EasyChair Preprint

№ 1938

A Framework for Implementing a Data Science Capability in a Military Intelligence System

Shaun Ball, Carien Van'T Wout and Rudolph Oosthuizen

November 11, 2019

24th ICCRTS

A Framework for Implementing a Data Science Capability in a Military Intelligence System

Topic
Track 3: Battlefields of the Future and the Internet of Intelligent Things

Authors

S.V. Ball
(CSIR, South Africa; SBall@csir.co.za)

C van't Wout
(CSIR, South Africa; CvtWout@csir.co.za)

Dr. R. Oosthuizen
(CSIR, South Africa; ROosthuizen@csir.co.za)

Point of Contact
S.V. Ball
(CSIR, South Africa; SBall@csir.co.za)

Name of Organization: Council for Scientific and Industrial Research
PO Box 395
Pretoria
0001
info@csir.co.za

A Framework for Implementing a Data Science Capability in a Military Intelligence System

S.V. Ball (Council for Scientific and Industrial Research)

M.C. van't Wout (Council for Scientific and Industrial Research)

Dr R. Oosthuizen (Council for Scientific and Industrial Research)

Abstract

Modern day conflicts give rise to complex problems that traditional military intelligence approaches and tools struggle to resolve. There is a need for prediction and/or forecasting in the military domain based on effective intelligence processing capabilities for pro-active measures as well as reactive responses. Intelligence processes and tools are becoming increasingly inadequate to support the decisions of commanders and other decision makers. The tsunami of data of the current age requires a system of new processing and analysis tools, with the supporting skills, to provide the intelligence required for making decisions about complex situations. The Internet of Battlefield Things (IoBT) and resulting Big Data is a reality for current and future military operations. The field of Data Science provides a foundation for processing and analytic tools, processes and skills. This paper assesses current literature on intelligence analysis and Big Data to define a framework that will guide the implementation of a Data Science Capability for modern military operations. The intelligence system is viewed from a sociotechnical system perspective to identify high-level requirements for implementation of a proposed Data Science framework for intelligence systems. The framework is derived from mapping the requirements of the traditional intelligence cycle to the various Data Science methods, tools and skills.

INTRODUCTION

Now the reason the enlightened prince and the wise general conquer the enemy whenever they move and their achievements surpass those of ordinary men, is foreknowledge. —Sun Tzu, The Art of War, 1963

Modern day conflicts give rise to complex problems that traditional military intelligence approaches and tools struggle to resolve. These intelligence processes and tools are becoming increasingly inadequate to support the decisions of commanders and other decision makers. The complexity stems from the current global interconnectedness that impact social, economic, and political changes (Smith 2016). Military decision makers require a prediction and/or forecasting capability based on effective intelligence processing for pro-active measures as well as improved reactive responses.

The world is witnessing an explosive growth in data being generated and communicated by people, systems, and networks. More data has been created in the past two years than in the entire previous history of mankind. By 2020, our digital universe of data will grow to 44 zettabytes (or 44 trillion gigabytes) which is ten times its size today (Suri et al. 2016, Baker & Henderson 2017).

In combination with the rising complexity of the problem space, the ever-growing Internet of Things (IoT) and social media platforms dramatically increase the information available for processing and analysis. Intelligence agencies require new tools, processes and skills to effectively analyse this abundance of information. Data science with Artificial Intelligence (AI) and other automated analysis tools can dredge through the gathered information to deliver the required intelligence products (Suri et al. 2016).

The field of data science is a wide area of study that includes big data, predictive analytics, cognitive computing, and machine learning. These technologies are currently being exploited for military applications, including surveillance, reconnaissance, intelligence analysis, command and control, threat evaluation, cyber security, and training (Svenmarck et al. 2018). Unfortunately, the intelligence community tend to resist new technologies (Byman 2016).

However, decision-making within combat operations can be characterised as having a "high regret" if the "wrong" decision is made, requiring a high degree of trust in the data science tools. Implementing of these tools into intelligence systems require a careful consideration of the expected difficulties. The complexity of this challenge is compounded by the need to consider military decision making in different combat scenarios, operating environments and levels of command. Intelligence agencies can progress beyond looking back at data to make projections. It is possible to drill deep into cause and effect, and determine the mathematical probability of future occurrences. This presents a transition from hindsight to foresight (Pearson et al. 2018).

This paper aims to provide an improved framework for the intelligence process by focussing on current best practices in data science. To achieve this aim, the paper will firstly clarify the concepts of military intelligence systems and the importance of intelligence for military decision-making. The challenges of current and previous approaches to producing military intelligence will then be investigated. This will be compared to data science's best practices to propose a framework that will guide the implementation of a Data Science Framework for modern military intelligence capabilities.

MILITARY INTELLIGENCE SYSTEMS

Attempts at gathering information for military planning is as old as war itself and spying is mentioned in the earliest historical accounts such as Homer's Iliad and the Bible, the network of spies and embassies used by the Romans, Chinese information gathering in theoretical documents, and accounts of reconnaissance by Hannibal and by Alexander the Great. As states and governments evolved over time, their intelligence capabilities evolved also into the highly complex and multi-faceted intelligence organisations of current day (Austin & Rankov 2002). Military Intelligence is defined by the New World Encyclopedia as:

> *"a military discipline that focuses on the gathering, analysis, protection, and dissemination of information of both strategic (long range actions intended to destroy military potential) and tactical (smaller operations of immediate significance in the field) value. This includes information about the enemy, terrain, and weather in an area of operations or area of interest, as well as information about political decision-making, military intentions, and dissidents. Intelligence activities are conducted both during peacetime and in war."* (New World Encyclopedia Contributors, 2018)

The Royal Canadian Air Force Doctrine on Intelligence, Surveillance and Reconnaissance (2017 p.vi) furthermore states that:

> "Intelligence, surveillance and reconnaissance (ISR) operations are a joint responsibility and key enabler to all other military operations and campaigns. Ultimately, ISR contributes directly to decision superiority, providing commanders and decision makers with timely and relevant information that supports the building of common and shared knowledge and understanding of the operational environment."

Military intelligence thus refers to the relevant and confirmed information which military planners use in decision-making and generally focus on information about the operational environment, hostile, friendly and neutral forces, the civilian population in an area of operations, and other broader areas of interest. Intelligence operations are executed throughout the political and military activity hierarchy (Ganger 2018). The type of topics covered in intelligence requirements include anything and everything that may influence the military mission in the physical, human or cyber domain.

An intelligence capability must include a set of functions which is described in the **Intelligence Cycle**. The intelligence cycle is a set of processes used to provide decision-makers (commanders) with useful information (intelligence). The cycle starts with a problem and consists of several processes including planning and direction, collection, processing & analysis and production, dissemination and utilisation/integration (Groupsense 2019). Intelligence cycle management is required for the overall intelligence function by guiding the activities of the intelligence cycle. Figure 1 below depict the intelligence cycle and provide some of the secondary activities present in each phase of the cycle. This section also provides a brief description for each of these phases in order to ensure thorough understanding of the system requirements going forward.

