very small at the cost of incurred additional computation. X. Zhang et al. (2018) proposed scalable multidimensional anonymization based on MapReduce but also showed its applicability to differential privacy. The multidimensional anonymization allows for balancing data obfuscation and utility when handling big data in the cloud.

Wang, Hu, Sun, and Huang (2018) summarized privacy preservation in Location-Based Services (LBS). LBS may disclose location/trajectory which can give up sensitive information and user's identity. For instance, edge/fog computing produces privacy leakage as mobile devices can be tracked without user awareness (Vaquero et al., 2019). Privacy protection models for LBS data include K-anonymity (and its derivatives) adopted to LBS based on grouping of K-1 users to achieve location perturbation of a certain user, differential privacy (Yin, Xi, Sun, & Wang, 2018), dummy-location selection (DLS) entropy-based algorithm, private information retrieval (PIR) adapted to secure the location privacy, dynamic grid system (DGS), and other algorithms (Wang et al., 2018).

4.2 Privacy Notice and Consent

According to GDPR privacy regulations, data subjects must agree for processing of their personal data based on privacy notice and consent, which should include all required information to make informed and fair consent decision, e.g., data controller's identity, purpose of data processing, involvement of automated decision-making including profiling and its potential consequences for the data subject (Denham, 2017). This can be problematic because it is not easy to briefly explain advanced analytics to people without a data science background. More innovative approaches to presenting privacy notices are needed. For instance, making simple to understand cartoon movie about data linking (NatCen, 2018) by the data processing organizations is a step in the right direction. Similarly, improved automated consent solutions should be considered. For instance, a sticky policy concept was introduced by Karjoth, Schunter, and Waidner (2002), in which PII is tagged with metadata that specifies individuals' preferences and possibly other organizational rules. It could communicate to data controllers subjects' acceptable purposes for the processing of personal data, type, and scope of permitted disclosure, and other rules (D'Acquisto et al., 2015). The implementation of sticky policy does not prevent from modification of the policies (Tang, 2008), hence, the use of encryption over the sticky policies is needed with the ongoing research in this area (D'Acquisto et al., 2015).

Increased control over the use of PII by the subjects can be facilitated by personal data stores (PDS). For instance, Chaudhry et al. (2015) presented a concept of PDS called Databox. PDS facilitates the gathering, storing, updating, correcting, analyzing, and sharing personal data. Data subjects have full control over their data even after the release (D'Acquisto et al., 2015). PDS would store personal data of individuals and could make it available to organizations on their behalf based on the specified detailed preferences. This could also mitigate the problem with privacy consent and notice due to the repurposing of data and when another organization would like to process the data. On the other hand, if individuals would prefer more direct control over their PII then data controllers would have to obtain permission to access personal data directly from the data subjects.

4.3 IT Security in Big Data

Safeguarding PII is closely related to national security (DA, 2016). Although privacy is not synonymous with security a poor cybersecurity policy is a threat to privacy (PCAST, 2014). It is very difficult to handle privacy at a sufficient level without good security practices (Jain, Gyanchandani, & Khare, 2016). It is recommended to encrypt data in flight and at rest to ensure data confidentiality and integrity (Naydenov, Liveri, Dupre, Chalvatzi, & Skouloudi, 2015). On the other hand, this increases computational overhead and potentially inferior utility due to the increased processing time of queries. Confidential data should be separated from insensitive data based on its classification to improve filtering and encryption. It is recommended to use legitimate sources and purchase authentic devices that adhere to desired security standards. Also, to ensure the validation of endpoint sources, it is recommended to use tamper-resistant devices with authentication capabilities. The use of on-premise trusted big data infrastructure lowers threats against the interception of information. Big data software must be tested for common vulnerabilities such as back doors, default credentials and weak authentication methods. Project Rhino as part of the Apache Hadoop ecosystem aims at better encryption support for Hadoop Core and its integration across other elements of the ecosystem such as HDFS, MapReduce and Hive (Hadoop, 2015).

