A REFLECTION ON DATA SCIENCE IMPLEMENTATION WITHIN THE FRENCH ARMY

A thesis presented to the Faculty of the U.S. Army Command and General Staff College in partial fulfillment of the requirements for the degree

MASTER OF MILITARY ART AND SCIENCE
General Studies

by

MATHIEU JUTTET, LIEUTENANT COLONEL, FRENCH ARMY
Engineering degree, École Spéciale Militaire de Saint Cyr, Coëtquidan, France, 2006

Fort Leavenworth, Kansas
2020

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-06-2020	Master's Thesis	AUG 2019 – JUN 2020

4. TITLE AND SUBTITLE

A Reflection on Data Science Implementation within the French Army

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Lieutenant Colonel Mathieu Juttet

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Command and General Staff College
ATTN: ATZL-SWD-GD
Fort Leavenworth, KS 66027-2301

8. PERFORMING ORG REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution is Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Data Science takes an increasing place in modern organizations as leaders believe data-enhanced decisions to define the paradigm of a new era. Techniques proceeding from it such as Artificial Intelligence are more and more employed within military institutions such as the French Army.

Implementing Data Science requires both technical assets allowing an adequate and effective collection, storage, and exploitation of data; and a competent workforce properly employed to optimize expected improvements.

Based on theoretical models, using a comparison with the U.S. Army, and taking in account the encompassing French ministry current project, this thesis explores recommendations for the French Army to successfully manage the implementation from an organizational perspective.

15. SUBJECT TERMS

Data Science, Artificial Intelligence, French Army.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. PHONE NUMBER *(include area code)*
(U)	(U)	(U)	(U)	85	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

MASTER OF MILITARY ART AND SCIENCE

THESIS APPROVAL PAGE

Name of Candidate: Lieutenant Colonel Mathieu Juttet

Thesis Title: A Reflection on Data Science Implementation within the French Army

Approved by:

_____, Thesis Committee Chair
Kenneth E. Long, D.M.

_____, Member
William K. McCurry, MBA

_____, Member
Gregory K. Butts, M.S., M.A., MMAS

Accepted this 12th day of June 2020 by:

_____, Acting Director, Office of Degree Programs
Prisco R. Hernandez, Ph.D.

The opinions and conclusions expressed herein are those of the student author and do not necessarily represent the views of the U.S. Army Command and General Staff College or any other governmental agency. (References to this study should include the foregoing statement.)

ABSTRACT

A REFLECTION ON DATA SCIENCE IMPLEMENTATION WITHIN THE FRENCH ARMY, by Lieutenant Colonel Mathieu Juttet, 85 pages.

Data Science takes an increasing place in modern organizations as leaders believe data-enhanced decisions to define the paradigm of a new era. Techniques proceeding from it such as Artificial Intelligence are more and more employed within military institutions such as the French Army.

Implementing Data Science requires both technical assets allowing an adequate and effective collection, storage, and exploitation of data; and a competent workforce properly employed to optimize expected improvements.

Based on theoretical models, using a comparison with the U.S. Army, and taking in account the encompassing French ministry current project, this thesis explores recommendations for the French Army to successfully manage the implementation from an organizational perspective.

ACKNOWLEDGMENTS

The author would specially like to thank Dr. Kenneth Long, chair of this thesis, for its helpful support to improve the methodology and make the whole content more relevant. Warm thanks are also addressed to Mr. Kenna McCurry and COL(R) Gregory Butts, committee members, without whom achieving this work would not have been possible.

A mention is also aimed at the office of Degree Program and more specifically to Dr. Robert F. Baumann, Dr. Prisco R. Hernández, and Mrs. Ann Chapman for their support and dedication to the MMAS program and for their availability.

Finally, the author acknowledges his loving wife Maÿlis and children Christine, Etienne, Marie, Agnès, and Jacques for their patience and comprehension for late work.

TABLE OF CONTENTS

Page

ACRONYMS

AFC	Army Future's Command
AI	Artificial Intelligence
AMSO	Army Modeling and Simulation Office
APCS	Applied Professional Case Study
CARL	Combined Arms Research Library
CDM	Chief Decision Maker
CDEC	*Centre de Doctrine et d'Enseignement du Commandement*
CEMAT	*Chef d'État-Major de l'Armée de Terre*
CFT	Cross-Functional Team
CICDE	*Centre Interarmées de Concepts, de Doctrine et d'Expérimentations*
CPX	Command Post Exercise
DEP	*Direction Etudes et Prospective*
DIRISI	*Direction Interarmées des Réseaux d'Infrastructure et des Systèmes d'Information de la défense*
DOTMLPF-P	Doctrine, Organization, Training, Materiel, Leadership, Personnel, Facilities, and Policies
DORESE	*Doctrine*, *Organisation*, *Ressource humaine*, *Equipment*, *Soutien*, and *Entraînement*
DS	Data Science
GAFA	Google, Amazon, Facebook, and Apple
HQ	Headquarters
HQDA	Headquarters, Department of the Army
IoT	Internet of Things
IT	Information Technology

LSCO	Large-Scale Combat Operations
MDO	Multi-Domain Operations
MGAT	*Major Général de l'Armée de Terre*
MMAS	Master in Military Art and Science
NATO	North Atlantic Treaty Organization
ORSA	Operational Research Systems Analysis
RETEX	*RETour d'EXpérience*
SaaS	Software as a Service
SQL	Structured Query Language
SVS	Shared Vision Statement

ILLUSTRATIONS

Page

CHAPTER 1

INTRODUCTION

Data Science, Artificial Intelligence (AI), Digitization, Autonomous Robots… are trendy concepts applicable and discussed in many domains. They arouse, on the one hand, fear or skepticism and, on the other hand, enthusiasm and hope in our future. This assessment is a starting point to this reflection, but what are actually those concepts referring to? Why are people and organizations interested in their use? What can be expected out of the related technologies? How do they link with military Art and Science and what type of consequence do they have on short, medium, and long terms? Such questions need answers before being able to move forward. This first chapter aims at defining these terms and concepts, their origin, and at raising the questions that this work aims at answering. An answer at first glance will also be provided in order to be refined throughout this academic work understood as a professional case study.

A Brief Historical Retrospective

Data and its importance to human endeavors can be tracked back to the beginning of history. If languages are among the first Data to have been made and processes by men, the invention of writing, first storable Data, is considered by historians as the turning point between pre-history and history and is dated around 5,000 BC.[1] Furthermore, Data has nowadays become part of our daily lives and can be defined as a piece of information exchangeable through sight, hearing, taste, smell, and touch as an

[1] John D. Kelleher and Brendan Tierney, *Data Science* (Cambridge: MIT Press Essential Knowledge Series, 2018).

artefact or as an immaterial intellectual object. Therefore, mostly unconsciously, humanity has become Data surrounded and most its activity consists in producing, exchanging, analyzing Data throughout life.

During centuries, human brains were the only mean to process Data. While some oral culture kept Data as a very volatile material, storage was mostly physically stored in libraries. Progress on the use of Data was slow until the European Renaissance and the birth of statistics and probability allowing analysis, sorting, and meaning-making out of Data. Mathematicians such as the Italian Gerolamo Cardano in the 16th century, the Frenchmen Blaise Pascal and Pierre de Fermat in the 17th century, or Adrien-Marie Legendre in the 19th century were pioneers in these fields.[2] In the 20th century, English mathematicians such as Karl Pearson or William Sealy Gosset accelerated the movement and brought knowledge as far as the ability to anticipate results thanks to probability, therefore allowing a form of prediction.

But progress was still missing in the field of storage. The birth of computing and numerical storage during the 20th century achieved a gigantic step for mankind, allowing storage of huge amounts of Data in a relatively small physical space; with an ability to process this Data and access it far faster and better than what it has been previously the case. Progressively, as quick as in a couple of decades, computers and networks have known constant hyperbolic performance progress. Computing skills became capable of storing, exchanging, and processing massive amount of Data in an ever-growing scale thanks to Data farms and supercomputers; therefore, allowing astonishing progresses that

[2] Kelleher and Tierney, *Data Science*.

has created the current situation: gigantic Data resources produced by numerical tools used as sensors and recorders to track everything related to our lives. In parallel, the birth of the Internet and mobile phone networks increased drastically the ability to access remote Data and exchange larger amount of it. The bandwidth of these networks currently allows the implementation of what is called the Internet of Things[3] (IoT) or Software as a Service[4] (SaaS) solutions as examples of the reached standards.

Fluidity, accessibility, and processability of Data as created a new era. Big Data has emerged and is characterized by its 3 "Vs": Volume, Variety, and Veracity. The ever-growing amount of Data has allowed numerical optimization to become part of the daily life of smartphone owners. People have started to hear about algorithms processing their Data and performing artificial intelligence. Companies such as the GAFA[5] are mastering the art of using Data to create autonomous software capable of predicting flows, recognizing faces or directing autonomous vehicles in dense traffic.[6]

In addition to the technical capacities, Data has now become a resource on its own, its mastering can be considered as a valuable good therefore making GAFA among

---

[3] The Internet of Things refers to equipment connected to low speed internet capable of reporting their current status. Refrigerators provide a self-explanatory illustration. Equipped models can, for example, order milk online when the stock is running low.

[4] In the Software as a Service model, the software is no longer physically stored and processed by a personal computer but remotely available through a network. As a consequence, it enables regular revenues for software developers.

[5] GAFA stands for Google, Amazon, Facebook, and Apple. These are among the most famous data related companies which business relies on the use of data.

[6] Pirmin Lemberger, Marc Batty, Médéric Morel, and Jean Luc Raffaëlli, *Big Data et Machine Learning* (Paris: Dunod, 2015).

the richest in the world. This always growing production has created a gigantic resource that is partly privately owned but also partly open-source. The expertise does not only rely on the ownership of Data, the science within algorithms is also gold nuggets. Algorithms consist of executing a set of instructions in order to achieve a task or produce a result. For example, a simplistic algorithm could be used to sort a list of names alphabetically. The combination of statistics, algebra, and coding has allowed algorithms to achieve more and more complex tasks such as recognizing a pattern on a picture like a cat or the numbers on a license plate, thus automatizing a task yesterday only performed by men.

Specialists have now mastered a technique called machine learning; algorithms optimize their own parameters by processing Data solutions. The term learning is used on purpose, it is just as if the algorithms were capable of learning on their own. In a way, the old dream of artificial intelligence is reached. More pragmatically, scientists have reached the lower level of intelligence. Human intelligence in its width and breadth is not already copied in its entirety but is definitely overmatched on specific tasks.[7] These techniques are implemented in our everyday lives and have already changed our surroundings and the way we behave, interact, communicate, and make decisions. After an industrial era throughout the 20th century, the 21st century is believed to mark the beginning of a numerical or information era and definitely impact our epoch.

[7] Lex Fridman, "Deep Learning State of the Art (2020)," MIT Deep Learning Series, Lex Fridman Artificial Intelligence Podcast + Clips, 10 January 2020, YouTube video, accessed 20 January 2020, https://www.youtube.com/watch?v=0VH1Lim8gL8.

Old and recent organizations are longing to implement these techniques in their core business as they identify powerful levers to optimize their effectiveness but also their cost or even the onerousness of some of their tasks. This implementation is known as digitization. It consists in implementing tools and processes that will allow the collection, storage, processing of Data in order to provision the workforce or the decision-makers with value-added information.

Military organizations, which are always searching for technological development, are among the most interested in Data science. Their need for technical dominance equals their appetite for quicker and optimized decisions. Moreover, it is interesting to remark that, just this once, armed forces' needs are not leading the technological progress but rather following civilian primacy in these fields. Understanding military positioning would not be complete without mentioning its ethical caution when it comes to the finality of autonomous weapon that can result from new algorithms.

Finally, the benefits armed forces can expect from Data science related skills are wide and the potential applications are numerous. Operational performance can be increased but the administrative aspect of the organizations is also concerned. Each warfighting functions have interests in the development of adapted tools. Intelligence, sustainment, command and control, and protection foresee applications which can change their efficiency or threaten their current doctrine. A technological race for dominance has

started opposing the Chinese model to the American one.[8] If developing such skills is a real technological challenge, implementing them is not an easier task.

Organizational Requirements

Implementing Data related techniques in organizations is no easy process. Several problems need to be solved such as: relevant and adapted Data collection through user friendly interface, Data storage and accessibility, development of suitable algorithms technics, exploitation of these algorithms, distribution of the optimized products or decision-making aid… Each of these tasks require a qualified workforce often rare and expensive. Each organization differs from another and can hardly organize itself in the same way that another would. Some companies have even developed an expertise in accompanying digital transformation processes to guide their customers. Military organizations do not differ from others. Their hierarchical structures or secrecy protection through different Data classification levels can even make the environment more complex to adapt to the very open-source mindset recommended by digitization. Taking into consideration the lethality encompassed within the military realm or nations' dependence on their military, the room for error is small.

The jobs related to these techniques are numerous and various. Network and Data base administrators, Data scientists (capable of statistics, algorithms coding, and Data cleaning), and software developers are some of them. Each are complex enough to be considered as a one and only job. The implementation is in itself an endeavor requiring

[8] Lee Kai-Fu, *AI Super-powers – China Silicon Valley, and the New World Order* (Boston, MA: Houghton Mifflin Harcourt, 2018).

an expertise that constitutes a job in itself. Project managers for digitization need a high technical mastering to lead successful change at affordable price. These facts allow a clearer understanding of Data Science and its actors. Data Science is here understood as a new domain born from the complexity of Data technics. The skills required are at the crossroad of many formerly autonomous sciences. Data Science is a notion born from the need for Data Scientists as people knowledgeable in the fields of statistics, coding, algorithms, Data representation, Data base, and others.