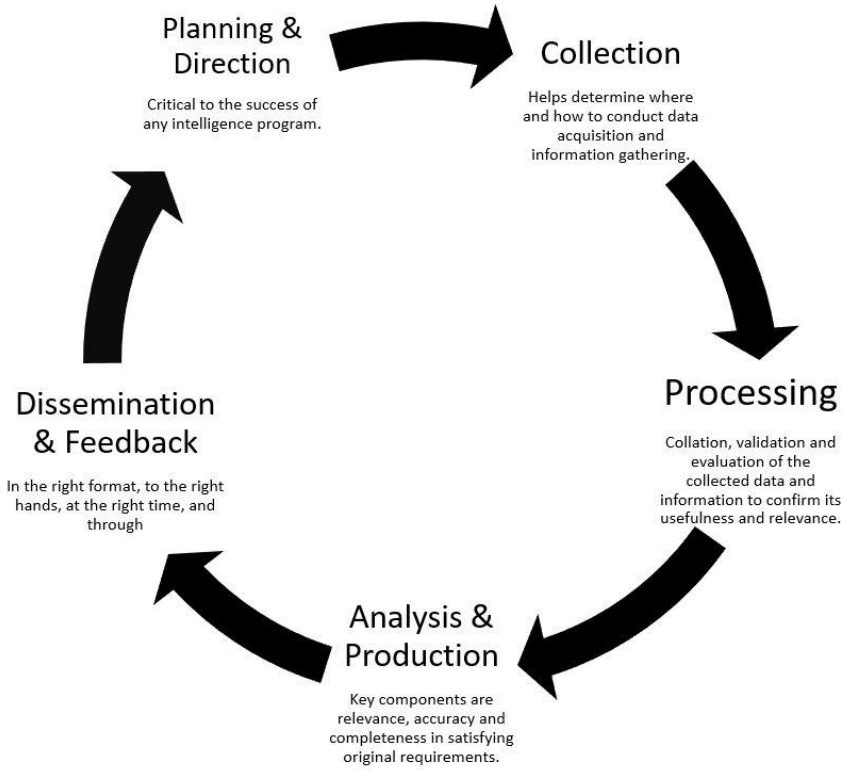

Figure 1: : Intelligence Cycle (adapted from: Groupsense, 2019)

- Planning and Direction. The first step generates and prioritises user information requirements to initiate the intelligence cycle. This includes an initial search to determine if existing information may meet their needs. These requirements guide the intelligence process. In the South African context this is referred to as the Intelligence Problem (similar to stating a research problem at the start of a research project). In data science terminology, this would be referred to as clarifying the business requirements.
- Collection. In response to requirements, the intelligence staffs develop an intelligence collection plan by identifying available sources and collection methods or sensors. The collection plan is then implemented towards acquiring the information needed.
- Processing. The collected information is processed by being evaluated (verified for level of validity), collated and integrated for exploitation. This is often seen as the first step of analysis as information is converted into the required format for analysis. In data science, this step would refer to data preparation and transformation to turn unstructured or raw data into a structured format for analysis.
- Analysis. Analysis turns collected information into intelligence assessments or products. It involves transforming the information into actionable intelligence by making deductions and conclusions about the current situation based on the knowledge and subsequent insight gained from the integrated picture of accumulated facts.
- Production. Intelligence is packaged into intelligence products - an appropriate format that meets the requirements of the specific user (decision-makers on different levels or functions), situation or technology.
- Dissemination. Dissemination places the intelligence product in the right hands, at the right place, at the right time and in the right format. Finished intelligence products take many forms depending on the needs of the decision-maker and reporting requirements. In terms of systems development, it is important to ensure secure and optimal information flow to enable effective intelligence dissemination.

- Feedback. The intelligence cycle is a closed loop; feedback is received from the decision maker and revised requirements issued. Once distributed, the intelligence product should not be ignored. This step thus involves evaluation of the intelligence by measuring the impact of the intelligence that was provided to the decision makers. Evaluative information and suggested improvements or newly identified requirements based on recent intelligence produced, must hence be fed back into the intelligence cycle in a dynamic manner.

This overview of the various interactive processes involved in an intelligence capability, demonstrates that developing and maintaining such a capability is a complex but crucial matter. The importance of military intelligence for defence forces is therefore reflected on as a consideration for future intelligence systems development, upgrading or maintaining.

Military intelligence is an essential military capability. The primary reason is that good intelligence provides a competitive advantage over an adversary. Elder (2007) makes a very clear statement about the importance of intelligence for military operations with reference to historical operational outcomes that have repeatedly demonstrated that smaller forces with less capable technologies than their adversaries, can win when their leaders' decisions are based on accurate intelligence. He thus states that military force and how it is employed are significant in driving combat outcomes, but that operational and tactical intelligence can have the most decisive impact on achievement of operational success, rather than numbers in force, technology or tactics (Elder, 2007).

As part of the process of Command and Control, commanders use the operational process of planning, tasking, execute and control the execution of the plan, and evaluate & assess continuously to re-appreciate and make changes to the operations. Intelligence and operational feedback (situational awareness) thus serves an immediate (in the present) as well as future-driven purpose as it informs the design and development of further capabilities to counter threats. The commander cannot successfully accomplish the activities involved in the operations process without information and intelligence. The design and structure of intelligence operations support the commander's operations process by providing him with intelligence regarding the enemy, the battle-field environment, and the situation. The operations process and the intelligence process are mutually dependent (Globalsecurity.org 2019 a).

Having a competitive advantage over one's adversary based on one's access to good intelligence is also referred to as having Information Superiority in modern literature. The United States Military Field Manual, 3 chapter 11 defines the term as follows: "Information superiority is the operational advantage derived from the ability to collect, process, and disseminate an uninterrupted flow of information while exploiting or denying an adversary's ability to do the same. Commanders exploit information superiority to accomplish missions." It is further explained that information superiority includes the ability to see first, understand first and act first – hence, to have the best data collection, effective analysis procedures and the ability to apply this to effective command and control (Globalsecurity.org 2019 b).

The nature of the information domain has changed over time and hence require changes to one's approach to harnessing it by means of one's intelligence capabilities. The NATO Allied Joint Publication-2 Joint Intelligence, Counter Intelligence and Security Doctrine (AJP-2) was updated to AJP(A) in February 2016. The main difference between the previous and current version being a change of focus from state-versus-state conflict to multiple smaller scale intervention and counterinsurgency operations. The more complex operational environment and the increasing number of factors that affect contemporary intelligence operations are now also considered. The updated version propagates the advent of the comprehensive approach to intelligence and the need for an increasing intelligence capability within NATO. AJP-2(A) also introduces the Joint Intelligence, Surveillance and Reconnaissance Concept and a revision of NATO intelligence requirement management and collection management (IRM&CM) functions. (NATO AJP-2 2016)

The requirement for intelligence for military operations, the importance thereof as well as the changes in the societal, information and technology environment suggest a need for changes in the current approaches to intelligence production towards consideration of the impact of Big Data, AI and Industry 4.0. This paper hence proposes a data science approach to the intelligence function, which will be presented in a framework in a later section. The next section provides some background in terms of the changes in the information domain and hence the demands on intelligence capabilities.