4.3.1 Encrypting Big Data

Encryption is a strong security measure that transforms data into the format viewable only to the authorized parties. Encryption can be used to secure data at rest and data in flight. Popular encryption schemes include symmetric, asymmetric, hybrid, and hashing algorithms. Because of its efficiency and relatively high level of security symmetric encryption schemes such as Advanced Encryption Standard (AES) (Daemen & Rijmen, 1999) are widely used in big data and cloud environments (D'Acquisto et al., 2015). It uses the same key for both encryption and decryption. RSA is an asymmetric cryptographic algorithm with two keys: one public to encrypt and one private to decrypt data. Because RSA is more computationally expensive it is used mainly for digital signatures or distributing secret keys. Hybrid encryption schemes combine asymmetric public-key encryption, e.g., RSA because of its quality key management with symmetric, e.g., AES because of its speed and efficiency. Secure hashing algorithm (SHA) is a family of cryptographic hash functions allowing for one-way encryption that can be run on digital data to ensure its integrity by comparing computed “hash” value to known and expected hash value. It gives no option to decrypt data. The use of strong hashing functions such as the SHA-2 family is advised. For instance, SHA-256 or SHA 512, are used in security applications with TLS and SSL protocols. They are also used for ensuring message integrity and storing passwords in databases. In practice, AES, RSA, and SHA can be used together to accomplish required encrypting functionality.

Big data technologies often require more granular data sharing policies among different user groups and flexibility in accessing encrypted data. For instance, attribute-based encryption combines access control with public-key cryptography allowing for sharing data among different user groups without compromising privacy (Goyal, Pandey, Sahai, & Waters, 2006). Functional encryption offers more flexibility as it is possible to control access to functionalities of the encrypted data allowing to decrypt only certain data for specific processing as prescribed by the secret key (Boneh, Sahai, & Waters, 2011). For instance, end-users via big data applications could only conduct a priori specified class of processing keeping the rest of the data confidential. Unfortunately, the practical implementation of functional encryption suffers from its performance and more work in this area is needed (D'Acquisto et al., 2015). User privacy and data security could be enhanced by encrypting sensitive data before outsourcing them to the cloud. Searchable encryption is a functionality that could allow searching over encrypted data without disclosure of personal information (Kamara, 2015). Property Preserving Encryption (PPE) encrypts data but preserves some properties related to, e.g., equity, order, and orthogonality. Searchable symmetric encryption (SSE) allows to search over encrypted data and can be used when the same entity encrypts and searches over the data (Curtmola, Garay, Kamara, & Ostrovsky, 2011). Public Key Searchable encryption (PEKS) can be used when an entity that encrypts data is different from the one searching over it (D'Acquisto et al., 2015). Future research on privacy-preserving algorithms for big data is ongoing. Homomorphic encryption allows computation over the encrypted data (Kamara, 2015). It is still not matured enough to be used in practical applications (Gentry et al., 2013), but it could in the future eliminate cloud providers from the privacy equation, which means only data providers would have visibility of the data. Zhang, Zhang, and Ma (2018) proposed a privacy-preserving low computational overhead conjunctive keyword search scheme over encrypted cloud data.

4.3.2 Access Control: Authentication and Authorization for Big Data

Access control ensures that only authorized users and processes can access data. To ensure that big data queries are executed by authorized users and processes only, it is imperative to implement user authentication and authorization. Classical role-based access control (RBAC) is still an efficient way to automate granting permissions to users based on their roles, but it can be difficult to administer in a large big data architecture serving many communities with many roles. Currently, HBase as part of the Hadoop ecosystem enables high granularity controls (Hadoop, 2015). Access control lists (ACLs) composed of users and groups are supported by HDFS, Hadoop Core, ZooKeeper, and HBase at the RPC layer via SASL, but they do not support attributes such as group membership, classification level, organizational identity, and user-defined attributes. The work on token-based authentication framework decoupling internal user and service authentication from external mechanisms supporting it aims to address these limitations (Hadoop, 2015). Granular access rules allowing to set access controls at the table, column, and even cell levels, which are very important when dealing with sensitive information in big data environments. Attribute-based access control (ABAC) holds promise for the big data access controls by providing finer-granularity, flexibility, and partial authentication based on the user having specific attributes and other data that create a context-aware decision regarding individual requests for access (Cavoukian, Chibba, Williamson, & Ferguson, 2015).