Another specificity of Data Science implementation worth mentioning is the leadership choices. Should the process be led from the above or below? Is the change originating from users or from a strategic choice? Are such reforms requesting a step by step change or, directly, a change as a whole? These questions can seem philosophical, nevertheless, they are actual and requires answers quite early in the process. Some facts on digitization state that the key to success relies in nearly every case on a user-led change and the creation of a project team empowered by its proximity with the leadership of the organization and skilled to act crosswise the organization.9

As a synthesis, the rise of the numerical age has created a great change. First, the storage of Data has become easier and requires fewer space. Second, the exchange of Data through networks has been drastically accelerated. Third, the ever-growing calculation skills of the electronic chip and associated computing power has created a vast shift of ability inn processing Data. Fourth, and as a consequence of the first facts, powerful algorithms have enabled autonomous complex activities yesterday performed

[9] Sopra Steria Consulting, *Livre Blanc de la Digitalisation* (Paris, 2016).

only by men and referred to as artificial intelligence, deep learning, or machine learning. These techniques are performed by specialists called Data Scientists. Implementing these techniques is a complex activity called digital transformation of digitization.

Data Science

Before going any further, some boundaries need to be defined. This thesis focuses on Data Science which is only a small portion of the overall digitization process. However, it is believed to be the heart of the numeric era as it includes the more value-added techniques among which figures Artificial Intelligence. Some additional characteristics that need to be stated about Data Science is the level of synthesis included in this function. It is a combination of statistics, coding, Data base management, but also algorithmic and many other skills. For all these reasons, Data Scientists are the heart of the digital process as their position allows them to overwatch the whole process and its requirements. Conducting the digitization of an organization requires such specialists and its complexity is understood to justify experts to conduct that job.

The French Army Situation

The emergence of computing techniques related to Data exploitation and their possible impact on the military effectiveness soon interested western armies including the French. At the same time, many questions dealing with ethics and politics applied to these means have raised and limit the will of going too far too fast. More pragmatically, the French Army takes its place within a Defense Ministry which holds responsibility for the overall organization of the forces. Therefore, for the French Army, no answer can be found on its own and a global approach needs to be applied especially as the scale of

Data is not without having some consequences on the effectiveness of the tools, meaning that a ministerial or a joint implementation including smart pooling of Data could produce more effective effects.

Being more specific on what motivates the French Army in focusing on this topic, the reader must acknowledge that an historical review on the evolution is required to create a better understanding. Throughout his evolution, man has been able to build an ever-growing dependency on techniques. The rise of smartphones era simplifies, personifies, and makes it easier to understand the reached point. Daily life has been deeply changed thanks to the ability to create Data, store, make available, sort, exchange, process, and exploit it. These facts are truisms. However, this evolution has been chaotic. Most organizations have tried to follow the movement by incrementally acquiring means and skills to follow progress and there is the problem. On the one hand, Data Science has emerged as a brand-new technology allowing machine learning and impressive artificial intelligence that can process Data that requires neither to be gathered in a unique platform nor to be strictly purposefully collected to be relevant. This progress, in a very large manner, has enlarged the scope of application of Data Science. On the other hand, many organizations focused on their core mission have just been using technical tools such as operations research / systems analysis (ORSA), or pre-developed purpose-focused solutions based on bundles of hardware and software allowing to take the benefits of a constantly-evolving domain without involving in hard-to-find balance. The recent progress has produced a change model called “digital transformation” or “digitization”; it aims at changing organizations from the inside to adapt them to the new numerical paradigm.

The French Army is one of these organizations. Since 2017, the French Ministry of Armed Forces has launched a Digital Transformation aiming at implementing the good assets, taking the benefits of new capabilities, and adapting to the new modernity that drives effectiveness and efficiency. The French Army is part of this reform as a major actor of the Ministry. However, for decades, it has owned tools related to Data such as an ORSA cell mainly focused on operational problem-solving, mathematical optimization, and simulation. More recently, since 2017, it has shifted from educating and recruiting operations researchers to Data scientists, thus acknowledging the understanding of the on-going change related to Data Science. The French Army also owns a small cell dedicated to digitization. It encompasses three main missions related to digitization: counseling the leadership, writing the ruling norms, and leading the development, experimentation, and implementation of new projects expected to lead the transformation.

What is this thesis all about?

Having stated this situation, the conditions are set to understand the questions that this thesis will try to address. Is the current situation satisfactory? Does it suit the problem? Can effective improvements be expected? How can the French Army optimize its organization to deal more efficiently with Data Science?

At first glance, it seems that the current situation can be improved. First, the Ministry's reform is allocating means and setting a framework that shapes the inner implementation of Data Science. This frame needs to analyzed in depth to identify the complementarity that can be built with the inner process. Second, the Data Science workforce is not only a technical resource, it might be the source of inspiration for developing new fields of exploitation, thus improving the overall effectiveness of the

French Army as Data Science can be expected to be. The initial step of developing experimental projects to initiate a broader reform can probably transition to a wider step of massive implementation. Third, indirect contributions to the digital transformation may be identified and easily implemented to provide support in the overall process.

Answering these questions will be conducted in a two-step process. The first one will encompass an extended literature review that allows a better understanding of the key terms and of the reasonable expectations for the French Army. The second step will consist of analyzing the U.S. Army's way to deal with that problem as the scholarship at Command and General Staff College offers to the author a great window on this organization. Without comparing models of different scales, and assuming that the U.S. Army can be considered as a reference in the technological investment and state of the art, some insights will allow a second round of refinement of recommendations allowing some practical mitigation to a theoretically applied in the first part.

Restrictions

The title of the thesis is, "A reflection on the best way to implement Data Science within the French Army." Even though the title is self-explanatory, the research topic deals with Data Science in its broadest sense, i.e. techniques related to the exploitation of numerical Data. It includes Data Analytics, Operations Research/System Analysis, Data Mining, Machine Learning, Deep Learning, Artificial Intelligence, and Digital Transformation of organizations. Before defining the terms, one can understand that the overall boundaries of the subject are very broad.

Data Science and the previously mentioned topics can become very technical. This study does not pretend to address these aspects of the subject as the author is not

qualified. In addition to this limitation, France's armed forces offer an environment framed by important though limited means and size that will naturally limit the effort led on Data Science. This academic work naturally aims at staying concrete and reality based.

Ultimate Remarks

The research project is significant to the military profession and other scholars because the process is ongoing in both French and US armies in a dimension never met before. It aims at providing concrete and applicable recommendations. The specificities of both the subject and the considered organization fills a gap in the literature and does not overlap any academic work. Therefore, this thesis can take its place within the numerous documents, articles, and thesis that have tried to sort and bring understandings to a complex (but not complicated) domain. The overall literature on the implementation of Data Science is very little indeed nonexistent. The research will attempt to fill this gap in the scholarly literature.

The author is qualified to explore this topic because of his engineering background, though not specific to computing, engineering provides him a scientific background which fits the technical nature of the subject. Moreover, he has been selected by the French Army to develop and acquire additional knowledge on Data Science and will probably take a full-time master's degree on this topic in the next few years. He is likely to hold positions related to this domain within French Armed Forces in the near future. However, this study is led without any biases linked to an already held function as it is fully based on the personal motivation of the author.

Recommendations 1 (R1)

Before transitioning to further analysis, finishing chapter 1 of the present thesis requires to formulate a first set of recommendations. The approach to these first recommendations should be that the current situation appeals for initial answers to the raised questions. This rough material will need to be further refined and/or amended. Three directions seem appropriate to be followed, they head to technical assets, manpower, and leadership.

First, the previous reflection has led to the assessment of the technical need for shared Data. Technical legacy provides the French Army with numerous Information Systems relying on dedicated Data base creating overlapping of Data, overload of entry for, and compartmentalization of the needs. The implementation of a Data Lake could reduce these frictions, the overall idea being to set the conditions for Data exploitation thus favorizing the Data Science processes implementation.

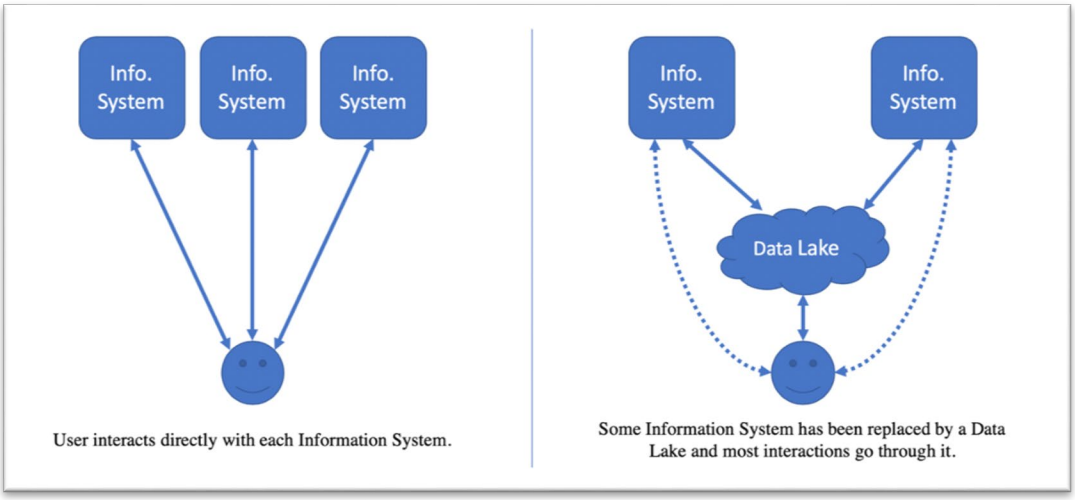


Figure 1. Data Lake Principle

Source: Created by the author.

Figure 1 depicts the general improvements performed by Data Lakes. Thanks to technology, all the information can be stocked in a unique place, which enables the most accessible collection and fastest use. The implementation of a Data Lake is a technical enabler of DS implementation. It sets the conditions allowing DS to be performed. The collected Data is linkable and massive. The algorithms can play their part. As organizations are always facing reluctance to change, it is essential to note that such a model allows a non-negligible decrease of redundant reporting for users. Data is valuable, increasing its accessibility will have positive effects. This recommendation acknowledges the technical difficulties raised by such an implementation. The main ones are managing the different levels of classification, adapting information systems to plug into this Data Lake, and allowing this lake to share with joint services.

Second, another important aspect of Data Science implementation is the skilled workforce. The French Army should recruit more Data Scientists and Data Analysts to effectively build its implementation. The trials conducted so far should be scaled to a larger use and recruiting is probably the most relevant track to achieve this acceleration in the implementation. This recommendation acknowledges the rarity of the resource and the prohibitive cost of such recruitments. Moreover, the need for technical specialists has to match the technicity of military organizations such as the French Army. Could the resource be found internally?

Third, organizing the implementation requires to think of suitable leadership. Creating a Data Science dedicated command appears to be a good solution to handle the numerous frictions and set abilities not only by advising commanders but by being able to act on their behalf. This command should also benefit from a specific place in the

hierarchy in order to be able to act throughout the whole organization. The quick improvement in effectiveness of the implementation depends on its scale and this is why empowering a unique command could help spreading the good practices and uniformize the processes.

This fist recommendations will be reconsidered throughout this thesis in order to be refined and completed if necessary.

CHAPTER 2

LITERATURE REVIEW

Foreword

The literature review explicitly describes the process followed during the documentation phase of the research. It articulates the intellectual process that has been conducted to perform this work. It is completed by the bibliography which gives a more exhaustive reference to the ideas that are presented in the following chapters. Due to the abundance of the sources and the time it needs to be sorted, readers of this thesis will understand that not all of the bibliographic references will be quoted in this chapter.

Data Science provides a wide-spectrum and large literary works that need to be analyzed carefully as it is a fast-evolving domain. Literature is quickly outdated or depicts an already passed reality. It can be technical and, as already mentioned, the author wants to limit the technical nature of this subject within the thesis. It can also be too vague or aiming at vulgarization. Finally, it can be very prolific as the subject is trendy and focused on many interests. The overall point is to identify relevant literature avoiding the two main pitfalls of being too generic, specified, or technical. Military relevance is also criteria not to be neglected, as the final goal should not be lost during the study. Going into too many details of each document that has been consulted would be too fastidious and counterproductive. Therefore, the aim of this chapter is to report the density and the main themes covered by the literature exploitation.

In addition to this chapter foreword, some comments are provided on the literature itself. First, it is mostly a technical literature based on the model of a science article. Second, there is a growing vulgarization literature which often explores

technology and the numerous expected and experienced applications. Most of the articles are related directly or indirectly to artificial intelligence and the dream or nightmare of uncontrolled machine: what can be named the "Terminator" effect. Third and last, there is an organizational literature which mostly depicts elements of strategy and applied or forecast doctrine. Following the first chapter logical path, the selected axes of analysis for the literature review will be successive focuses on Data Science, the French approach, prospective, and the U.S. approach.