FROM FIRST INDUSTRIAL REVOLUTION TO INDUSTRY 4.0

Human history has marked the First Industrial Revolution when the steam engine was developed during the 1700's which allowed production to be mechanised. The Second Industrial Revolution started with the invention of electricity and other scientific advancements in the 1800's that led to mass production. The Third Industrial Revolution occurred with the emergence of computers and digital technologies from the 1950's which led to access to and consuming of vast amounts of information as well as increased automation of manufacturing and changes in industries such as banking, energy and communications (McGinnis 2018). The next chapter in human advancement, is the Fourth Industrial Revolution.

In the Information Age knowledge-based societies and high-tech global economies wherein the modernisation of information and communication processes has become the driving force of social development. The development of personal computers enabled access to and ability to share and store information for increasing numbers of workers. Furthermore, connectivity between computers within organisations led to the ability to share and access greater amounts of information at different levels (Cafrene 2016). Alberts (2011) add that complexity and change are the defining principles of the Information Age.

During the Information Age, that started in the 1970's, and especially with the internet becoming publicly available in 1991, access to information became easy and societies became consumers of huge amounts of data and dependent on the internet for the majority of the aspects of their lives (Weebly.com 2019). The Information Age is also referred to as the digital age as it refers to the period of history where a shift took place from traditional industry (which started with the Industrial Revolution) to an economy based on information computerisation. This is said to have started with the Digital Revolution that occurred between the 1950's and 1970's.

The development of new technology to improve people's lives, unfortunately also bring with it new vulnerabilities and challenges (Ryan 2017). New challenges occur in both the defensive stance (what are own new vulnerabilities and how can it be mitigated?) and the offensive stance (how can own capabilities be optimised for maximum competitive advantage?). The focus of this paper will be on the offensive stance – optimising own capabilities.

The phrase Fourth Industrial Revolution was first introduced by Klaus Schwab, the executive chairman of the World Economic Forum, in a 2015 article in Foreign Affairs. Schwab describes this fourth era to include technologies that combine hardware, software, and biology (cyber-physical systems) through advances in communication and connectivity. This era is expected to be marked by breakthroughs in emerging technologies in fields such as robotics, artificial intelligence, nanotechnology, quantum computing, biotechnology, the Internet of Things, the Industrial Internet of Things (IIoT), decentralized consensus, fifth-generation wireless technologies (5G), 3D printing and fully autonomous vehicles (https://en.wikipedia.org/wiki/Technological_revolution).

The dawn of the new era of Industry 4.0, brings with it a requirement to merge the cognitive abilities and complexities of human understanding, with Information and Communication Technologies (ICT), as well as other technologies. This implies that military intelligence and conducting military operations should also consider the implications of technological revolutions for capability development – new paradigms are required. From a Command and Control perspective, operations in "smart cities" are expected to present significant challenges due to increased complexity of the battlefield (Suri et al. 2016).

Research and innovation in new technologies are usually funded and applied by the defence industries of nations. In current times, most modern military forces have started and/or applied new research and development of technologies that fall within the description of Industry 4.0. This paper proposes an approach to military intelligence that is aligned with this new technological age.

CHALLENGES WITH PAST AND CURRENT APPROACHES TO MILITARY INTELLIGENCE SYSTEMS

The need for information superiority, implies the necessity to constantly upgrade and optimise one's national and military intelligence capability. In order to achieve this, one must apply a reiterative process of evaluation of past and present capabilities. Effective solutions are developed by evaluation and testing of own systems, as well as other's approaches through research, to record lessons learned and identify discrepancies.

The challenges with previous approaches to military intelligence are investigated in this section in terms of the elements of the intelligence cycle, which include the intelligence problem & intelligence requirements, data collection, processing, and dissemination. Two major past approaches will be discussed, and have been given names for description in this paper:

- Approach 1: Before the Digital Revolution; World Wars Approach
- Approach 2: After the Digital Revolution; Fish-Trawler Approach

Challenges with the World Wars Approach

Military intelligence progressed through three stages in history. Intelligence was first institutionalized by general staffs in the nineteenth century; during World War I, communication intelligence (COMINT) by means of radio intercepts gave it importance; and during World War II and the Cold War, it played such a significant role that intelligence officers gained equality in rank with combat commanders. Intelligence could, however, never supersede operations for intelligence in war works only through force. Intelligence may optimise operations by enabling economy of effort and providing a competitive advantage, but in the end, force is necessary for victory. This remains true in all kinds of operations, including the war on terrorism, involving non-state actors in which intelligence plays a major role. (Kahn, 2006).

United States (US) intelligence activity in the field of COMINT was sporadic prior to 1917 with little of its history recorded. The recorded history of American cryptanalysis began with the entry of the US into World War I. During this time intelligence codes and ciphers, even those used to carry the most sensitive information, were by current standards naive. It was hand-coded and hand-applied cipher systems usually overlying double-entry code books. These activities required human skills and patience but not the elaborate electronic and tabulating devices used today. Consequently, the codes which this government “cracked” from 1917 to 1919 were handled by a small group of lexicographers, mathematicians, and people who had acquired some background in what was then the hobby of cipher construction. (Safford, 1952)

In summary, Intelligence problems during this age were based on conventional warfare thinking and included aspects such as finding out the adversary’s strengths, firepower, movements, log lines, doctrine and plans. Mostly human intelligence (HUMINT) and COMINT were employed for collection due to little technology being available compared to today. Processing was done by humans in a paper and pencil methodology as well as making use of ciphers and cryptoanalyis. Dissemination of intelligence was also done via rudimentary methodology such as Morse code, humans (e.g. the Navajo Code Talkers) and even homing pigeons. This conventional approach relies heavily on human resources, skills and insight and was labour intensive, with little available technology for support.

The complex environment of present day with its rapid advances in as well as dependence on technology and access to vast amounts of information implies that this approach may no longer be effective. Some intelligence organisations, however, still follow this approach by sending human “collectors” to go out and obtain the information required as per the collection plan, with little utilisation of automated data analysis processes. Furthermore, this approach may still be executed in a fragmented way rather than applying holistic systems thinking. With the conventional approaches commonly used by defence organizations, it is difficult to bring together all of the data for analysis. Typically, organizations look at data in isolated silos (such as training, personnel, equipment or supply chains) and then theorize about the impact those areas may have on one another. With advanced analytics, they can break down those silos to see the larger picture, with the full scope of interrelationships.

The United Kingdom’s Intelligence Doctrine (JD 2-00) describes the conventional approach to intelligence as having fixed lines and boundaries between departments that include rules for inter-agency co-operation - a closed system in effect. Such a system will become more difficult to sustain in the 21st Century and, while useful for certain problems, it is not flexible enough to deal with problems that are more complex. An intelligence process based on this conventional approach does have utility as the bedrock of intelligence activity, but will have to become more flexible to effectively deal with asymmetric problems. The adaptive approach is described as requiring a flexible and more open system, where agencies work together each with its own focus, but contributing to a bigger picture. Agency collaboration and information sharing necessitates common protocols, but these should be agile: focussed on working together rather than articulating the obstacles to working together. (United Kingdom Ministry of Defence 2011)

The World Wars approach of the pre-digital age can be summarised as having the following characteristics:

- Human resources and skills stood central;
- Very little technology available for intelligence use (compared to today);
- Collection primarily by means of HUMINT and COMINT delivered a limited amount of facts (data).
- Limited to zero automated machine analysis – analysis was done by learned /skilled humans with paper and pencil method – making deductions and conclusions from facts;
- Dissemination methods was also limited and required mostly human activity, e.g. Navajo code talkers, Morse code sending, physical couriering, etc

Challenges with the Fish-Trawler Approach

The Fish-Trawler Approach was applied during the information age and is still being applied in many organisations today, including intelligence organisations. It is the opinion of the authors that this approach still has some value in specific types of business organisations, but that it is not the most effective approach to intelligence. This approach is characterised by the following:

- Utilising of technology stands central with strong dependence on automated ICT processes.
- Ease of access to vast amounts of data.
- Acquisition and storage of such huge amounts of data, including invalid or less valid data.
- Application of computerised analytic tools based primarily on statistics.
- Visualisation of the results in charts and Geo-Intelligence products, e.g. heat maps.
- Bottom-up approach that tend to be undirected.