4.3.3 Data Provenance and Monitoring for Big Data

Data provenance provides background knowledge allowing for interpretation and confirmation of appropriate data use within specified contexts (Ram & Liu, 2008). The provenance in big data provides audit-compliance-accountability, which allows confirming the origin and authenticity of the data, qualifies assertions, tracks how data are being processed, and explains unexpected results (D'Acquisto et al., 2015). Because provenance data is not aggregated it requires a high level of protection against disclosure. Monitoring to protect against Distributed Denial of Service (DDoS) attacks during distributed data processing can be done by using ingress filtering, rate limiting, and network traffic monitoring and analysis, e.g., using Wireshark and SPLUNK. Big data distributed computation nodes, distributed databases and applications should be monitored using logs and detection applications. A potential solution to the problem resides in access control and query systems (Davidson et al., 2011), e.g., provenance-based access control (Park, Nguyen, & Sandhu, 2012). Currently, most databases allow for automated scanning of data and logs. Ongoing work in Project Rhino aims at defining consistent log formats and developing audit log processing tools capable of transforming logs into different industry-standard forms so they could be used for compliance verification activating anti-policy violation mechanisms (Hadoop, 2015).

4.4 Privacy Evaluation

Privacy evaluation should be used to verify implemented privacy-preserving mechanisms. Approaches to evaluating the protection of privacy rights include privacy impact assessment (PIA), privacy maturity models (AICPA/CICA, 2011), privacy certifications, IT security testing, testing against re-identification attacks (Shamsi & Khojaye, 2018), and algorithmic auditing.

5. RECOMMENDATIONS AND FUTURE WORK

Harnessing the value of big data could facilitate the creation of innovative solutions, but this requires elimination or at least mitigation of burdens to conduct research using large military datasets. Different ways to handle privacy requirements are possible and some amendments to military regulations regarding big data use cases can help with directing the change. On the battlefield, decision-making supported by big data analytics could make a difference between life and death. Regulation-based privacy workarounds infused by the DA to balance privacy laws and policies allow partially for taking advantage of new technologies within specific military communities. Listed by AR 25-22 blanket routine uses simplify handling of PII for counterintelligence and terrorism scenarios. All properly classified data and data used for statistical purposes that do not make decisions about the rights, benefits, and entitlements of individuals are also exempted. DA cybersecurity regulation instructs that "…users will have no expectations as to the privacy or confidentiality of any electronic communication, including minor incidental personal uses. The Army reserves and will exercise the right to access, intercept, inspect, record, and disclose any and all electronic communications on Army IT, including minor incidental personal uses, at any time, with or without notice to anyone, unless prohibited by law or privilege" (DA, 2019a, p. 30). Similar rules for other military communities could enable big data analytics, but they could also increase privacy risks. For instance, additional exempts on training uses and mission data could be added to AR 25-22. Moreover, since privacy and IT/cyber regulations overlap, a single integrated regulation should be considered to avoid any ambiguities.

NIST 800.53 privacy control AR-7 advocates the design of IT systems with automated privacy controls, which is similar to the PBD concept in GDPR. The DoD/DA privacy policies and regulations should follow PBD guidance on efficient handling of PII within big data landscape enabling the development of appropriate IT infrastructure with privacy-preserving data discovery and data linkage services. Various technological advances protecting PII in big data could be implemented. DoD organizations may benefit from expanding common authoritative personnel data by storing all training and mission data about individual service members in their "avatar PDS" representations. The purpose, privacy consent, and privacy notice functions in big data technologies create requirements to enable interoperability between big data controllers and SORN. For instance, an automatic system of records could be synchronized with data services so that military managers can adjust PII rules in near real-time, which can take immediate effect on datasets availability. This should provide dynamic tractability off PII controlled by organizations and allow linking stated purpose for PII processing with automatic management of privacy notice and consent.

NIST 800.53 privacy controls regarding the minimization and retention of data should be revisited, especially in the context of testing, training, and research (DM-3). These controls seem to contradict the big data principle and could mislead DA policies and regulations. Moreover, one should be able to automatically link big data analysis results back to benefit data subjects, e.g., by using PDS stores for individual military members linked with DoD research enterprise.

NIST 800.53 privacy control DM-1 only mentions anonymization, but various anonymization techniques could find applicability in the context of data exchange with coalition partners, hence, more adequate guidance is needed. DoD privacy agreements specifying privacy protection techniques like anonymization between US and coalition partners relevant to training and test events would be useful to harness the power of big data. Moreover, adequate interoperability standards are needed to exchange real-time big data. In this context research is needed to develop solutions that allow exchanging big data during training exercises, possibly including the anonymization of PII algorithms and CDS in cases when data is exchanged across domains at different classification levels. Research on a framework for combining PBD, security by design (SBD) and interoperability by design (IBD) concepts is needed in the context of developing big data systems supporting military organizations across M&S training, test, experimentation and analysis communities.