First, the general topic of Data Science will be further depicted to build definitions and a more elaborate understanding upon which the overall research will be based. Second, the French Ministry of Defense's documents on digitization will be more precisely studied to develop the frame within which the final recommendations for the French Army could be done. Third, a focus on prospective is required as it fuels the understanding of the way Data Science could be useful and, therefore, guides implementation. Fourth, mirroring the second point on the French approach, the U.S. approach will be further investigated to highlight similarities, gaps, or specific points worth mentioning.

Data Science

Many of the definitions and related concepts have been previously defined in Chapter One. Nonetheless, it seems necessary to state that Data Science is an inclusive term which achieves the needed balance between being too specific or too generic.[10] This term notably encompasses Artificial Intelligence and all related notions such as Machine

[10] Kelleher and Tierney, *Data Science*.

Learning, Deep Learning, neuron networks, and other technical terms that will not be defined.

A process surfaces from Data Science. It starts with the collection and the storage of Data. New software or techniques such as Hadoop or MapReduce[11] allows the ability for mining a huge Data resource without a need of a too complex formatting thus revolutionizing the access to a Data base. Formerly, Data bases were formatted in very strict ways to be explored through SQL[12] language. The evolution implemented by the previously quoted software is therefore called no-SQL, explicitly referring to the progress of no longer being necessary to use the SQL mindset in Data bases.[13] Though, a cleaning and preparation of Data in its content more than in its form is still needed before further processing Data Science algorithms. Data Lake's concept referenced in Chapter One, is clearly allowed by this no-SQL mindset. Algorithms are logical loops that constitutes the engine of the process; this is the place where the actual benefit is done. More specifically, algorithms are at the thick of Artificial Intelligence even if the benefit comes from the whole process. After the result is obtained, it needs to be shaped in order to be presented to the user who is often a decision-maker. These steps can look easy and logical from the past few sentences. Nevertheless, they are the fruit of decades and even centuries of research in mathematics as well as development of powerful computing tools. This outcome of the literature reading provides a first understanding of the way organizations

[11] Lemberger et al., *Big Data et Machine Learning*.

[12] SQL stands for Structured Query Language.

[13] Lemberger et al., *Big Data et Machine Learning*.

should align to match the Data Science process in order to take the best advantages of its powerful skills. Collect, Store, Prepare, Process, Present can be memorized as the five steps of Data Science.

From an organizational perspective, Data-optimized decision knows no limit as far as field of application. It could be spread throughout organizations in their entirety as the benefits are often exceeding the expectations. Using different words to express it, one can see Data bases as mines in which are hidden golden nuggets of information. Data Science is the not only the digging tool, it is more to be understood as the tool that will manage the whole mine in order not only to extract nuggets but to produce ingots. Data Mining is also a term that can be found in Data Science related documents. It refers to the ability of gathering Data from large Data bases that were not created at the same period, not from the same purpose, and that are overlapping. The preparation of Data and the care to put on the collection of Data is therefore key to success. In addition, quantity of Data is a quality in its own and so implementing Data Science should allow the widest collection possible as its basic paradigm relies on Big Data. The more Data availability, the more optimization, automation, and artificial intelligence improved tools can be created.

The French Approach

Understanding the French approach is another important point of this literature review. As partly stated in Chapter One, the French military system and especially its Army component has inherited of an ORSA background. This science formalized during the Second World War by the U.S.A. and the United Kingdom, is nothing more than a mathematical tool to optimize solutions. The most famous historical cases are, first, the

optimization of the geographical settlement of the newly created radars in the southern part of England during the Battle of Britain; second, the optimization of the ratio of military escorts within naval convoys across the North Atlantic Ocean during the Battle of the Atlantic. Afterwards, these mathematical tools have mostly spread and developed within the industry. However, it has kept a non-negligible place within Armed Forces. More recently, it has, for example, been used to develop the logistical scheme for withdrawing materiel from Afghanistan in the years 2012-2014. Its benefit was optimizing the flow of materiel in order to maintain suitable military units available while optimizing the logistical flow of materiel back to France taking in account the limited time before redeploying them in Sahel and the difference of costs between air and sea routes. Within the French Army, simulation tools have been aggregated to ORSA skills, leading to the creation of a cell which main requirement was to manage mathematically complex and Data-centric topics.

Focusing on organization[14], ORSA was integrated as of 1956 within the French Army; Simulation appeared in the early 1980s-simulation office has not evolved significantly within the French Army model. Close to higher echelon of leadership, the workforce has varied between thirty to sic personnel.

[14] Antoine Naulet, *Les données : quels enjeux pour les armées ? Approche croisée s'appuyant sur le cas de l'armée de Terre* (Paris: École de Guerre, Promotion général de Gallois, 2017).

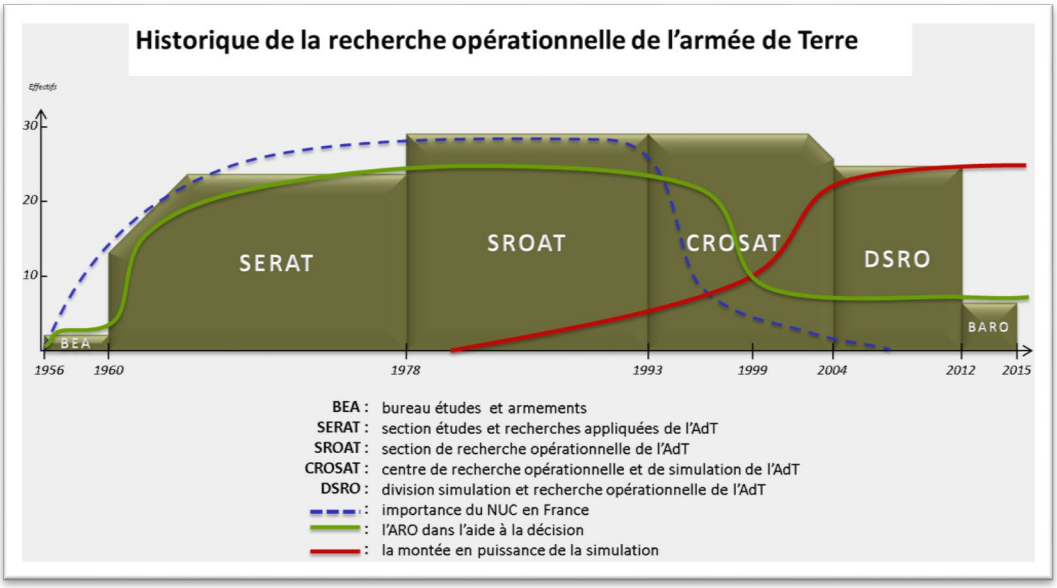


Figure 2. Evolution of the ORSA Cell within the French Army

Source: Antoine Naulet, *Les données: quels enjeux pour les armées? Approche croisée s'appuyant sur le cas de l'armée de Terre* (Paris: École de Guerre, Promotion général de Gallois, 2017).

Figure 2 represents three main pieces of information. The first one depicts the evolution of the name of the cell. The second is the evolution of the associated manpower. The third is the evolution of the balance between ORSA and simulation focuses.

Recently, as of 2016, the appearance of more and more computing ability to process Data has pushed the ORSA-simulation cell to turn itself to Data Science as an additional tool to an already existing pool of competencies; therefore, bestowing with another skill, an already heavily burdened cell. The risk of confusion and lack of clarity of the its missions is high and aggravated by the complexity of the subject they handle which is far from being clear to many.

In addition to this cell and as briefly mentioned in Chapter One, a policy officer dedicated to Digital Transformation also takes its place within the French Army

Headquarters (HQ). This colonel, directly reporting to the Major Général de l'Armée de Terre (MGAT), a typically French function that can be translated as a 3-star general acting as the chief of staff of the Army HQ, or second in command of the French chief of staff of the Army, owns the responsibility for three main missions already stated in Chapter One. This policy officer owns a critical place which does not put him a favorable situation. First, he relies on nearly no staff officer, therefore lacking the ability to explore in depth constantly evolving concepts. Second, he is not directly commanding the ORSA-Simulation-Data Science cell which is own by another branch of the HQ, therefore minimizing his role and ability to order or prioritize tasks. Particularly, he has no power to implement and change the structure and the organization. This policy officer, colonel ranked, constitutes an interesting nucleus for expanding the place of Data Science.

The digitization process initiated in 2017 by, Minister of Armed Forces, Florence Parly is very generic. It identifies the necessity of implementing a Digital Transformation and orders the whole Ministry to organize and starts this implementation. This reform is mostly limited to organizational and administrative boundaries. Within the French institution, the role of the Minister of the Armed Forces is built to report President of the Republic on the organization and administration of the Military, and to report to the Prime Minister who is in charge of the National Defense. Therefore, logically, the reform is more focused on how to frame the tools that will make the Ministry of Armed Forces digitalized. Main concerns addressed are Data pooling, Date Scientists recruiting, policy making looking for Data protection and privacy respect, and autonomous weapon and the ethical danger they represent.

As a synthesis on the current reform, this digital transformation is a policy that aims at the French defense organization's transition to Data led organization and identifies it as a key to effectiveness and success. It mainly aims at transforming the organizational part of the Army rather than its operational component which is naturally including the latest technological improvements on the battlefield since the battlefield digitization of the 1990s with the command and control structure relying on numerical tools to lead troops into battle.

As a global synthesis on the French approach, the French Army is, on the one hand, a subordinate actor of a digital transformation reform ordered and conducted by the French Ministry of Armed Forces which mainly focused on an organizational revolution. On the other hand, the French Army has historically built and aggregated an ORSA-simulation-Data Science capability that owns the advantage of its long 60-year existence but lacks of a clear position in the organizational hierarchy.

Prospective

The literature which tries to anticipate the future of Data Science is prolific. Most of it focuses on artificial intelligence, which is wide and skilled enough to be believed to be applicable to every field of action of the military organization. However, it is worth taking the time to explicitly describe the main paradigm behind this prospective literature: Data Science will go further than what is expected.[15] The progress relies on the fact that every single system of our everyday life should be developed to, among other

[15] Jean-Christophe Noël, *Intelligence artificielle : vers une révolution militaire?* (Paris: IFRI, 2018).

functions, produce Data in order to create a massive resource that can be exploited by increasingly powerful Data Science. It seems not exaggerated to talk of a new era as Data are no longer produced to achieve a specific goal but rather as a unique and precious resource that will feed the next progress whatever it aspires to achieve.

Without going in too long details, the main topics emerging from prospective readings are ethics, as far as men replaced by machines; ruptures in the evolution of human life on earth, as far as how much human life and organizations are going to change; but also, military applications, as far as how Data enhanced warfare is going to be decisive.

The U.S. Choices

The unclassified literature depicting the U.S. strategy related to Data Science and Artificial Intelligence is not surprising. The U.S. Armed Forces are acknowledging the gigantic potential of such Data-related techniques. The expected benefits are assumed to mainly contribute to ensuring the technological dominance of the U.S. forces and its allies. It is worth mentioning that, as a general trend, American literature is more positive than the French one which appears more suspicious about new technologies. To be more specific, it seems that the French culture is more reluctant on technological implementation than the American mindset which seems more technophile. This is no surprise when familiar with doctrine and military history where Anglo-Saxon culture differs from the French one.

Therefore, the U.S. Army is deeply involved in implementing Data Science throughout its organization in order to get the benefits of the use of its related techniques. The creation of Army Future's Command (AFC) is symptomatic of the U.S. involvement.

Even if not constituting a whole Cross-Functional Team (CFT), the U.S. Army has created an AI Task Force dedicated to the subject of AI. Another interesting initiative concerns the military Academy of West Point which has created an academic major on Data Science. On the contrary of the French model ORSA, Simulation, and Data Science are kept separated and the two first concept have kept their place throughout the organization.

As a synthesis, the U.S. Army has effectively shifted towards the future. The need to renew its major materiel program synchronizes with a doctrinal update resolutely headed towards Large Scale Combat Operations (LSCO) and Multi-Domain Operations (MDO). Data Science[16] is expected to take all its place within the future and the pursuit of technological dominance relies on it.

[16] It worth mentioning that most American publications do not use the term Data Science used this thesis. They broadly employ AI as a generic term more understandable for the general public.

CHAPTER 3

RESEARCH METHODOLOGY

Applied Professional Case Study

The method that will be used throughout this thesis is the Applied Processional Case Study (APCS). It has been developed by Dr. Kenneth Long from the faculty of the Command and General Staff College. The APCS methodology benefits form several years of iteration and, therefore, constitutes a strong methodological reference for the present thesis.

Figure 3. APCS Reference

Source: Kenneth Long, Ph.D., "The Applied Professional Case Study (APCS)" (Department of Command Leadership, U.S. Army Command and General Staff College, Fort Leavenworth, KS, 2016).