The rate of growth in data sources available for collection as well as the analysis thereof is probably the main challenge facing intelligence professionals today. The magnitude of potentially relevant or useful information for analysis is overwhelming for current intelligence systems, while more is being generated and stored daily. However, **more data, if not properly used, does not support the intelligence process**. Dedicated processing and analysis tools are required to uncover important facts from the mountain of information (Knopp et al. 2016, Szeligowski 2018). With really big data, the data is there before it is demanded. As a result, the approach many organisations take is to collect all available data and then search for significant patterns in the data.

The operational environment of the military also tends to be data-rich. Ubiquitous sensing and social media platforms have made more data available for analysis than ever before in history. Technology enables the creation and storage of data in unimaginable quantities (Knopp et al. 2016). The current bottom-up approach require intelligence tools to trawl through vast datasets at a high speed to identify specific signatures (Couch & Robins 2013, Szeligowski 2018).

Applying a mostly linear approach to analysis does not provide reliable insight into the current real world of complex systems. A systems approach must therefore be applied to intelligence capability development in order to produce effective intelligence. “A system is a collection of elements or components that are organized for a common purpose” (techtarget.com/definition/system 2019). The Lexico.com (2019) dictionary defines a system as “a set of things working together as parts of a mechanism or an interconnecting network; a complex whole”. A system is thus a collection of elements, components or parts that are organised for a common purpose and in interaction with one another.

General systems theory includes broadly applicable concepts and principles, rather than concepts and principles applicable to only one domain of knowledge. A distinction is made between dynamic / active systems *and* static / passive systems. Active systems are activity structures or components that interact in behaviours and processes whilst passive systems are structures and components that are being processed. The field is related to systems thinking, machine logic and systems engineering. (https://en.wikipedia.org/wiki/Systems_theory 2019)

Current Intelligence capabilities that still apply the linear approach to analysis and/or apply this in silo’s of focus areas will thus achieve limited insight. Such application does not provide the theory (conceptual approach) to understand a multi-actor competition unfolding in a complex environment. The existing and limiting conceptual approach to intelligence centres on scientific theory that holds the sum of a system’s parts equal to its whole; therefore, study of components leads to understanding of the system. This approach

is inconsistent with contemporary understanding of complex systems. It discounts the dynamics of the system generated from interactions between components and subsequent synergy of the system. It assumes that linearized systems models will accurately approximate the real system. The increase in interconnectedness that became reality since the end of the Cold War makes this assumption unlikely. Joint doctrine presumes simplicity in the adversary network, e.g. determining Centre of Gravity (COG). It does not employ cognitive skills suited to understand a complex environment. This approach thus still uses deductive logic, for the analyst to confirm expected behaviour based on a theory (or adversary doctrine) held to be true, rather than interpreting and explaining the adversary (Smith 2016)

Descriptive analytics sort through and summarize raw data to make it understandable, through spreadsheets, charts, reports and other kinds of presentations. Descriptive analytics commonly help defence organizations determine the current situation and past trends, with the aim of guiding future actions. Essentially, these analytics provide hindsight and don't explain why things happen (Suri et al. 2016). In summary, the biggest challenge of the Fish-Trawler Approach to intelligence systems is:

- the challenges associated with Big Data (Volume, Variety, Veracity, Velocity);
- the big data challenges include spending more resources than necessary to collect and analyse vast amounts of data (whilst the "less is more" approach may provide better results);
- a largely un-directed methodology which may include collection and processing of less valid data;
- linear and statistical analysis that include invalid/ less valid data may result in less reliable answers;
- over-reliance on technology (machines) and automated analysis with limited inductive reasoning by humans.

The next sections will provide a brief overview of current best practices in data science and considerations for an alternative approach to military intelligence systems in order to overcome some of the challenges highlighted for the World Wars approach and the Fish-Trawler approach.

CURRENT BEST PRACTICES IN DATA SCIENCE

The study and application of Big Data spawned the interdisciplinary field of Data Science which combines the domains of operations, mathematics, and computer science as well as several ancillary fields such as social, organizational, intelligence, economics, psychological and cognitive sciences (Baker & Henderson 2017, Knopp et al. 2016). Data Science analytics is the scientific process of transforming data into insights for making better decisions (Gauthier & Kaluzny 2018).

Data Science represents the science of learning from data, and is defined as a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms (structured and unstructured). Data Science is the application of specific algorithms for extracting patterns from data through traditional statistics, data analysis, machine learning, artificial intelligence, pattern recognition, and database systems (Vasant 2013, Leek 2013, Leskovec et al. 2014, Knopp et al. 2016, Donoho 2017, Gauthier & Kaluzny 2018).

Data Science applies sophisticated semi-autonomous or autonomous analytic techniques and tools to discover deeper insights to make predictions or recommendations about solving problems. This requires a *prior understanding* of the complexities in data sets (Jani 2016). Advanced analytic techniques include data mining, machine learning, pattern matching, forecasting, visualization, semantic analysis, sentiment analysis, network and cluster analysis, multivariate statistics, graph analysis, simulation, complex event processing, neural networks. Advanced analytics goes beyond descriptive analytics ("what happened?") to provide predictive analytics ("what will happen?"), and prescriptive analytics ("what should we do?") (Gauthier & Kaluzny 2018).

Business intelligence and data analytics are new means of increasing efficiency and enabling decision support in defence organisations (Gauthier & Kaluzny 2018). The intelligence community has come to recognise Data Science as an important new technology (Frank 2017). Based on Data Science successes in the civilian sector, defence forces start to leverage its data to increase situational awareness and achieve information dominance over its adversaries. Fully leveraging all the data generated in military networks is essential to outmanoeuvring adversaries in all domains and unpick the digital 'fog of war' (Baker & Henderson 2017, Pearson et al. 2018). Data Science benefit information processing for intelligence in the following ways (Jani 2016, Ganger 2018):

- Collection, sharing and characterizing of data become automated. This may help to break down data silos as multi-disciplinary analysis require multi-disciplinary data.
- Processing time for analysing complex data structures can almost be reduced to being real time. This helps in anticipating surprise.
- Incisive analysis to maximize insights from dynamic datasets.
- Presentation of result can be refined to only key features that lead to effective decision making.