To address big data privacy concerns more work is needed to analyze potential solutions. Yu (2016) highlighted needs for work on better measurements of privacy, theoretical frameworks of privacy, scalability, and efficiency of privacy algorithms, and algorithms to deal with heterogeneous data sources in the context of privacy preservation. Commercial implementations of big data technologies are often offered as cloud-based solutions. Although this paper focused on implications arising from current privacy policy and regulations relevant future research should consider implications arising from using and combining different types of big data and cloud technologies. Mell and Grance (2011) categorize clouds based on service models used as software-as-a-service, platform-as-a-service, and infrastructure-as-a-service, and based on the deployment model as private, community, public and hybrid. Multi-cloud solutions may be needed for large enterprises due to location, disaster back-up, and law constraints (Petcu, 2013). Kratzke (2018) discussed trends in cloud technology architectures over the last decade with a direction towards making computing more resource-efficient starting from virtualized bare-metal machines, through more fine-grained container-as-a-service, to function-as-a-service. Pahl, Brogi, Soldani, and Jamshidi (2017) presented a systematic review of cloud container technologies. Lynn, Rosati, Lejeune, and Emeakaroha (2017) provided a review of serverless computing and available serverless platforms. A major opportunity for high impact research includes quantum computing for unconditional privacy-preserving.

REFERENCES

AICPA/CICA. (2011). *Privacy Maturity Model* Retrieved from http://www.cil.cnrs.fr/CIL/IMG/pdf/10-229_aicpa_cica_privacy_maturity_model_finalebook_revised.pdf

Balaban, M. A. (2015). The Support of the Acquisition Defense Governance Using Constructive M&S. *The 2015 ITEA Test Technology Review (TTR)*.

Balaban, M. A., Mastaglio, T. W., Sokolowski, J., & Ezell, B. (2014). *Exploration of Soldier Morale Using Multi-Method Simulation Approach*. Paper presented at the ITSEC, Orlando, FL.

Balaban, M. A., & Mielniczek, P. (2018). *Hybrid conflict modeling*. Paper presented at the Proceedings of the 2018 Winter Simulation Conference.

Boneh, D., Sahai, A., & Waters, B. (2011). *Functional encryption: Definitions and challenges*. Paper presented at the Theory of Cryptography Conference.

Cavoukian, A., Chibba, M., Williamson, G., & Ferguson, A. (2015). The Importance of ABAC: Attribute-Based Access Control to Big Data: Privacy and Context. *Privacy and Big Data Institute, Ryerson University, Toronto, Canada*.

Chang, W. L. (2018). *NIST Big Data Interoperability Framework: Volume 4, Security and Privacy*. Retrieved from https://doi.org/10.6028/NIST.SP.1500-4r1

Chaudhry, A., Crowcroft, J., Howard, H., Madhavapeddy, A., Mortier, R., Haddadi, H., & McAuley, D. (2015). *Personal data: thinking inside the box*. Paper presented at the Proceedings of the fifth decennial Aarhus conference on critical alternatives.

Clinger, B., Sipes, C., Blomberg, N., Burch, R., Lanman, J., & Todd, J. (2018). *Automating the Training Feedback Paradigm with Intelligent After Action Review*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Craven, P. L., Oden, K. B., Landers, K. J., Shah, A. J., & Shah, J. A. (2018). *Man-Machine Interoperation in Training for Offensive Counter Air Missions*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Curtmola, R., Garay, J., Kamara, S., & Ostrovsky, R. (2011). Searchable symmetric encryption: improved definitions and efficient constructions. *Journal of Computer Security, 19*(5), 895-934.

D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y.-A., & Bourka, A. (2015). Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics. *arXiv preprint arXiv:1512.06000*.

DA. (2007). Regulation 25-400-2 "Information Management: The Army Records Information Management System (ARIMS)". *Headquarters, Department of the Army, Washington, DC, 2*.

The Army Privacy Program AR 25-22, (2016).

Army Cybersecurity, (2019a).

DA. (2019b). US Army System of Record Notices (SORN). Retrieved from https://www.rmda.army.mil/privacy/sorns/

Daemen, J., & Rijmen, V. (1999). AES Proposal: Rijndael.

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Metayer, D. L., Tirtea, R., & Schiffner, S. (2015). Privacy and data protection by design-from policy to engineering. *arXiv preprint arXiv:1501.03726*.