This method is defined by its author with the following terms:

> Case studies are a useful method for conducting qualitative research for problem areas that are human-centric, dynamic, volatile, and contain a mix of stakeholders, interests, variables and information concepts that demand a deep understanding of context in order to produce informed policy choices (Creswell, 2009, 2013, 2014; Yin, 2014).
>
> Case studies are appropriate and useful when the goal of the research is to inform or persuade policy makers about the rich and deep context of a setting associated with complex human issues where the goal is to take informed action to improve the situation. They make an explicit trade-off in favor of making informed policy choices in a strictly defined setting over the purpose of generating broad theoretical knowledge that has a wide applicability beyond the boundaries of the chosen setting. Case study designs look at how different types of knowledge and the associated methods of gathering, analyzing and making sense of information can be mixed to create a rich and deep understanding of the research area. (Hancock & Algozzine, 2011; Gagnon, 2010).
>
> Long (2016) describes the practical application of the case study method within the MMAS program over a five-year period. He provided a set of critical decisions that can guide a case study design to satisfy the purpose of either informing or persuading policy decision makers. By addressing each of the key points of that design model the MMAS case study can make a systematic, consistent and aligned argument for their research design."[17]

This method recommends to use five chapters to pursue the case study:

[17] Kenneth Long, Ph.D., "The Applied Professional Case Study (APCS)" (Department of Command Leadership, U.S. Army Command and General Staff College, Fort Leavenworth, KS, 2016).

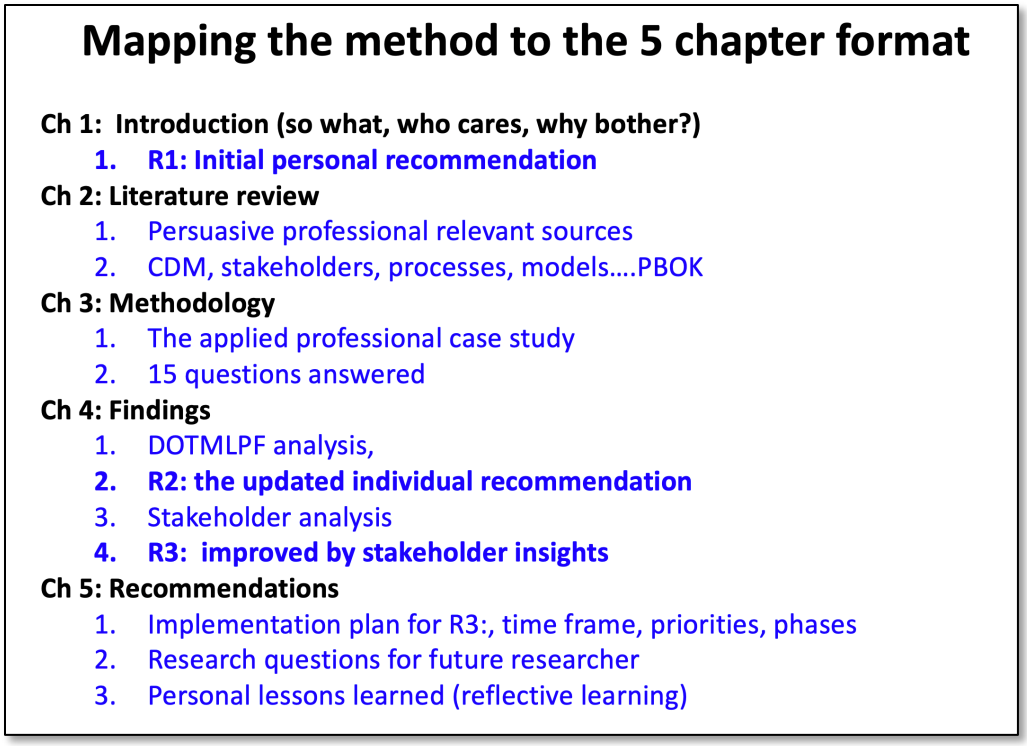


Figure 4. Dr. Long's 5-Chapter Format

Source: Kenneth Long, Ph.D., "The Applied Professional Case Study (APCS)" (Department of Command Leadership, U.S. Army Command and General Staff College, Fort Leavenworth, KS, 2016).

This method also frames the research by the following questions that will be answered in the following paragraphs:

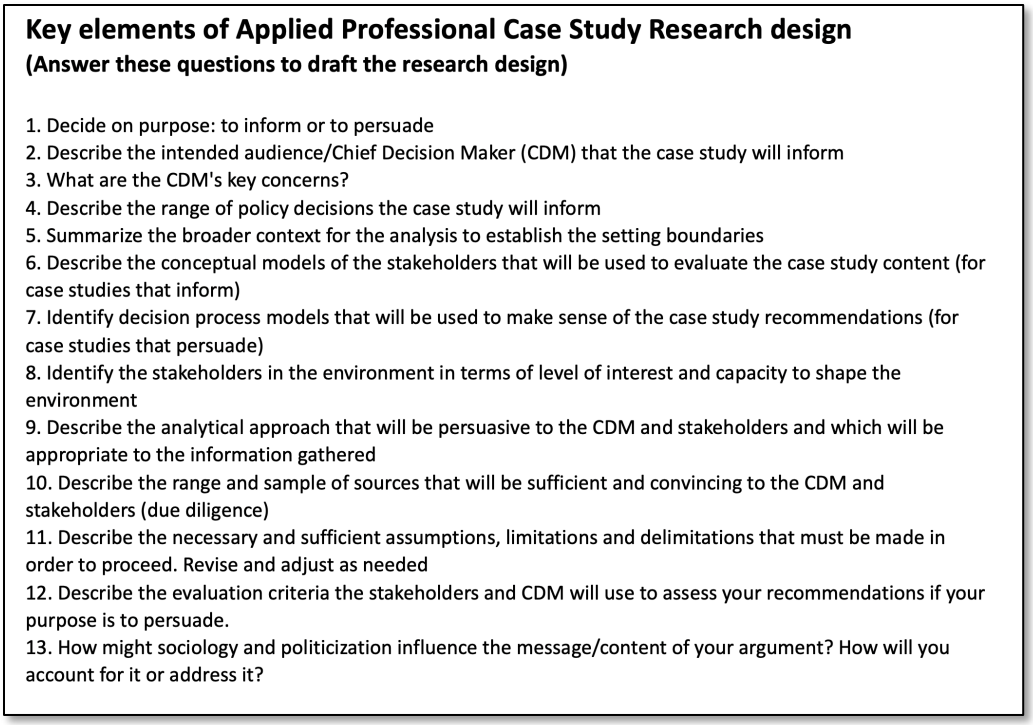


Figure 5. APCS Research Design

Source: Kenneth Long, Ph.D., "The Applied Professional Case Study (APCS)" (Department of Command Leadership, U.S. Army Command and General Staff College, Fort Leavenworth, KS, 2016).

The author has chosen to follow this methodology for its validity conferred by several successful MMAS thesis in the former years and because it perfectly suits the pursued objective of conducting a reflection on a subject of current concern while taking advantage of a scholarship embedded within the U.S. Army Command and General Staff College to draw comparison as it will match the concept of "stakeholder" developed in Dr. Long's APCS method.

Methodology

The method that is going to be used is mainly qualitative. It will rely on readings and theoretical knowledge that will support the analysis of Data Science and its

requirements to be implemented within the French army organization. This methodology based on a holistic approach to this new science will provide a relevant and exhaustive intellectual image of the subject.

Moreover, as the expected applications of Data Science are numerous and not yet totally framed because technology is rapidly evolving, DOTMLPF-P[18] methodology will be applied to explore implementation in the largest way. Doctrine seems far from our preoccupation but frames the whole use and the expected benefits. Organization is the heart of this study. Training is closely related to doctrine and is probably one of the ways to feed the algorithms of machine learning with easily gathered Data. Materiel will be one of the aspects of the proposal as the organization will rely on some physical assets. Leadership and education are also in the thick of the process as Data Science is by nature connected to decision making. Personnel is a major aspect as it is related to Human Resources that statistics and figures relevantly illustrates to better understand the sociology of the manpower. Data Science can doubtlessly reveal some interesting trends to act on. Facilities will be considered in the way of producing (Internet of Thing) and stocking (Data farms) Data. Policies will be addressed because the Army cannot be considered outside of its joint and organizational surroundings which claim regulation and compatibility.

Finally, as a way to refine the recommendation, the APCS method is advocating in favor of the identification of a stakeholder to confront the refined recommendations to

[18] DOTMLPF-P is an ordinary acronym within the U.S. Army. It refers to a Force Management concept and stands for Doctrine Organization Training Materiel Leadership Personnel Facilities and Policies.

an external point of view. The author has identified the U.S. Army as the ideal stakeholder to theoretically compare the ways of improvement identified in this work for the French Army to the choices made by the U.S. Army. This stakeholder comparison naturally acknowledges the differences between to institutions of different scale. However, the author firmly believes that this comparison is relevant and could produce interesting refinements. This access to the U.S. Army references on Data Science will mostly be achieved through research at the CARL.[19] comparison. Nevertheless, conversational type interviews with personnel from The Research and Analysis Center (TRAC) of AFC located on Fort Leavenworth post has been conducted and helped the author of this thesis to better understand the background and the way the different organizations articulate themselves together. Moreover, AFC has been tasked to lead the shift from the industrial era to the Data era which is perfectly in accordance with the thematic of this thesis. The Master in Military Art and Science located at Fort Leavenworth offers an ideal framework to use the U.S. Army as the stakeholder.

APCS Research Design

As already mentioned, the APCS offers a 13-question research design methodology. The following sub-paragraphs address these questions. The answers to follow are specifically enriching the Chapter Three of this thesis by explicating the use of the APCS performed in this thesis.

[19] Combined Arms Research Library of Fort Leavenworth.

Purpose of the Thesis

This thesis aims at providing recommendations to improve or optimize the implementation of Data Science within the French Army through tangible actions. Therefore, the purpose of this thesis is to persuade the policy officer in charge of Digitization to take these recommendations into account. Even though the main objective is clearly stated, the author acknowledges the information dimension of this work. Providing a developed analysis from an organizational perspective is the secondary goal of this work. It also aims at lightening the best way to use Data Science and propose a U.S. Army comparison as it can be considered as the most advanced and the best funded ally.

Intended Audience of the Case Study

The policy officer in charge of digital transformation and Data for the French Army is the first targeted audience of the case study. As he directly reports to the *MGAT* in charge of the overall organization of Army HQ and the French Army, the author acknowledges that ideas and arguments from this thesis are indirectly addressed to the *MGAT* and further on to the *CEMAT*[20] (*Chef d'État-Major de l'Armée de Terre*). However, in a logic of respecting the hierarchy, and as the policy officer has the task to propose the suitable organization, it is logical that he is picked to be the Chief Decision Maker (CDM) of the APCS methodology.

[20] The *CEMAT* is the, 4-star French general Chief of Staff of the Army.

Key Concerns of the CDM

The CDM's main concerns consist in looking at three different directions, all of them of the same importance and challenging his function. The first direction deals with his perimeter of responsibility. He is the primary advisor of the MGAT and the CEMAT on digitization and Data Science. He is in the same time leading two small teams; one in charge of implementing experimental digital projects that are actually creating the digitization, and the other of the Data policy of the French Army. These two fields of competency are very time consuming and can be considered as a burden for a policy-officer who does not rely on any staff.

The second concern is its position within the hierarchy. He is directly attached to the MGAT as an advisor but lacks staff power to perform real tasks. Moreover, most of the Data related units of the Army, such as ORSA, are not under his direct control. Finally, even while being a colonel with a senior task, he does not weight as a general officer to convince his higher hierarchy.

The third and last concern is the joint and Ministry of Defense (MoD) perimeters of his tasks. The joint and ministerial levels are requesting the French Army to implement the reform. The French Army is expecting its Data policy officer means to endorse these actions with only limited means and a great deal of action to be performed.

To synthetize, the CDM is overstretched between the Army, the Joint and Ministry levels, the actual and time-consuming conduct of implementation of experimental Data related projects, and writing the Data policy of the French Army. All of this prevents him from being able to successfully convince the Army's higher

command to create necessary change. This thesis therefore aims at providing him support by adding to his arguments some insights from an external analysis.

Range of Policy Decisions Informed by the Case Study

This case study mainly focuses on the organizational level and all its components. The targeted subjects are the nature and volume of the required workforce, its organization and its just location within the French Army hierarchy to reach the expectable gains offered by an efficient use of Data Science. The recommendations that will be stated in this thesis should provide structured and convincing arguments in favor of a helpful and adapted though not costly reorganization.

Broader Context for the Analysis

The boundaries of the subject are framed by the French Ministry of Armed Forces' ongoing strategy. There is no question either on the will to be successful, or the clarity of the MoD digitization. However, it seems that there is a gap of understanding on the opportunities that arise from these new techniques within the Army. In addition, the Army wants to stay at the top of technological capabilities to comfort technical dominance of its possible adversaries. It means that the will to implement Data Science is not at stake, it has been already decided and needs now to be implemented. This is why the subject is really looking at the optimization of this implementation mostly from an organizational perspective. The analysis is not going to go any further than a global approach to understand the components of a successful implementation.

Conceptual Models of the Stakeholders

Stakeholders can be sorted in two main categories that will approach Data Science from two very opposite states of mind and conceptual model. Data Science stakeholders are for the most part understanding their techniques as key to improve efficiently the processes and results throughout the realm of their organizations and more generally of their surrounding life. Most of them are not afraid of the consequences of using Data Science as they understand how it works and see neither phantasmagoric nor unethical consequences of its use. They are firm believers who tend to minimize the technical aspects and then show such enthusiasm that they may frighten their counterparts. At the opposite, leaders with no scientific background tend to be contemptuous of technical tools and minimize their potential impact. As a consequence, they show fearfulness in allowing effectively more space to Data Science. In addition to this natural tendency, they react skeptically to Artificial Intelligence that they often mix up and misunderstand.