Artificial Intelligence (AI) is a major component of Data Science. The field of AI is a wide area of study that includes developments of big data, predictive analytics, cognitive computing, and deep learning. These are enablers for improving the intelligence level of machines to support information analysis (Szeligowski 2018). Modern Military Systems will benefit from the application of AI. AI can compress the OODA loop by augmenting situation awareness and decision making (Szeligowski 2018).

It must, however, be considered that applying AI to military intelligence systems also present challenges to be weighed in feasibility assessment and/or must be managed. Military operations place unique demands on AI tools. Often, military operations are conducted in new settings with new requirements and sparse data sets that afford limited training of AI algorithms. Being multi-domain in nature (i.e. land, maritime, air, space and cyber domains), the tools have to link across highly dispersed elements operating in complex environments. Also, coalition operations with international participants having common aims tend to operate under different policies and levels of trust. Furthermore military operations may be undertaken against an active opponent, who is deliberately attempting to deceive the AI systems; and the adversary may be using their own AI systems to control the deception operation (Pearson et al. 2018).

CONSIDERATIONS FOR A NEW APPROACH TO INTELLIGENCE SYSTEMS

Intelligence is crucial to the development of understanding. It requires systems, architectures and practitioners flexible enough to operate in complex environments and a command climate that promotes collaboration and creates the organisational structures required to achieve fusion at the point of need (United Kingdom Ministry of Defence, 2011).

Shifting the intelligence approach from what an adversary is doing and estimating what that adversary will do next to ***why*** an adversary or operational problem has emerged is a significant shift from the direction of over 100 years of intelligence development (Smith 2016). This would require a systems theory approach to analysis in order to arrive at reliable insight. Smith (2016) proposes three adaptations to the Joint Force approach to intelligence to keep intelligence relevant in the contemporary environment:

- Intelligence must ground its conceptual approach in theory that accounts for the characteristics of complex environments, not in the archaic scientific methods of Descartes and Newton.
- Shift from a reliance on a deductive cognitive method to one inclusive of inductive and abductive methods.
- Intelligence must retain existing analytic techniques, but broaden analysts' skillsets to include advanced analytic techniques and apply these techniques in a manner logically consistent with the characteristics of complex systems.

Alexandra Novosseloff and Olga Abilova (2016) did a study for the UN to improve its intelligence function. They argue that, in reforming its analytical capacities and capabilities, the UN should focus, first and foremost, on improving its current structures and on strengthening information analysis and sharing more than information collection. They add that the UN should prioritize developing a comprehensive information-management system rather than new intelligence infrastructure, which they fear most member states would be likely to oppose for reasons of funding and politics. These recommendations imply that better intelligence is based on better information management and analysis, rather than new technology (hardware and software products).

PROPOSED DATA SCIENCE FRAMEWORK FOR MILITARY INTELLIGENCE SYSTEMS

The approach to producing military intelligence by applying contemporary data science best practices, proposes inclusion of the following principles:

- Directed (focussed) approach to data acquisition – less is more.
- Business requirements (desired end-state) to provide direction.
- A common data model to support interoperability and data preparation (structuring).
- Data preparation with human participation and inductive reasoning.
- Systems theory approach to analyse data effectively towards prediction and forecasting for maximum human insight.
- Linked to effective C2 system for interactive decision support.
- Effective visualisation of intelligence products for decision-support.
- Application of sociotechnical systems principles.

Proposed Intelligence Systems Framework

Given the new sources of information and development of new tools, the intelligence process may change to a new format, as seen in figure 2 below. This is an adaptation of the traditional Intelligence process as per the Intelligence Cycle discussed above. The figure depicts the proposed data science approach to producing military intelligence with a description of the steps in the process thereafter. Counter-terrorism will be used as an example topic to illustrate the steps in the process.

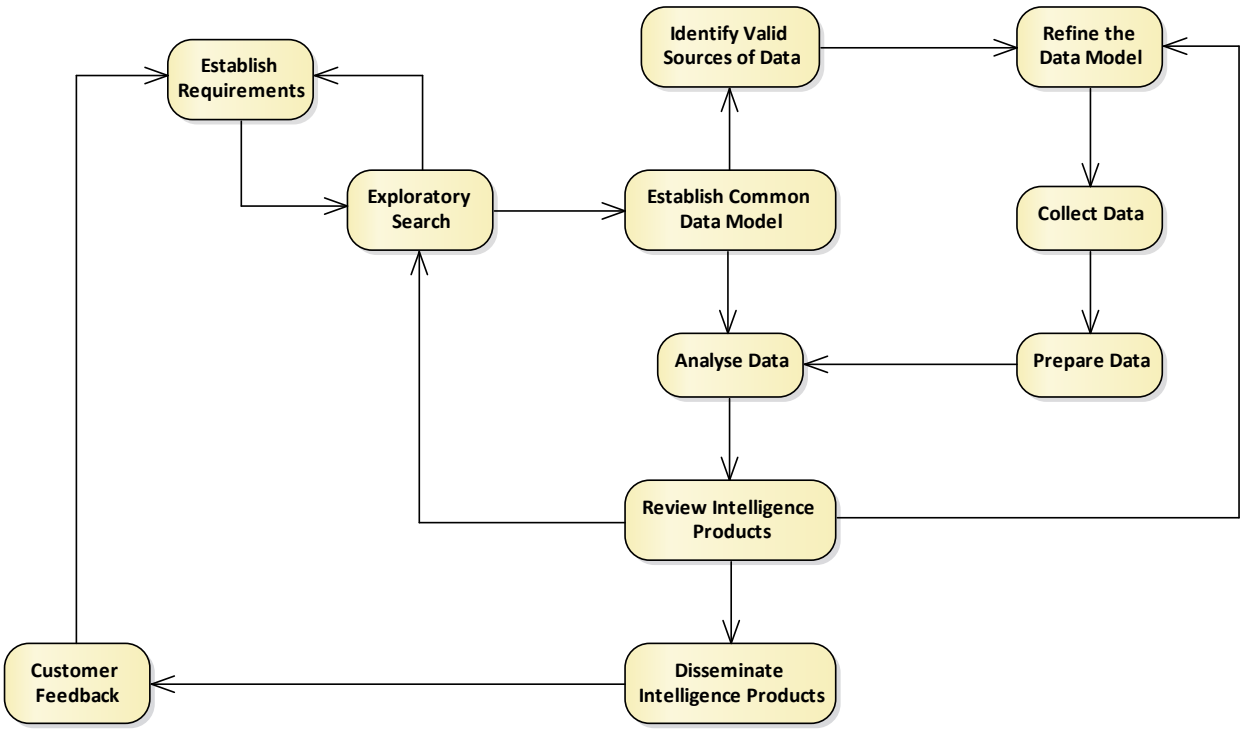

Figure 2 : Proposed Intelligence Systems Framework

Military Intelligence needs to develop and validate a process to guide how to integrate data science capabilities into operations. Data mining and analysis is an iterative process that needs to be customized and streamlined. This process model combines the functions from the intelligence and targeting models with a typical data science process, building on common functions and addressing gaps.