Danner, B., & Djahandari, K. (2008). *Cross Domain Solution Policy, Management, and Technical Challenges*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Davidson, S. B., Khanna, S., Roy, S., Stoyanovich, J., Tannen, V., & Chen, Y. (2011). *On provenance and privacy*. Paper presented at the Proceedings of the 14th International Conference on Database Theory.

Denham, E. (2017). *Big data, artificial intelligence, machine learning and data protection*. UK-ICO.

DoD Policy 5545.02 for Congressional Authorization and Appropriations Reporting Requirements, (2008).

DoD 5400.11 Privacy Program Directive, (2014a).

DoD Information Collections 8910.01, 30 C.F.R. (2014b).

DoDI 8500.01 Cybersecurity, (2014c).

DoDI 8510.01 Risk Management Framework (RMF) for DoD Information Technology (IT), (2014).

Durlach, P. J. (2018). *Can we talk? Semantic Interoperability and the Synthetic Training Environment*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Dwork, C. (2008). *Differential privacy: A survey of results.* Paper presented at the International Conference on Theory and Applications of Models of Computation.

EU. (2016a). Commission Implementing Decision (EU) 2016/1250 of 12 July 2016. *Official Journal of the European Union, L 207/1*.

General Data Protection Regulation, (2016b).

EU. (2018). General Data Protection Regulation.   Retrieved from https://gdpr-info.eu/

Freeman, J., Nicholson, D., Squire, P., & Bolton, A. (2014). *Data & analytics tools for agile training & readiness assessment.* Paper presented at the Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).

Gentry, C., Goldman, K. A., Halevi, S., Julta, C., Raykova, M., & Wichs, D. (2013). *Optimizing ORAM and using it efficiently for secure computation.* Paper presented at the International Symposium on Privacy Enhancing Technologies Symposium.

Goyal, V., Pandey, O., Sahai, A., & Waters, B. (2006). *Attribute-based encryption for fine-grained access control of encrypted data.* Paper presented at the Proceedings of the 13th ACM Conference on Computer and Communications Security.

Graham, C. (2012). *Anonymisation: managing data protection risk code of practice*.

Greenberger, A. (2016). Apple's 'Differential Privacy' is About Collecting your Data—but not your Data.   Retrieved from https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/

Hadoop. (2015). Project Rhino.   Retrieved from https://github.com/intel-hadoop/project-rhino/

Harder, B., Blais, C., & Balogh, I. (2017). *Conceptual framework for an automated battle planning system in combat simulations*. Paper presented at the Proceedings of the 2017 Winter Simulation Conference, Las Vegas, Nevada.

Harron, K. (2016). Introduction to Data Linkage.   Retrieved from https://adrn.ac.uk/policies-procedures/protecting-privacy/?Data-linkage

Huber, Z., Winslow, B., Chiang, J., & Aziz, A. (2018). *Objective Stress Monitoring for Live Training Exercises*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data, 3*(1), 25. doi:10.1186/s40537-016-0059-y

Jensen, R., & Ramachandran, S. (2018). *Data-driven Training Development: Deriving Performance Constraints from Operational Examples*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Jung, K., Park, S., & Park, S. (2014). *Hiding a Needle in a Haystack: Privacy Preserving Apriori Algorithm in MapReduce Framework.* Paper presented at the Proceedings of the First International Workshop on Privacy and Secuirty of Big Data.

Kamara, S. (2015). Encrypted search. *XRDS: Crossroads, The ACM Magazine for Students, 21*(3), 30-34.

Karjoth, G., Schunter, M., & Waidner, M. (2002). *Platform for enterprise privacy practices: Privacy-enabled management of customer data.* Paper presented at the International Workshop on Privacy Enhancing Technologies.

Kratzke, N. (2018). A brief history of cloud application architectures. *Applied Sciences, 8*(8), 1368.

Li, N., Li, T., & Venkatasubramanian, S. (2007). *t-closeness: Privacy beyond k-anonymity and l-diversity.* Paper presented at the Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on.

Lynn, T., Rosati, P., Lejeune, A., & Emeakaroha, V. (2017). *A preliminary review of enterprise serverless cloud computing (function-as-a-service) platforms.* Paper presented at the 2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom).

Marrou, L., Glenn, C., & Nielsen, K. (2018). *Exploring Cloud-Based Terrain Generation Services*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.

Mont, M. C., & Thyne, R. (2006). *A systemic approach to automate privacy policy enforcement in enterprises.* Paper presented at the International Workshop on Privacy Enhancing Technologies.