Decision Process Model

If this thesis convinces the policy officer, he may decide to use its arguments by proposing an informative proposal to his higher command and more specifically the MGAT. If positive feedback is given, it may then go on a decision briefing with a board of interested stakeholders such as the CEMAT, the ORSA cell, and any cell dealing with Data that may belong to a Data Science department.

Identified Stakeholders

Starting from the policy officer as the identified CDM of this case study, several circles can be identified. First is the Data Science community, which can be found within

the ORSA cell located in another division of the Army Staff or outside of the Army at the joint or interagency level. This circle is assumed to be supportive and considered as biased by Army leaders. Second is the external sphere in which civilian leaders are the ones who may be in contact with the military world like the defense industry, but also other militaries already involved in that domain such as the U.S. Army. Their testimonies and/or support have to be understood as a lever to convince that Data Science is a tool that needs to benefit from a specific implementation organization. Finally, the hierarchical surroundings of the French Army with the Minister and the joint framework is characterized by its will to achieve a digitization. It frames the whole process as it has clearly a head start on the Army.

Analytical Approach

This thesis will be read and judged through several lenses. The first one will be the persuasiveness of the argumentation making sense out of technical thus dubious tools. The second lens will be the realism of the proposal, especially in the cost domain both human and financial. The third lens will be the comparison with the U.S. Army model as it defines a real reference when it comes to organization and interoperability. The expected quality of this work should be clarity and realism to provide exploitable results. In the best case, it is possible for this work to be presented to the Chief of Staff of the Army to convince him of the possible benefits that can be taken out of Data Science.

Range and Sample of Sources

The expected range and sample of sources that may positively support this thesis can be defined by its resulting clarity. Open sources precisely analyzed and synthetized

would offer a solid argumentation in favor of the recommendations. Sorting and explaining historical layers and analyzing expectable applications of Data Science in the French Army will support this thesis. Academic works on the subject are mostly accessible through libraries such as the Combined Arms Research library of Fort Leavenworth. The U.S. Army example can be analyzed through open source publication such as organigrams and publicly published strategies.

Assumptions, Limitations and Delimitations

The main assumption on which this works relies is that Data Science implementation can be improved within the French Army. Moreover, it is assumed that leaders may be convinced by arguments on a more centralized Data Science resource and would have the ability to implement such a change without interfering with the ministerial and joint implementations. Therefore, a clear delimitation is set on the French Army and this work does not aim at proposing recommendations to any other organization. As far as limitations, this work being led from an academic perspective will miss the factual and exhaustive approach to be implemented without additional work. It aims at providing an intellectual support through argued recommendations.

Evaluation Criteria

This thesis and its recommendations will be evaluated through the lenses of several criteria. The first is organizational. Military structure is always evolving and every domain claims the legitimacy of a dedicated branch as close as possible to the top leaders of the organizations; this thesis' recommendations are probably going to go in that direction. The structure evolutions should be balanced and justified. The second lens

is result focused. Data Science is mostly attractive by its expected gains of effectiveness; both administratively and operationally. The recommendations should illustrate the variety of the fields of application. The third and last lens is financial. Resources, both human and financial, are limited. The recommendations should be realistic and promote efficiency.

Sociology and Politicization Influence

This thesis will not address any form of politics and will purely stay within the organizational realm. Being mostly focused on the French Army, it will have low impact on the sociology itself even if Data Science can be excepted to simplify the load of repetitive work of the French Army workforce on an everyday basis.

CHAPTER 4

FINDINGS

This chapter's purpose is to elaborate on the results of the research. The first three chapters of this thesis presented the purpose of the study, defined the topic of the thesis and established its boundaries. Chapter 4 will present the findings to the research questions. The APCS method proposes a two-step methodology. The first consists in a DOTMLPF-P analysis to dissect the different approaches. This analysis will lead to an update of the initial recommendation that was stated in Chapter 1. The second step will consist of a stakeholder analysis to complete and supplement the first results. As already mentioned, the selected stakeholder is the U.S. Army. This will result in a final refined recommendation that will conclude this chapter.

DOTMLPF-P Analysis

The notion of Data Science is going to be processed through DOTMLPF-P to illustrate the scope and scale of the benefits that can be expected in the French Army. This analysis aims at providing both an understanding and also some insights on the associated required organization. To be more specific, Data Science is the object of the analysis and organizational concerns related to the French Army are a perspective that will be systematically added to each domain analysis.

DOTMLFP-F is a tool mostly used in Force Management to develop a solution to an identified gap. It refers to the French acronym *DORESE* (Doctrine, Organization, Human Resource, Equipment, Sustainment, and Training). This thesis considers the

optimization of the implementation of Data Science within the French Army as the subject of the following analysis.

Doctrine

Doctrine is constantly evolving and reflects the theoretical use of an Army. The actual writers and actors of doctrine are spread throughout the French Army but the *Centre de Doctrine et d'Enseignement du Commandement* (*CDEC*) is the organization that is responsible the doctrine process and acts as the final authority to validate and publish the documents. It is nested in its joint equivalent the *CICDE* (*Centre Interarmées de Concepts, de Doctrine et d'Expérimentations*).

Doctrine is defined out of experience, best practices, new materials, lessons learned (*Retour d'Expérience* or *RETEX* in France which slightly differs from the American concept). Those multiple sources constitute large amounts of Data; therefore, Data Science can be helpful. Data Science can be expected to support this process by performing an analytical study of the different sources that build doctrine. This support can only be expected if numerical Data are massively produced in training activities to depict the way activities are performed. The creation of a Data Lake would absolutely match the intent of gathering training related information that could directly match doctrine purpose.

Data Science support to doctrine is mostly indirect as it is not expected to perform the redaction itself. Field experimentations or computer assisted implementation are great Data producers that can be used to test the effectiveness of a choice or its relevance as far as integration within the already existing related doctrines. Moreover, simulation software using doctrine as a parameter of an Artificial Intelligence to automatically

execute missions as a virtual opponent or subordinate can be a great tool to test new doctrine. This referenced technology is already partly implemented in the French Army through the SOULT software that is used as a simulation tool for Command Post Exercise (CPX). Before being implemented, doctrine can be tested and compared to other or former doctrines.

Training activities especially in national training centers, could provide their support by allowing an enhanced production of Data. Best practices, or successful patterns could be identified and helped the development of not identified doctrine. The optimization of behaviors even with a complex number of parameters is definitely in the skills of Data Science.

Organization

Organization is one of the main focus of the present thesis as it studies the improvements of the implementation of Data Science. However, to avoid any confusion this paragraph focuses on the benefits that the French Army organization can get out of Data Science; not on the way to organize Data Science. Administrative tasks are here understood as being included in the notion of organization.

As already stated, Data Science can produce its effects in several fields of activity. As far as organization, Data Science can be applied in a purely organizational way through the different networks of the Army to support the administrative numerical life of the Army. As earlier explained, the world of Data base has known a revolution thanks to the new powerful tools to research Data base such as Data Mining. This really puts at stake the concept of task-dedicated information systems. Data Science capabilities have surpassed the old information systems concept which was relying on dedicated Data

base, exploited by a single software, and requiring its own Data entries. Information has now become a shared value of which mutual use benefits the efficiency of the whole organization. The Data collected for a specific purpose should also be exploitable for other purpose, thereby saving time and energy spent several times so far. The example of administrative proof of entitlement speaks to everyone especially military personnel frequently moving. Data Science can be expected to improve that situation, bringing administrative repetition to an end.

Data Science also allows for focus on a large collection of Data through the multiplication of sensors. Connected objects, geo-localized sensors, or more simply activity trackers embedded in smartphones of desk computers could perform these reporting tasks automatically. Data Science is capable of exploiting this Data, thus optimizing everyday effectiveness of office work such as task organizing, statistics providing, or automating repetitive tasks. This aspect of Data Science is fully part of the ongoing ministerial reform that was launched in 2017: digital transformation.[21] Even if this reform is led for the entire minister, the Army needs to adapt it to his reality by taking its part in developing the especially relevant Data collectors. Easing users' life is a great criterion to develop or prioritize new projects. Reporting actions, often manually and repetitively typed, are great targets for Data Science.

New tools have to be implemented throughout the entire organization to make it more numerical and increase the size of the Data. The related improvement is expected to be incremental. Easing users' experience is the guideline to achieve this digitization

[21] French Ministry of Defense (MOD). *Orchestrer la transformation numérique du ministère des armées* (Paris: MOD, 2019).

properly.[22] Classification of Data raises a non-neglectable concern in term of sharing. There is no general answer to this problem as anonymizing Data is made less and less realistic by Data Science and cross-checking methods. However, this limitation should not prevent the implementation of Data Science but rather push the French Army to rethink its classification model. This point will not be further expanded in this thesis.

An additional comment should be developed on the consequences of such a reform. From an organizational perspective owning the information has been a sign of authority; digitization is putting that paradigm at stake. Data is fueling the organization and its effectiveness and efficiency. Hierarchy can be expected to be influenced by such a reform as the information is going to shape differently the environment. Even if it is hard to elaborate very precisely on this aspect, it is important to note that Data Science enablers, even if technicians, should find their place close to the commanders as their action shapes the decision and therefore the whole organization.

Training

Training is the Army's tool to create combat effectiveness. Assessing it is key to a commander who owns the readiness decision. Progressive schedules and mandatory activities are nowadays defining standardized training activities. They are constructed out of experiences and expected operational engagements. Data Science can improve this situation by personalizing training. Instrumenting training could allow the creation of massive Data related to training. Data Science can surely identify causality links or relations between training events and operational results. Therefore, taking in account as

[22] Sopra Steria Consulting, *Livre Blanc de la Digitalisation*.

many parameters as possible can be usefully taken in account to optimize the training. Past experience, formerly performed training or operational activities, qualification levels of the personnel manning the unit are examples among many others that actively take part in the combat effectiveness' assessment. Matching the training activities to the optimized combat effectiveness benefit could save time and money. This analytical approach is not expected to disqualify the leaders who righteously own the final assessment on readiness but to provide them with a decision tool enlightening their decisions.

As already mentioned, instrumenting training activities to obtain more Data can also have effects on doctrine. A good example is geo-localizing members of a unit in a training center. Relations between movement success in mission can identify successful courses of action or most effective tactical procedures. Data Science can be of great help to analyze training. Data Science can also be used to source training and, more specifically, in the best way to optimize the resource.

Materiel

Reflecting on the implementation of Data Science within the French Army is basically a non-material activity from a Force Management perspective. It mostly concerns organizational matters and human resource. However, servers, computers, Data Lake, and the rare and costly qualified workforce can be considered as a materiel solution to solve an information gap. In addition, Data Science has a role to play in the performance of new materiel acquired by the French Army. Therefore, this paragraph covers several aspects. First, the implementation of Data Science in the materiel of the Army and the related operational and sustainment benefits will be explored. Second, the focus will turn on the materiel Data Science is relying on.

Starting with the materiel aspect, developing Data Science compatible materiel is a source of operational improvement. Collecting and sharing Data from the operational surroundings of actors such as dismounted soldiers of vehicles can be useful in terms of intelligence and decision making. The collaborative combat expected out of the French Scorpion program can be associated with this concept of Data enhanced battlefield. New generations of materiel are also believed to enable predictive maintenance and logistics. Forecasting a breakdown before it happens or optimizing the stocks are already possible. Artificial Intelligence application can allow drone swarms, autonomous robots capable of following units, or even autonomous cars leading to terrestrial drone.[23] However, those last points are still under experiments and are exceeding the boundaries of this study. From an organizational perspective, such evolutions argue in favor of, firstly, deploying Data Science teams within the Force Management process, and, secondly, systematically mentioning Data Science compatibility to functional specifications of new materiel. Moreover, the ability to deploy Data Scientists in operations would probably be a notable consequence to be forecast.

Moving to the materiel Data Science is relying on, Information Technology (IT) equipment is the largest part with which Data Science is performed with materiel such as computers or servers. It is complemented with algorithms that are made of on-the-shelf purchased capabilities and, in certain cases, super calculators. Therefore, implementing Data Science from a materiel perspective requires a slight effort on IT equipment as far as calculation is required. However, those systems are part of a larger plan of equipment

[23] Jean-Christophe Noel, "Intelligence artificielle: vers une nouvelle révolution militaire," *Focus stratégique*, no. 84 Ifri (October 2018): 82.

that is not only supporting Data Science. As far as cost, in comparison to other equipment program, Data Science materiel costs are very limited not to say neglectable.

Leadership

From a leadership perspective, Data Science can be considered as a new tool to support decision making. The implementation of Data Science can materialize in the fields of leadership as follows. A first step could be implemented in military education programs where Data Science benefits could be exposed in their capabilities and limitations. Commanders could also be counseled by Data Scientists who could take their place within staff. The load of work represented by this advising task can probably lead to a dual hatted function within the staff. Without any doubt, leadership benefits from Data Science quicker analysis but should be careful with overreliance and dependence which can lead to lack of questioning and loss of responsibility.

Personnel

Data Science can be very beneficial to the personnel aspect of the military. It is, for example, a tremendous tool to support human resource management. At the macroscopic scale, the flow of men and women are more easily managed when their general behaviors can be forecast. Workforce retainment within a professional Army is key to build operational effectiveness and can be better understood thanks to Data Science. Managing necessary movement of numerous executive leaders to homogenize the organization is another necessity that Data Science can be supported by Data Science. At the microscopic scale of the administered people, Data Science can allow a more

tailored management, by handling negative consequences so far banned by general national policies.