Establish Requirements

The intelligence cycle begins with a question, which guides a user's need for specific information. In data science this is the critical step in the process. It can be compared to the research-initiating question of a research project. In military intelligence terms, this translates to the Intelligence Problem of an operation. These are the higher-level question(s) that the commanders need to answer in order to make sound decisions for effective command and control. In the proposed model, different missions will have unique intelligence problems or business requirements. These questions will delineate a specific data domain or topic for analysis. In the domain of counter-terrorism, the intelligence problem may be any of the following:

1. What is the current terrorism threat?
2. What are the modus operandi of terrorist threat group(s)?

This step encapsulates the planning, directing, and decision activities of the intelligence process. The goal is to establish what data science outputs are needed. It is important that these business requirements be well clarified to ensure that it is correctly understood by the intelligence function as the desired end state of the intelligence process. The data science team should think through the analytical questions, how they relate to one another, how they support the mission, and when and what analytical outputs are needed.

Exploratory Search

Once the requirement(s) has been clarified, a directed and systematic literature review, including searching of already existing intelligence products and available (trusted) online data must be done on the specific topic and domain. The hypotheses and associated metrics defined in the requirements function drive what data must be acquired, how much data is needed, and how often it is updated. Once collected, the data is transformed for follow-on analysis. During this step, some initial useful elements of information may be uncovered already and could be presented in ad-hoc information briefings to the commander and/or relevant stakeholders. This illustrates the essential element of human insight in the sociotechnical systems of Industry 4.0.

The data science team can then form hypotheses that help measure the progress of moving from the current state to the end state. Once hypotheses are identified, the data science team can draft measures, or metrics, that later confirm or fail to confirm it. Skills required in this phase include data science, intelligence expertise, network systems engineering, mathematical modelling, human behaviour modelling, and simulation engineering.

The exploratory search of literature and web-based sources will create initial insight into the specific domain and must result in further research objectives for the intelligence process – the intelligence problem must be translated into further intelligence requirements. This initial insight may result in clustering of specific information elements in the domain, identifying specific jargon, synonymous terms and different existing models as well as relationships between constructs within the domain. It should further be noted that this initial exploratory search (data collection) may already yield relevant answers to be applied for decision-making, without requiring more intensive analysis.

In terms of the counter-terrorism domain example, the exploratory search would already deliver lists of currently identified terrorist groups with various perspectives on how they are described and categorised. Reports on events committed by such groups also describe different aspects from different perspectives which need to be analysed and translated into specific required data elements that would describe a standardised model to describe modus operandi. The latter output will also require knowledge or exploration of the military domain in order to have insight into how a military commander would want the modus operandi of a hostile group described.

Establish Common Data Model

The required data to inform the rest of the intelligence and targeting process needs to be explicitly specified. The output of the exploratory search step must include sufficient insight into the domain to conceptualise a common data model for the intelligence system. It is proposed that such a data model is based on an ontology. This ensures that specific domain with its attributes and relationships between elements, objects,

constructs and terms are semantically linked. The data model includes the required data elements for the specific domain and how it inter-relates towards providing answers to the initial intelligence requirements.

In terms of the counter-terrorism domain example, a full ontology description is beyond the scope of this paper, but figure 3 below provides an example of a part of such an ontology. It is important to note that one's common data model should be customised and unique enough to adhere to the military principle of maintenance of the aim. The data model contain a client-relevant taxonomy and defined relationships, but also contain standardised elements, e.g. chosen upper ontologies to enable interoperability with other stakeholders and/or exploitation of already existing valid data models, sources and analysis frameworks or analytics tools. Furthermore, the ontology should be seen as dynamic as it should be constantly reviewed, refined and further developed with strict adherence to the principle of maintenance of the aim.

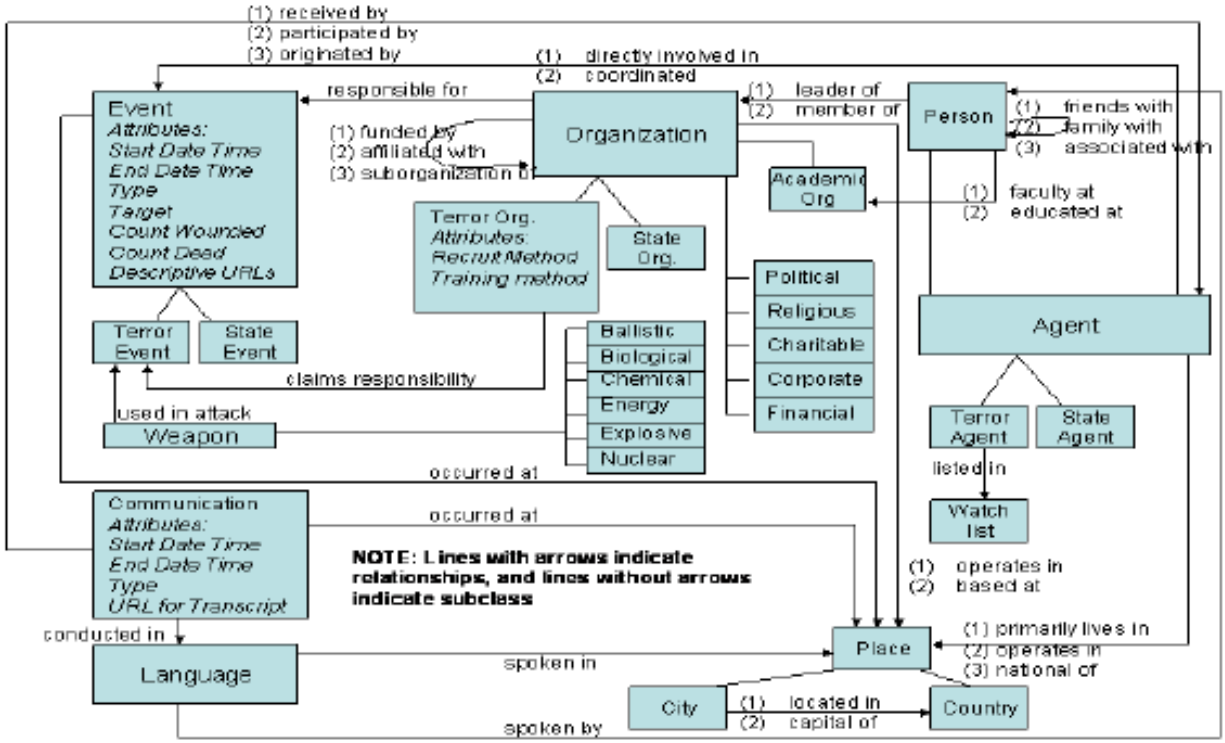

Figure 3: Counter-Terrorism Ontology Model (Aleman-Meza et al. 2009)

Identify Valid Sources of Data

The first version of the common data model will reflect required fields for data acquisition – the lowest level of elements of information or the pieces of a puzzle required to complete the picture. In traditional intelligence terms, this step of the process is called compiling the collection plan. It includes identifying relevant and valid data sources to collect the required data elements from as well as what sensors to deploy for this purpose, i.e. where data can be found and how will it be retrieved.