N. S. Research (Producer). (2018). *About linking data* [Retrieved from https://www.youtube.com/watch?v=JRRRPUXQWbc

Naydenov, R., Liveri, D., Dupre, L., Chalvatzi, E., & Skouloudi, C. (2015). Big data security-good practices and recommendations on the security of big data systems. *European Union Agency for Network and Information Security (ENISA), Greece*.

NIST. (2013). 800-53 Rev 4, Security and Privacy Controls for Federal Information Systems and Organization.

Pahl, C., Brogi, A., Soldani, J., & Jamshidi, P. (2017). Cloud container technologies: a state-of-the-art review. *IEEE Transactions on Cloud Computing*.

Park, J., Nguyen, D., & Sandhu, R. (2012). *A provenance-based access control model.* Paper presented at the Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference on.

PCAST. (2014). *Big Data and Privacy: A Technological Perspective*.

Petcu, D. (2013). *Multi-Cloud: expectations and current approaches.* Paper presented at the Proceedings of the 2013 International Workshop on Multi-cloud Applications and Federated Clouds.

Ram, S., & Liu, J. (2008). A semiotics framework for analyzing data provenance research. *Journal of Computing Science and Engineering, 2*(3), 221-248.

Shamsi, J. A., & Khojaye, M. A. (2018). Understanding Privacy Violations in Big Data Systems. *IT Professional, 20*(3), 73-81.

Shanley, M. G., Crowley, J. C., Lewis, M. W., Leuschner, K. J., & Masi, R. (2007). *Supporting Training Strategies for Brigade Combat Teams Using Future Combat Systems (FCS) Technologies* (Vol. 538): Rand Corporation.

Soria-Comas, J., & Domingo-Ferrer, J. (2016). Big Data Privacy: Challenges to Privacy Principles and Models. *Data Science and Engineering, 1*(1), 21-28. doi:10.1007/s41019-015-0001-x

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10*(05), 557-570.

Tang, Q. (2008). On using encryption techniques to enhance sticky policies enforcement. *DIES, Faculty of EEMCS, University of Twente, The Netherlands*.

The Privacy Act of 1974, 16 C.F.R. (1974).

US_Congress. (2002). E-Gov Act of 2002, section 208.

Vaquero, L. M., Cuadrado, F., Elkhatib, Y., Bernal-Bernabe, J., Srirama, S. N., & Zhani, M. F. (2019). Research challenges in nextgen service orchestration. *Future Generation Computer Systems, 90*, 20-38.

Wang, S., Hu, Q., Sun, Y., & Huang, J. (2018). Privacy Preservation in Location-Based Services. *IEEE communications magazine, 56*(3), 134-140. doi:10.1109/MCOM.2018.1700288

Weiss, M. A., & Archick, K. (2016). US-EU data privacy: from safe harbor to privacy shield: Congressional Research Service.

Woodard, T., & Enloe, M. (2018). *Learning Applications for Modeling, Simulation, and Training*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Wu, X., Wu, T., Khan, M., Ni, Q., & Dou, W. (2018). Game Theory Based Correlated Privacy Preserving Analysis in Big Data. *IEEE Transactions on Big Data*, 1-1. doi:10.1109/TBDATA.2017.2701817

Yin, C., Xi, J., Sun, R., & Wang, J. (2018). Location Privacy Protection Based on Differential Privacy Strategy for Big Data in Industrial Internet of Things. *IEEE Transactions on Industrial Informatics, 14*(8), 3628-3636. doi:10.1109/TII.2017.2773646

Yu, S. (2016). Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data. *IEEE Access, 4*, 2751-2763. doi:10.1109/ACCESS.2016.2577036

Zarsky, T. Z. (2016). Incompatible: The GDPR in the Age of Big Data. *Seton Hall L. Rev., 47*, 995.

Zhang, L., Zhang, Y., & Ma, H. (2018). Privacy-Preserving and Dynamic Multi-Attribute Conjunctive Keyword Search Over Encrypted Cloud Data. *IEEE Access, 6*, 34214-34225. doi:10.1109/ACCESS.2018.2823718

Zhang, X., Qi, L., Dou, W., He, Q., Leckie, C., Ramamohanarao, K., & Salcic, Z. (2018). MRMondrian: Scalable Multidimensional Anonymisation for Big Data Privacy Preservation. *IEEE Transactions on Big Data*, 1-1. doi:10.1109/TBDATA.2017.2787661