Turning to Data Science workforce, Data Scientists constitute a rare and costly resource. Being Data Science qualified requires time investment. Implementing Data Science within an organization does not only require Data Science knowledge but expertise on the core skills and tasks of the organization. This is the reason why the Army needs, in the same time, Data Scientists capable of military expertise. This double competency is hard to develop. Data Scientists jobs can be complemented by Data Analysts who owns limited skills but suitable to use a system that is already deployed. These two levels of qualification argue for more flexibility and provide insights on the way to organize the Data Science Workforce. A core of Data Scientists could beneficially be complemented by Data Analysts. One can then imagine a core of centralizes Data Scientists and Data Analysts compensating for them in a decentralized way. Concentrating a rare resource is probably the only way to afford Data Scientists. Contracted support complements the requirements and reinforce the capabilities of action. A third or fourth level could even envision when thinking of sensitized personnel trained to relay the action of experts in units or HQ. Organizing these four levels in an alternation of centralized and decentralized activities offers a flexible course of action to develop a viable implementation. Flexibility and ability to weight efforts and to act in a prioritized way seem to be the more suitable way to implement Data Science within an organization. As a synthesis, the envisioned taxonomy of Data Science personnel is Data Scientists, Data Analysts, contracted specialists, and sensitized internally trained personnel.

Facilities

Facilities have no direct link with Data Science. However, by extrapolating the need for large amount of descriptive Data, facilities can play their part in favor of Data Science. The IoT offers an interesting way to equip facilities with sensors capable of complementing the Data reporting by transmitting Data of the everyday life.

Data Science also requires additional facilities as specific places to install servers supporting a Data Lake. Such infrastructures are very energy-consuming facilities. Additional reflection on their implementation could be required. Although Data Science is by nature promoting Data sharing, additional protection of farm of servers has also to be considered from physical and cyber standpoints. Such types of concern must be considered as to the early steps of the digitalization process.

Policies

Last but not least, policies should support Data Science by providing a favorable environment which helps to deal with the complexity of Data. As already mentioned earlier, levels of classification are a major impediment to the sharing of date. Specific policies have to be implemented to handle properly this difficulty. Protecting the privacy of Data is also an imperative for the French Army even if the military statute gives specific rights to the Army as far as Data ownership is concerned. As far as the Army environment is concerned, implementing Data Science implies to stick to the internal regulations of the Ministry of Armed Forces, but also to national and supranational ones such as the European Union. This problematic has already clearly been identified and is currently addressed.

Recommendations 2 (R2)

This first holistic analysis reveals the exhaustive nature of Data Science. The revolution induced by its employment is not only a matter of topic or field of application, Data Science can literally be used everywhere, for limited costs and potentially great effects. The framework of this thesis shapes the recommendations: the French defense is already conducting a digital reform and its Army component inherits from ORSA background. The result of this analysis leads the author to a couple of refined recommendations compared to the ones stated at the end of Chapter 1. Recommendation 2 is going to be explicated in four sub-paragraphs as follows; one, refine the initial recommendation (R1); two, define a singular path within the ongoing reform; three, defeat distaste for technology; four, dare a new organization.

Refinement of R1

Recommendation 1 was developed in three main domains; implantation of a Data Lake, acquisition of a skilled workforce, and creation of a dedicated command. These three ideas have mostly been confirmed by the exploitation of the literature and the DOTMLPF-P analysis conducted in Chapter 2 and first part of Chapter 4. Nevertheless, it is necessary to briefly recapitulate the identified complements to the initial ideas.

Concerning the implementation of a Data Lake, the idea has been confirmed and some additional conditions have been identified. A Data Lake need to be following policies including the protection of the privacy of Data and the different levels of classification. Therefore, instead of a single Data Lake, several would surely need to be created to match the different levels of classification. Data lakes will also require additional protection both physical and cyber.

Concerning the skilled workforce, the need for experts to be able to master simultaneously Data Science and the French Army military specificities has led to the identification of four levels of qualification; Data Scientists, Data Analysts, contracted specialists, and internally trained non-specialist actors. These actors should act in a succession of centralized and decentralized activities allowing prioritized actions and weighted efforts.

Finally, concerning the dedicated command. The idea has not been specifically challenged. Needs for being able to counsel properly commanders confirm the importance for the dedicated command of being empowered by its hierarchical position. As a consequence of the lack of reference on this specific point, the study of the U.S. Army should focus specifically on this organizational point and lead to confirmation or denial.

Define a SINGULAR PATH within the Ongoing Broader Reform

The broader framework offered by the French Ministry of Defense is a great opportunity for the French Army.[24] It aligns a whole system on the new era of Data considered by many as having the scale and scope of a new industrial revolution. Following this reform is not only a duty and a task, it is a starting point based on the smallest common denominator. This is why, from an army perspective, this reform should the starting point for further reform. In no case it should be considered as sufficient for the French Army. The idea of going further does not aim at jumping into science fiction, but at betting on the success of Data Enhanced organization that has

[24] MOD, *Orchestrer la transformation numérique du ministère des armées*.

already been proven as successful within the civilian world. However, the peculiarity of an army requires a personal adaptation. If the Ministry reform is mostly organizational and administrative, the Army should consider implementing Data Science in more than its staff works. Core activities such as training, human resource, intelligence, sustainment are fields of application that could benefit from these new technologies. This point is therefore not a contradiction but a supportive point to the ongoing reform. The choices so far are the good ones but should be accelerated and intensified.

Defeat Distaste for Technology

The French Army, because of its history, its national mindset, and the academic backgrounds of its leaders, tends to see war more as an art than a science. This tendency can interfere with the implementation of techniques that can be associated with scientism. Artificial Intelligence, more as a word than as a reality, raises fears even if the current skills are more of an automatization than an actual duplication of human intelligence. Backed by ethical legitimate arguments, opponents to Data Science exists, often misunderstanding the reality of can be done and how. In the style of Colonel Ardant du Pics, they mostly argue than men are the only valor in the organization and that they are key to success and the keystone of resilience whereas technical tools are illusions of solutions but in fact only tools. The author of this thesis shares this idea. The real treasure of the French Army is its men and women, not its Data. However, the French Army is based on a technological dominance to deter and defeat its opponents. This reality cannot be done by halves. Today, technical dominance proceeds from quicker decision-making and better optimized solutions which is exactly the realm of Data Science.

The impulsion should be given from the top leaders and a tool to achieve that would be to add a science history program in the military academies. Military thinkers are studied to build an understanding of the evolution. General culture is also often learned through the lenses of historical evolution. Science is often neglected although it has accompanied and shaped the evolutions of our societies. To be more precise, after passing the exam of the *Ecole de Guerre*, the author of this thesis was astonished not to be able to assign a date, a domain, and a major discovery to each names of supposedly famous French scientists painted all around the first level of the Eiffel Tower.[25] This reality is a shame. Officers are not only thinkers, they are also actors and therefore should master acting and related science application.

[25] As a friendly challenge to the reader, the author reproduces the list of these names as they appear on the monument: Seguin, Lalande, Tresca, Poncelet, Bresse, Lagrange, Belanger, Cuvier, Laplace, Dulong, Chasles, Lavoisier, Ampere, Chevreul, Flachat, Navier, Legendre, Chaptal, Jamin, Gay-Lussac, Fizeau, Schneider, Le Chatelier, Berthier, Barral, de Dion, Gouin, Jousselin, Broca, Becquerel, Coriolis, Cail, Triger, Giffard, Perrier, Sturm, Cauchy, Belgrand, Regnault, Fresnel, de Prony, Vicat, Ebelmen, Coulomb, Poinsot, Foucault, Delaunay, Morin, Hauy, Combes, Thenard, Arago, Poisson, Monge, Petiet, Daguerre, Wurtz, Le Verrier, Perdonnet, Delambre, Malus, Breguet, Polonceau, Dumas, Clapeyron, Borda, Fourier, Bichat, Sauvage, Pelouze, Carnot, and Lame.

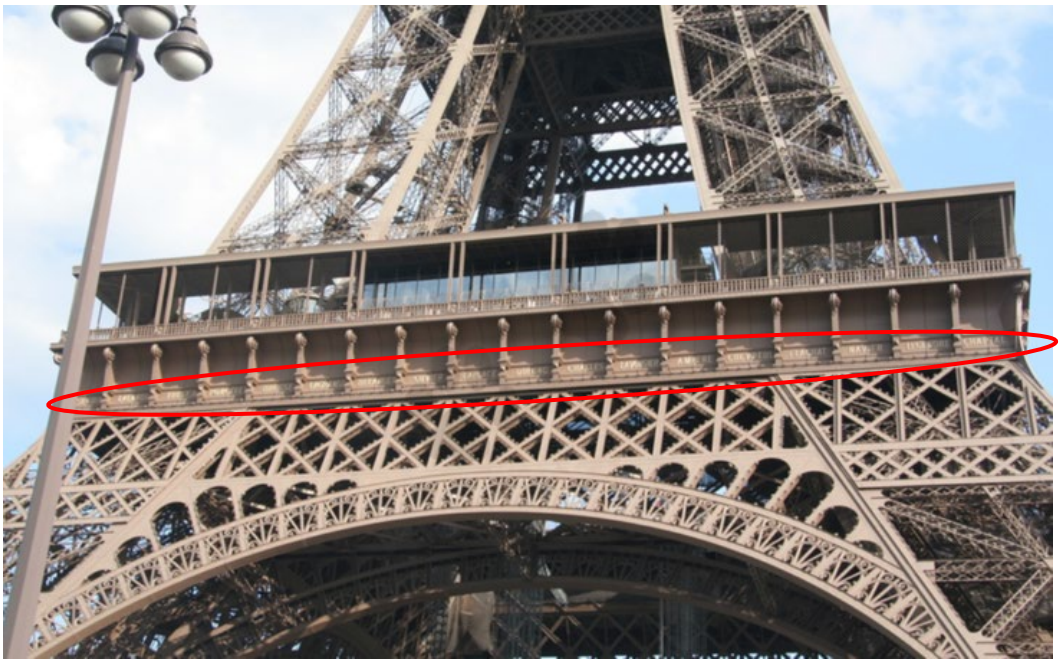


Figure 6. Eiffel Tower with Scientists' Names

Source: Personal collection of the author.

Dare a New Organization

ORSA can be considered as a mathematical tool to optimize decisions. Simulation is a way to provide support to leaders. Developing digital tool to facilitate the everyday life is great to build a Data enhanced organization. Dedicating a whole office to the regulations related to Data, its storage, its protection, and its use is definitely a necessity. Having a general officer in charge of the numerical transformation would be a good proof of the importance recognized in Data.

Although these tasks are all useful and good within themselves, there is a need to federate the talents and improve the organization to enable the benefits. The French Army could create a whole cell within its headquarters to deal with Data, its needs and requirements. This cell should be manned enough to be able to implement Data Science.

Centralized Data Scientists could provide their support to every field and organize, administer the common tools, and regulate the domain. Data Science is cross functional and should therefore occupy a resulting position. Its ability t*o interact with all the domains of the Army should lead to the decision of putting a general officer at the head of this cell with enough experience and credibility to be able to interact with the heads of other divisions of the headquarters.

Finally, as our previous analysis has identified a very extended spectrum of application for Data Science, it seems relevant to acknowledge that a new era has started with Data. Is it a new paradigm? The author of this thesis believes it. Whatever word, the Army wants to put on this new domain. It is highly recommended to admit the significance of Data Science as well as its cross-functional characteristics and its ability to cover all domains and be a benefit for all. The last associated recommendation is to identify Data Science as a game changer that requires a voluntary appropriation.

U.S. Army Analysis

Being embedded for one year as a student of the Command and General Staff Officer Course at Fort Leavenworth, Kansas is of no value to witness the U.S. Army and understand its doctrine and approach to the current and future understanding of war. When it comes to prospective trends and technologies, the U.S. Army is known for its voluntary effort to keep the important advance it has acquired at the cost of massive investments and resolute budget efforts. The American Way is also renowned for being technology friendly and for its willingness to rely on a technological dominance. One can still wonder if the U.S. Army choices are relevant for the French Army. An answer could be that the unanimous recognition of the U.S. Army as the world leader in the field of

defense seems to be a sufficient proof of reliability. Moreover, NATO[26] and the SVS[27] between the U.S. and the French Army are respectively multilateral and bilateral agreements identifying interoperability as a key of success for cooperation. Therefore, more than a subjective view of the author, the U.S. Army model is relevant to the French one and Data Science makes no exception. The need for an enhanced compatibility between these two armies is assumed by the two parties. As a consequence of those arguments, using the U.S. Army as a stakeholder to compare and contrast the recommendations seems relevant. However, as already mentioned, the author fully acknowledges the limitations of such a process as in no way, the French Army by its means can be compared to the U.S.'s. The objectives of the following paragraphs are limited to a compare and contrast process. The synthesis presented below relies on the reading of a selected literature made explicit in the bibliography of the present thesis. In addition, as already mentioned, the author has been able to conduct an interview with a TRAC civilian employee on the 6th of September 2019. This interview has allowed the author to better understand the historical evolution of the U.S. structure in the fields of Data Science which is not made explicit in the doctrine and strategy documents read.

Overall Organization

The U.S. Army approach on Data science is very similar to the French Army one. The ORSA and simulation backgrounds are also existing and are organized in organizations such as the already mentioned TRAC and the Army Modeling and

[26] North Atlantic Treaty Organization.