A crucial consideration in this step is the reliability of the data sources, as well as the validation of the data collected. Sources identified for data acquisition must be evaluated for reliability in order to ensure obtaining of valid data. Furthermore, data can be validated if the same results are obtained from more than one reliable source. It is important to note that some elements of data are valid and reliable based on the dedicated nature of the sensor used for collection, e.g. a special reconnaissance operator acquiring directed data items by means of observation or audio/ audio-visual recording. This step is essentially the directed element to this approach. It is the application of a traditional intelligence principle using web-based sources and data science technologies.

Refine the Data Model

Before being able to operationalise the intelligence system, the common data model needs to be trained and refined. This is done by using a valid test dataset. Such a dataset will be manually obtained by a human functionary, e.g. by an intelligence functionary scraping a specific reliable source and preparing the data by turning unstructured data into a structured dataset based on the common data model. The test dataset may present new insights that can be applied to further refine the data model. It can also result in the first level of training of the model if inferences are made by the system based on the rules defined in the data model.

Collect Data

Tools and processes must be developed to acquire valid data from multiple reliable sources, using various sensors. This is a significant task involving capturing data from sensors and moving and consolidating this data to a persistent data store. Smart data collection will ensure effective use of sensors.

The data is secured to prevent the enemy from manipulating the data to deceive analytics. The data is then reformatted and cleaned. Reformatting includes actions to store the data in a format that is compatible with the persistent data store. Then, data is parsed, translated, and mapped from its native schema into the schema of the persistent data store. Data cleaning, also known as data normalization, is the process of ensuring data integrity and involves deliberate steps to address incomplete, duplicate, or inconsistent records.

Prepare Data

The data acquired from various sources and using different sensors, will largely be unstructured data. Unstructured data is defined as being of various formats and non-standardised terminology. Data therefore needs to be prepared by transforming it into standardised format for importing into the common data model. Data may also need to be sorted into the correct sequence for being imported into the system database. As the intelligence system becomes more mature, more automated tools may be developed and added to the system to enable automated data preparation.

This step should be approached from a sociotechnical perspective as it requires interaction between the human and technology, with increasing levels of automation. The configuration of data depends on effective conceptualisation, software development and set-up by the human specialists.

Analyse Data

The methodology, models and analytics tools used in one's intelligence system should be based on the domain data model and most importantly on arriving at the answers to the questions identified in the initial steps. Hence, the common data model dictates the analytics and analysis methodology. The structuring of the data in databases is also according to the common data model. This dependency of analysis on the common data model is indicated with the arrow in the model from the data model to the analysis process.

This step also requires high levels of human interaction; for decisions on when and how to deploy which appropriate analysis technique or application and to ensure valid interpretation towards reliable insight and hence high quality intelligence.

The analytic techniques of decomposition, link analysis, pattern analysis, trend analysis, technical analysis, anticipatory analysis, recomposition, and synthesis apply in complex environments. Analysts can use these techniques to explain activities that have happened, which informs an understanding of how the system has behaved or is behaving. However, analysts should not use these tools to linearize the system for forecasting purposes. (Smith 2016)

Rapid sense making requires analysts to generate assessments within minutes to provide precise and actionable intelligence to decision makers. To automate and speed up as much of the analysis process as possible, a human analyst can arrange analytical tools in a sequence known as a model, which can then be automatically run to answer the same question repeatedly, with modified parameters and a different set of data. Scripts are authored and edited to test the hypotheses. The data science team initially runs and tests these scripts on a subset of data loaded on local computing resources (i.e. a local cluster, server, or workstation). Once tested on an extract of the data, the data science team uploads the scripts into a production data science computing environment. Once the scripts are complete, the data science team can inspect their output to verify if the outputs matches intended behaviour. The ultimate product of this phase is

a set of verified analytics (script outputs) that potentially answer hypotheses from the Establish Requirements phase.

The products for this phase are permanently deployed analytics running in the data science production environment, regularly consumed as part of broader intelligence operations workflows. This involves making the new analytics, previously tagged as developmental, fully accessible to all relevant stakeholders on the production system. The analytics are integrated into dashboards and similar tools and fully documented. The analytics are carefully secured to ensure the enemy does not compromise our data-driven decision-making processes. In the long term, the data science team should report their efforts to the broader community and archive any results.

Review Intelligence Products

The goal of this phase is to determine if the hypotheses are answered by the analytics. The team makes comparisons and selects the analytics answering the hypotheses in the shortest amount of time with the least probability of error. If the analytics do not meet the requirements, then the scripts are redesigned. The ultimate product of this phase is a set of analytics that allows a decision maker to answer the hypotheses. Skills required in this phase include data science, computer science, mathematics, statistics, machine learning, information visualization, and intelligence expertise.

Newly processed information is reviewed to identify high-value and time-sensitive information that can immediately support the mission. The products of this stage are results from currently deployed analytics, charts, and scripts that immediately answer current or past hypotheses. The exploitation phase should also include visualization through automated charts and maps that track aggregate trends in the data.

Disseminate Intelligence Products

The format of presenting and disseminating the answers to the initial intelligence problem is very important. Military commanders do not want to be flooded with the details of the data elements or the process of analysis, but need a reliable answer to their intelligence problem timeously. How to present such answers in the most effective and unambiguous manner (ensuring correct interpretation) must be carefully considered and conceptualised. The intelligence system may also allow the command and control user (decision-makers) to run more specific queries based on the data model and drilling down into more detail regarding the results of analysis as presented in intelligence products.

Customer Feedback

The outcome of this process is a review of the deployed analytics' performance, validity, relevancy, and data sources. Performance data (latency, accuracy and resource consumption) is compiled and reported for each analytic. Each analytics' data source is reviewed to ensure their integrity. The data science team should also collect and review usage data about how users apply the intelligence products – and how valuable the output was for decision-making.

A change in usefulness could equate to a training deficiency, a loss of confidence in an analytic, or changing information requirements. The original hypotheses are reviewed to ensure they are still relevant to the organization's mission and operations. If these have changed, the entire process is restarted to address evolving requirements. Skills required during this phases include data science, data architecture, and information technology administration.

CONCLUSION

Military Intelligence has always been crucial for military decision-making. Modern day conflicts give rise to complex problems that traditional military intelligence approaches and tools struggle to resolve. Traditional and current intelligence processes and tools are becoming increasingly inadequate to support the decisions of commanders and other decision makers. The traditional approach to intelligence before the digital age was human-intensive and lacked the technological advancement of the data acquisition and analysis tools of present day. After the onset of the digital age, organisations and even intelligence capabilities became overly dependent on ICT for acquisition and analysis of big data. A tendency developed to just trawl vast amounts of data due to its availability and to apply mostly linear models and descriptive statistics to arrive at outputs. This paper proposes an alternative model based on data science principles and especially that of socio-technical systems. This requires marrying up of the human intensive traditional approach with the

machine-intensive big data approach, to arrive at a directed approach to intelligence systems aimed at yielding more reliable outputs by acquisition of more valid data (albeit less in quantity), and more human interaction for inductive analysis techniques based on systems theory.

REFERENCES

Abilova, O. and Novosseloff, A., 2016. Demystifying intelligence in UN peace operations: Toward an organizational doctrine.

Alberts, D.S. 2011. The agility Advantage: A Survival Guide for Complex Enterprises and Endeavours. CCRP Publication, USA.