[27] Shared Vision Statement.

Simulation Office (AMSO) at HQDA[28] level. These structures are historic ones that can be tracked back to the apparition of their related techniques as valuable assets to the military terrestrial endeavor. In addition to the already mentioned actors, G6 of HQDA is also providing an interesting reflection on Data as a new valuable resource.[29] This overall organization does not seem to lead to interesting insights concerning this thesis except for AFC, a new actor created in 2018, which requires a dedicated analysis.

Army Futures Command

The creation of the AFC in July 2018 answers a need to face several challenges. Among them are the return to LSCO, MDO, and the programs to renew the Big Fives (reference to the five pieces of equipment that framed the 1980s: Apache and Black Hawk helicopters, M1 Abrams main battle tank, Bradley infantry fighting vehicle, and Patriot missiles). Identified gaps have led to creating CFTs within AFC to work on each equipment project.

[28] Headquarters, Department of the Army.

[29] Office of the Army Chief Information Officer/G-6, *Army Data Strategy* (Washington, DC: HQDA CIO/G-6, 2016).

AFC's CFTs
Long Range Precision Fires
Next Generation Combat Vehicles
Air and Missile Defense
Soldier Lethality
Synthetic Training Environment
Network Command, Control, Communication and Intelligence
Assured Positioning, Navigation and Timing
Future Vertical Lift

Figure 7. AFC's CFTs

Source: Army Futures Command, "About," U.S. Army, accessed 20 April 2020, https://www.army.mil/futures#org-about.

Technologies related to Data and more especially Artificial Intelligence are embedded within each of the CFT of the identified subject on which AFC is working. Data Sciences are at stake in many CFTs. They appear more clearly in the CFT related to Network. It also matches the 5th Army Priority Research Area.[30] Looking at those priorities underlines the high reliance on the competencies of Data Science.

[30] Army Futures Command, *2019 Army Modernization Strategy: Investing the Future* (Austin, TX: Army Futures Command, 2019).

Army Priority Research Areas
1. Disruptive Energetics: Greater than 2x energetic energy over smaller footprints.
2. RF Electronic Materials: Taking advantage of optical and thermal properties of diamond materials for directed energy.
3. Quantum: Optimized information transfer, sensing, and communication with unparalleled security.
4. Hypersonic Flight: Aerodynamics, materials, and processes.
5. Artificial Intelligence: Increasing speed and agility in which we respond to emerging threats.
6. Autonomy: Maneuverability and off-road mobility of platforms.
7. Synthetic Biology: Reactive and responsive skins/spectrally selective materials/anti-materiel properties.
8. Material by Design: Protection overmatch against future threats.
9. Science of Additive Manufacturing: For next generation munitions for increased range and lethality.

Figure 8. AFC's Priority Research Areas

Source: Army Futures Command, *2019 Army Modernization Strategy: Investing the Future* (Austin, TX: Army Futures Command, 2019).

The U.S. Army Combat Capabilities Development Command Army Research Laboratory is an AFC component in charge of research, it produces documents on Machine Learning, Deep Learning, Data Science, and Artificial Intelligence.[31] Data is envisioned to be the enabler of new technology development and, as such, to be key to the success of the Army. AFC is focused on future and so is capable of implementing Data Science related techniques.

Artificial intelligence has seen the creation of a dedicated Task Force under the umbrella of AFC. This structure is working in close relation with universities and civilian specialists. AFC is providing a strong support to Data Science and its related application.

[31] Michael Lee, Ramakrishna Valisetty, Alexander Brueuer, Kelly Kirk, Brian Panneton, and Scott Brown, *Current and Future Applications of Machine Learning for the U.S. Army* (Adelphi, MD: U.S. Army Research Laboratory, April 2018).

It helps and accompanies the Data revolution. AFC also owns a 1-star general in charge of artificial intelligence. This leadership choice is a powerful sign of the importance granted to Data Science. The AI task force has released its areas of interest.[32]

AFC's AI Task Force areas of interest
Autonomous platforms
Artificial Intelligence and Machine Learning
Data visualization and synthetic environments
Assured Position, Navigation, and Timing
Sensing
Communications & networks
Computation
Internet of Things
Protection
Human Performance
Underpinning Methodologies

Figure 9. AFC's AI Task Force Areas of Interest

Source: U.S. Army Artificial Intelligence Task Force. "Areas of Interest," accessed 20 April 2020, https://armyfuturescommand.com/wp-content/uploads/2019/11/AITF-Scope-Areas.pdf.

These non-exhaustive areas of interests are matching the previously identified applications of Data Science. They are very materiel focused and the mission of AFC

[32] U.S. Army Artificial Intelligence Task Force, "Areas of Interest," accessed 20 April 2020, https://armyfuturescommand.com/wp-content/uploads/2019/11/AITF-Scope-Areas.pdf.

clearly is clearly underlying. Concerning the present thesis, no novelties or gaps appear concerning the expected benefits of Data Science.

To conclude on AFC, this structure, even if heavily polarized by materiel development, offers a very positive opportunity to implement new technologies such as Data Science. The transverse organization, open on the civilian world, offers flexibility to the U.S. Army. As far as Data Science, it is interesting to note that it takes an important place and has a general officer in charge. However, the U.S. Army does not merge the administrative and operational aspect of Data Science. The possibilities offered by Data Science are on the one hand carried by AFC for operational purpose and by G6 HQDA for the administrative dimension of the U.S. Army. This is a major insight for this thesis. Moreover, it seems that no confusion nor merging of ORSA, Simulation, and Data Science are existing within the U.S. Army.

Building an Educated Workforce

The identification of Data Scientists as a key resource for the U.S. Army is not a new trend. Without precise figures, the overall workforce dedicated to Data related tasks is large and the investment has to be understood through the lens of the concept of Network Centric Warfare, a concept depicted by Vice Admiral (retired) Cebrowski in 2005 when he was at the head of the Office of Force Transformation.[33] The transition to the Information Era is ongoing for at least 20 years within the U.S. Army. Civilian and

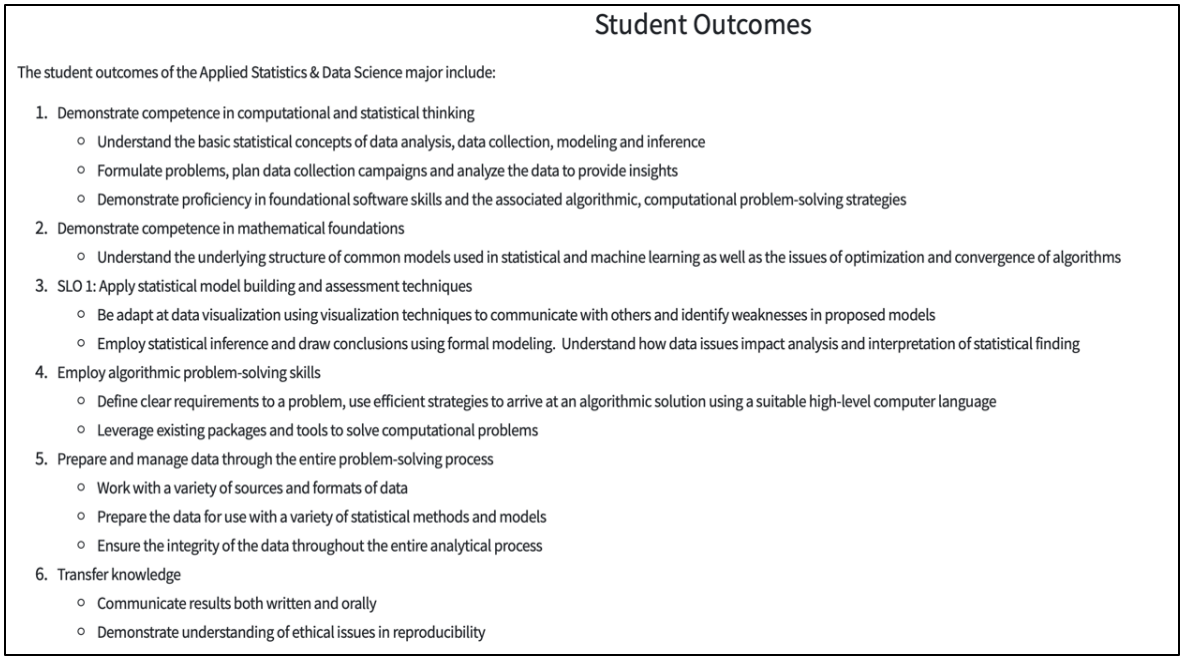
[33] Office of Force Transformation, Secretary of Defense, *The Implementation of Network Centric Warfare* (Washington, DC: Secretary of Defense, 2005).

military personnel is recruited for its academic background in Data related fields and allow the U.S. Army to exploit this domain.

In addition to that, internal solutions are developed. West Point has created a major on Data Science. The future officers can select this new major and thus the Army is generating competencies at a minimal cost. Without any doubt, such academic background will irrigate the future U.S. Army structures where these West Point graduates will be posted.

The following figure reveals the expected outcomes of this major.[34]

[34] U.S. Military Academy West Point, "Applied Statistics & Data Science Major," accessed 21 February 2020, https://www.westpoint.edu/academics/academic-departments/mathematical-sciences/applied-statistics-and-data-science.



Student Outcomes

The student outcomes of the Applied Statistics & Data Science major include:

1. Demonstrate competence in computational and statistical thinking
   - Understand the basic statistical concepts of data analysis, data collection, modeling and inference
   - Formulate problems, plan data collection campaigns and analyze the data to provide insights
   - Demonstrate proficiency in foundational software skills and the associated algorithmic, computational problem-solving strategies
2. Demonstrate competence in mathematical foundations
   - Understand the underlying structure of common models used in statistical and machine learning as well as the issues of optimization and convergence of algorithms
3. SLO 1: Apply statistical model building and assessment techniques
   - Be adapt at data visualization using visualization techniques to communicate with others and identify weaknesses in proposed models
   - Employ statistical inference and draw conclusions using formal modeling. Understand how data issues impact analysis and interpretation of statistical finding
4. Employ algorithmic problem-solving skills
   - Define clear requirements to a problem, use efficient strategies to arrive at an algorithmic solution using a suitable high-level computer language
   - Leverage existing packages and tools to solve computational problems
5. Prepare and manage data through the entire problem-solving process
   - Work with a variety of sources and formats of data
   - Prepare the data for use with a variety of statistical methods and models
   - Ensure the integrity of the data throughout the entire analytical process
6. Transfer knowledge
   - Communicate results both written and orally
   - Demonstrate understanding of ethical issues in reproducibility

Figure 10. West Point's Applied Statistics and Data Major's Student Outcomes

Source: U.S. Military Academy West Point, "Applied Statistics & Data Science Major," accessed 21 February 2020, https://www.westpoint.edu/academics/academic-departments/mathematical-sciences/applied-statistics-and-data-science.

This provides a great example of the flexibility and adaptability the new technologies are requiring. The U.S. Army is making the best out of its existing means.

Recommendations 3 (R3)

After this short and brief analysis of the U.S. Army situation, previous recommendations can be further refined and some additional comments have appeared. The following paragraphs will successively discuss the organizational component of Data Science and more specifically the initial recommendation (R1) of Chapter One and the proposal of creating a dedicated command and add some valuable ideas to the need for a skilled workforce.

Insights from the U.S. Army Data Science Organization

If the U.S. Army is considering the need for implementing Data Science and actively investing in that field, it does not seem to over emphasize the need of a dedicated command structure. The creation of AFC, which has to be understood as a great sign of investment in the future, is very materiel focused and complements existing ORSA and simulation structures that are not questioned. The AFC's AI Task Force is capable of acting on a larger perimeter than just AFC and is therefore a valuable tool. Its general officer materializes the importance of the domain and gives more weight and implementation potential.

Administrative implementation of Data Science seems to be left to G6 HQDA which is the structure in charge of IT solutions and organizations of the network. This distribution of tasks reveals an interesting choice. However, this choice is not transposable to the French Army as, mostly for scale reasons, the French Networks are administrated and managed by a joint directorate (*DIRISI*[35]).

The U.S. Army's choice in term of Data Science organizations seem to be divided into two main categories; the operational one assigned to AFC, and the structural one left to G6-HQDA. AI Task Force seems to be able to act in support of G6-HQDA. From this understanding emerges interesting adaptations to the former recommendation on the dedicated command.

[35] *Direction Interarmées des Réseaux d'Infrastructure et des Systèmes d'Information de la défense* stands for Joint Directorate of Infrastructure Networks and Information Systems.

Before formulating a definitive organizational recommendation for the French Army, one needs to have a look on the mirroring French AI Task Force. Located at the ministerial level, it provides, once more, a valuable framework for the Army but is too far from the applied expectation.[36]

France has already conducted is major materiel renewal reflection. The Scorpion program is under deployment within the forces. Creating an equivalent of AFC would not be adapted to the French structure. However, the need to handle the questions related to the future is still there. As a result of this review, the need for a fully autonomous Data Science command is questioned. The author would like to amend the initial recommendation on an autonomous command to a French Army Task Force attached to EMAT and commanded by a brigadier general. This task force could be in charge of Data Science implementation and would cover both operational and structural environment. ORSA and Simulation would stay as they are right now and will not merge with Data Science.