Aleman-Meza, B., Sheth, A., Palaniswami, D., Eavenson, M. and Arpinar, I., 2009. Semantic Analytics in Intelligence: Applying Semantic Association Discovery to determine Relevance of Heterogeneous Documents. 5. 10.4018/978-1-59140-935-9.ch020. Available at: https://www.researchgate.net/figure/ National- Security- and-Terrorism-Part-of-SWETO-Ontology_fig1_244441023

Austin, N.J. and Rankov, N.B., 2002. Exploratio: Military & political intelligence in the roman world from the second punic war to the battle of Adrianople. Routledge., N.J. and Austin, N.J. and Rankov, N.B.,

Baker, J.W. and Henderson, S., 2017. The Cyber Data Science Process. The Cyber Defense Review, 2(2), pp.47-68.

Byman, D., 2016. Intelligence and its critics. Studies in Conflict & Terrorism, 39(3), pp.260-280.

Cafrene, J. 2016. Characteristics of digital age. Available at: http://thecomplexgroup4.blogspot.com/2016/05/characteristics-of-digital-age.html

Couch, N. and Robins, B., 2013. Big data for defence and security. Occasional Paper, Royal United Services Institute, pp.801-23.

Donoho, D., 2017. 50 years of data science. Journal of Computational and Graphical Statistics, 26(4), pp.745-766.

Elder, G. 2007. Center for the Study of Intelligence, CSI Publications, Studies in Intelligence, vol 50 no2, Intelligence in War: It Can Be Decisive. Available at: https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol50no2/html_files/Intelligence_War_2.htm

Frank, A., 2017. Computational social science and intelligence analysis. Intelligence and National Security, 32(5), pp.579-599.

Ganger, R., Coles, J., Ekstrum, J., Hanratty, T., Heilman, E., Boslaugh, J. and Kendrick, Z., 2018. Army Intelligence Data Science Workshop Summary Report (No. ARL-TN-0931). US Army Research Laboratory Aberdeen Proving Ground United States.

Gauthier, Y. and Kaluzny, B., 2018. Data Analytics: The Same Old Next Big Thing for Military OR?. Defence Research and Development Canada= Recherche et développement pour la défense Canada; DRDC/RDDC.

Globalsecurity.org, 2019. US Army Field Manual 2-0. Chapter 4: Intelligence. Available at https://www.globalsecurity.org/intell/library/policy/army/fm/2-0/chap4.htm

Globalsecurity.org, 2019. US Army Field Manual 2-0. Chapter 11: Information Superiority. Available at: https://www.globalsecurity.org/military/library/policy/army/fm/3-0/ch11.htm

Groupsense, 2019. How To Use The Intelligence Cycle To Secure Your Brand. Available at: https://www.groupsense.io/intelligence-cycle/

Jani, K., 2016. The Promise and Prejudice of Big Data in Intelligence Community. arXiv preprint arXiv:1610.08629.

Kahn, David. 2006, The Rise of Intelligence. Foreign Affairs, vol. 85, no. 5, pp. 125–134. JSTOR, www.jstor.org/stable/20032075.

Knopp, B.M., Beaghley, S., Frank, A., Orrie, R. and Watson, M., 2016. Defining the Roles, Responsibilities, and Functions for Data Science Within the Defense Intelligence Agency (No. RR-1582-DIA). RAND National Defense Research Institute Santa Monica United States.

Leek, J., 2013. The Key Word in 'Data Science' Is Not Data, It Is Science. Simply Statistics.

Leskovec, J., Rajaraman, A. and Ullman, J.D., 2014. Mining of massive datasets. Cambridge university press.

Lexico.com, 2019. What is system? - Definition from lexico.com. [online] Available at: https://www.lexico.com/en/definition/system (2019 [Accessed 13 Jun. 2019].

McGinnis, D. 2018. What Is the Fourth Industrial Revolution? Available at: https://www.salesforce.com/blog/2018/12/what-is-the-fourth-industrial-revolution-4IR.html)

NATO - AJP-2, 2016. NATO Allied Joint Publication-2 Joint Intelligence, Counter Intelligence and Security Doctrine. Available at: https://standards.globalspec.com/std/9994887/ajp-2

New World Encyclopedia contributors. 2018 Oct 5, Military intelligence. New World Encyclopedia, [cited 2019 Jun 16]. Available at: http://www.newworldencyclopedia.org/p/index.php?title=Military_intelligence&oldid=1015088.

Pearson, G., Jolley, P. and Evans, G., 2018. A Systems Approach to Achieving the Benefits of Artificial Intelligence in UK Defence. arXiv preprint arXiv:1809.11089.

Royal Canadian Air Force, 2017, Doctrine on Intelligence, Surveillance and Reconnaissance. B-GA-401-002/FP-001, 2nd Ed. Available at: at http://www.rcaf-arc.forces.gc.ca/en/cf-aerospace-warfare-centre/aerospace-doctrine.page.

Ryan, L., 2017. Cyber security in the digital age: The threats are real, keep yourself safe. Available at: https://thenewsrep.com/91793/cyber-security-in-the-digital-age-the-threats-are-real-keep-yourself-safe/

Safford, L.F., 1952, A Brief History of Communications Intelligence in the United States. Available at: http://www.worldwar2facts.org/communications-intelligence-history.html)

Smith, F.A., 2016. The Importance of Why: An Intelligence Approach for a Multi-Polar World. (Norfolk VA United States: National Defense University Norfolk).

Suri, N., Tortonesi, M., Michaelis, J., Budulas, P., Benincasa, G., Russell, S., Stefanelli, C. and Winkler, R., 2016, May. Analyzing the applicability of internet of things to the battlefield environment. In 2016 international conference on military communications and information systems (ICMCIS) (pp. 1-8). IEEE.

Sun Tzu, 1963, The Art of War (New York: Oxford University Press) 144.

Svenmarck, P., Luotsinen, L., Nilsson, M. and Schubert, J., 2018, May. Possibilities and challenges for artificial intelligence in military applications. In Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision Making Specialists' Meeting.

Szeligowski, R., 2018. Cognifying the OODA Loop: Improved Maritime Decision Making. Gravely Naval Research Group, Naval War College Newport United States.

Techtarget.com. (2019). What is system? - Definition from WhatIs.com. [online] Available at: https://searchwindowsserver.techtarget.com/definition/system [Accessed 13 Jun. 2019].

United Kingdom Ministry of Defence; Development, Concepts and Doctrine Centre. 2011. Joint Doctrine Publication 2-00 (JDP 2-00) (3rd Edition), Understanding And Intelligence Support To Joint Operations. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/311572/20110830_jdp2_00_ed3_with_change1.pdf

Vasant, D., 2013. Data Science and Prediction. Communications of the ACM: 64-73.

Wikipedia, History of Espionage. [cited 2019 Jun 16]. Available at https://en.wikipedia.org/wiki/History_of_espionage.

Wikipedia, Technological Revolution. [cited 2019 Jun 10]. Available at: https://en.wikipedia.org/wiki/Technological_revolution

Wikipedia, Systems Theory. [cited 2019 Jun 10]. Available at: https://en.wikipedia.org/wiki/Systems_theory

Weebly.com, 2019. History of Technology – Information Age. Available at: https://historyoftechnologyif.weebly.com/information-age.html