Enlarging the Qualified Workforce for Limited Costs

The U.S. Army analysis enlightens an academic solution to enlarge the workforce identified need at limited cost. Saint-Cyr military academy could adapt such a program to the existing engineering degree in order to create favorable background for further development of Data Scientists, thus, complementing the already existing research chair on cyber in Coëtquidan.

---

[36] French Ministry of Defense (MOD), *L'intelligence artificielle au service de la défense – Rapport de la Task Force IA septembre 2019* (Paris: MOD, 2019).

These two additional recommendations fulfill the role initially identified for using the U.S. Army model as an external stakeholder. This second half of Chapter 4 has been valuable to confirm the former chapters of this thesis.

CHAPTER 5

RECOMMENDATIONS

The present chapter is based on the recommendations developed in the previous chapters. It aims at defining an implementation plan arranging in space and time these recommendations. In addition to that and to conclude the work, Chapter 5 will also address possible tracks to follow for future researchers and look back on the reflective learning the author has achieved with this work.

What can be expected out of implementing proposed recommendations?

Before addressing the implementation plan, it might be interesting to formulate the expected outcomes of implementing the identified recommendations for the French Army. It is important to note that Data Science is not going to solve all the current challenges of the French Army and should not be expected to. Data Science is an unavoidable characteristic of our time and should consequently be taken in account as an additional factor of the surrounding environment.

First by taking in account the formulated recommendations, the French Army would voluntarily play its part in a mission that has been assigned to the French Armed Forces by the Defense Minister herself, assuming that designating a policy officer cannot be considered as sufficient regarding the ambition of the reform.

Second and as already associated with the defined concept of Data Science, the French Army could look forward to benefiting from two different perspectives. The fastest gain to develop should be with the daily organic activities of the organization. Working faster through more Data sharing and less redundancy. On a much longer

perspective, implementing Data Science within operational activities can also be beneficial to the whole French but the pace to do so must be slower, based on experience and taking Ethics in consideration each time the use of force is considered.

Third and last, the recommendations identified are founded on the use of cross functional expertise employing mutualized and flexible means. The French Army's limited means matches the idea of not overinvesting in Data Science but rather optimizing the ratio investment-benefit. The recommendations are founded on that principle and thus matches the means' framework of any Data Science reform.

As a synthesis, Data Science is unavoidable for the French Army. Its implementation requires to follow pragmatic rules that this thesis' recommendations are taking in account.

Implementation Plan

Implementing the recommendations of this thesis is a complex process that will require relatively important means. Moreover, the final recommendations reached in Chapter 4 should not lead to neglect the constant evolution that needs to be tracked within the Data Science domain. The French Army is expected to evolve during the implementation process and therefore the implementation plan will be limited to a prioritized list arguing in favor of a chronology that seems logical to the author. This chronology could optimize and accelerate the benefits.

This thesis identifies organization as the key of success for the implementation. The first step that is recommended is therefore to simultaneously create a Data Science Task Force directly attached to the MGAT and man it with newly recruited Data Scientists. The primary tasks given to this task force could be to lead the implementation

of a Data Lake within the French Army; to conduct bilateral warfighting functions analysis with associated branch schools study and prospective directions (*DEP*)[37]; to develop experimental projects of Data Science and implement their results throughout the Army; to develop a write the associated regulations; to train an additional workforce within the Army's HQs and organizations to complement its action with sensitized personnel.

As a second step, the Data Science task force his pool of competency should then start a decentralized analysis within the Army units to propose optimizations directions and projects to optimize the collection and the use of Data. Each implementation program should include plan to take benefits and participate in the built-up of national Data Lakes increasing the growth of shared Data to increase the potential of benefits. An important point is also the identification and training of local representative who belongs to beneficial units. The proposed methodology is to conduct a first round of optimization assessment at the warfighting function before conducting a second round at the Regiment, Brigade HQ, and Division HQ level. These assessment missions should prioritize their focus following the logic of the more beneficial fields of application. Data collection capability is probably the most effective criteria and should be considered first as it constitutes the primary requirement to make Data Science effective. A second criteria should then be the focus on the repetitive and time-consuming tasks in order to generate fast gains that will convince the users and favors the whole process. Finally, a

[37] *Direction Études et Prospective*.

third and last criteria to follow could be to follow commander's assessment on the most beneficial path to follow in order to improve the operational effectiveness of units.

The third step could be recurrent and consisting in producing a yearly report on the evolution of Data Science. This report should be composed of the following elements: overall Data Science's situation, actions performed in favor of the French Army over the past year, actions envisioned for the following year.

The fourth and final step could be to assist and orientate Saint Cyr academy in creating a Data Science option within the Engineering degree in order to stimulate new ideas and provide educated officers adapted to the current concerns. These students could usefully involve in internship to support the Data Science task force.

This methodology constitutes a way to prioritize the implementation as means both human and technical are not going to be sufficient to fulfill all the identified needs at the same time. As a final comment the Data Science task force should be manned enough to be able to conduct centralized development and support actions while conducting decentralized activities to develop and elaborate local solutions. A combination of not more than 10 Data Analysts and Scientists could probably achieve these actions, leading to an overall volume of 30 men and women for the Task Force.

Research Questions for Future Researchers

It is assessed as probable that the subject of Data Science may be of some interest to other researchers in the coming years. Even if this thesis is mostly focused on the French Army, many of the depictions, analysis, and recommendations may be valid for other organizations. Moreover, the subject itself can be further investigated. Research questions for future researchers can head in many directions that can be sorted in many

types. The first one focuses on the expected benefits; one can start reflecting on implementation by warfighting functions or materiel solution with questions such as how does Data Science (or Artificial Intelligence) can benefit to the Movement and Maneuver or Fires warfighting functions? A second path to be followed is ethics, and the reliability of Data Science. This field arouses questions like: What are the ethical limitations of artificial intelligence? How far can a military organization go in implementing autonomous Data-related technologies? A third direction could be to actualize this thesis as Data Science is a constantly evolving domain. The focus can be put on the organizational perspective and enriched by civilian organization case study comparisons to improve military choices.

To synthetize this, future researchers can explore many options of study. Whether widening the subject to the ever-growing spectrum of application, or narrowing it to a specific case in relation with a specific subject. The current situation, as depicted, is assumed to be a lasting and sustainable expansion. The questions mentioned above will definitely expand to new challenges and new fields of application. The technologies and related algorithms will also improve skills associated with Data Science.

Reflective Learning

Researching, reading, and writing on Data Science is a tremendous learning opportunity. As a science domain, the perspectives are gigantic and the field of applications are nearly without end. The author is convinced that Data Science and its ongoing implementation is already bringing a new type of revolution to the military that can refer to both the concepts of military revolution and the revolution in military affairs. It is at the same time changing the techniques used to build technical dominance and

enlarging the spectrum of conflict by creating new weaknesses and opportunities for adversaries. Implementing Data Science is, at the same time, reinforcing dominance and weakening our defense policies by relying on easily targetable technologies as globalized and interdependent.

As an officer, this thesis has consolidated my intent and will to serve in the domain of Data Science and Artificial Intelligence. Modern militaries relying on limited resources and limited size such as the French Army and many others Western militaries are investing in this technological dominance to maximize the ratio investment-payoff. A reflection on this bet is that it needs to be pursued with smartness as the optimization effect goes along implementing new weaknesses.

As a researcher, working on an endless amount of literature has been a great challenge. Sorting it and prioritizing readings has been a path of improvement that will surpass the MMAS endeavor and produce effects in the staff officer realm.

As a final word, addressing to a potential reader, I would like to recommend MMAS project as a key element of my CGSC learning experience. There is nothing like working on his own on a subject that we have chosen from our personal interests. In an environment where everything is graded to motivate students, taste for self-learning, professional improvement, and personal investment is a successful alternative.

BIBLIOGRAPHY

Understanding Data Science

Ayres, Ian. *Super Crunchers*. New-York: Bantam, 2008.

Fridman, Lex. “Bjarne Stroustrup: C++” Artificial Intelligence (AI) Podcast, Lex Fridman Artificial Intelligence Podcast + Clips, 7 November 2019. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=uTxRF5ag27A&list=PLrAXtmErZgOdP_8GGztsuK9nrraNbKKp4&index=20.

———. “Deep Learning State of the Art (2020).” MIT Deep Learning Series, Lex Fridman Artificial Intelligence Podcast + Clips, 10 January 2020. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=0VH1Lim8gL8.

———. “Elon Musk: Tesla Autopilot” Artificial Intelligence (AI) Podcast, Lex Fridman Artificial Intelligence Podcast + Clips, 12 Avril 2019. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=dEv99vxKjVI&list=PLrAXtmErZgOdP_8GztsuKi9nrraNbKKp4&index=33.

———. “Eric Weinstein: Revolutionary Ideas in Science, Math, and Society” Artificial Intelligence (AI) Podcast, Lex Fridman Artificial Intelligence Podcast + Clips, 20 Mars 2019. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=2wq9x2QcZN0&list=PLrAXtmErZgOdP_8GztsuKi9nrraNbKKp4&index=35.

———. “François Chollet: Keras, Deep Learning, and the Progress of AI” Artificial Intelligence (AI) Podcast, Lex Fridman Artificial Intelligence Podcast + Clips, 14 September 2019. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=Bo8MY4JpiXE&list=PLrAXtmErZgOdP_8GztsuKi9nrraNbKKp4&index=30.

———. “Gilbert Strang: Linear Algebra, Deep Learning, Teaching, and MIT OpenCourseWare” Artificial Intelligence (AI) Podcast, Lex Fridman Artificial Intelligence Podcast + Clips, 25 November 2019. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=lEZPfmGCEk0&list=PLrAXtmErZgOdP_8GztsuKi9nrraNbKKp4&index=17.

———. “Jeff Hawkins: Thousand Brains Theory of Intelligence” Artificial Intelligence (AI) Podcast, Lex Fridman Artificial Intelligence Podcast + Clips, 1 July 2019. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=-EVqrDlAqYo&list=PLrAXtmErZgOdP_8GztsuKi9nrraNbKKp4&index=40.

———. “Stuart Russell: Long-Term Future of Artificial Intelligence” Artificial Intelligence (AI) Podcast, Lex Fridman Artificial Intelligence Podcast + Clips, 9 December 2018. YouTube video. Accessed 20 January 2020. https://www.youtube.com/watch?v=KsZI5oXBC0k&list=PLrAXtmErZgOdP_8GGztsuK9nrraNbKKp4&index=55.

Kai-Fu, Lee. *AI Super-powers – China Silicon Valley, and the New World Order.* Boston, MA: Houghton Mifflin Harcourt, 2018.

Kelleher, John D., and Brendan Tierney. *Data Science*. Cambridge: MIT Press Essential Knowledge Series, 2018.

Noël, Jean-Christophe. *Intelligence artificielle : vers une révolution militaire?* Paris: IFRI, 2018.

Pirmin Lemberger, Marc Batty, Médéric Morel, and Jean-Luc Raffaëlli. *Big Data et Machine Learning*. Paris: Dunod, 2015.

Sopra Steria Consulting. *Livre Blanc de la Digitalisation*. Paris, 2016.

The French Approach

French Ministry of Défense (MOD). *Schéma directeur de la transformation numérique – volet stratégique*. Paris: MOD, 2018.

———. *Orchestrer la transformation numérique du ministère des armées*. Paris: MOD, 2019.

———. *L’intelligence artificielle au service de la défense – Rapport de la Task Force IA septembre 2019*. Paris: MOD, 2019.

Naulet, Antoine. “Les données : quels enjeux pour les armées ? Approche croisée s’appuyant sur le cas de l’armée de Terre.” Mémoire, École de Guerre, Promotion général de Gallois, Paris, 2017.

Reflecting on Prospective

Army Futures Command. *2019 Army Modernization Strategy: Investing the Future*. Austin, TX: Army Futures Command, 2019.

Bates, Stephen J.K., Major. “Artificial Intelligence: A Revolution Waiting to Happen.” Master’s Thesis, U.S. Army Command and General Staff College, Fort Leavenworth, KS, 2016.

Berman, Eli, Joseph H. Felter, and Jacob N. Shapiro. *Small Wars, Big Data – The Information Revolution in Modern Conflict.* Princeton: Princeton University Press, 2018.

De Landa, Manuel. *War in the Age of Intelligent Machines*. Cambridge: The MIT Press, 1991.

Din, Allan M. *Arms and Artificial Intelligence – Weapon and arms Control Applications of Advanced Computing*. New-York: Oxford University Press, 1987.

Domingos, Pedro. *The Master Algorithm* – How the Quest for the Ultimate Learning Machine will Remake our World. New-York: Basic Books, 2015.

Huggins, Kevin. *Military Applications of Data Analytics*. Boca Raton: CRC Press Taylor & Francis Group, 2019.

Noël, Jean-Christophe. "Intelligence artificielle : vers une nouvelle révolution militaire." *Focus stratégique*, no. 84 Ifri (October 2018) : 82.

Methodology

Long, Kenneth, Ph.D. "The Applied Professional Case Study (APCS)." Department of Command Leadership, U.S. Army Command and General Staff College, Fort Leavenworth, KS, 2016.

The U.S. Approach

Feige, Erich. "The Army Needs Full-Stack Data Scientists and Analytics Translators." *War on the Rocks*. 14 February 2020. Accessed 21 February 2020. https://warontherocks.com/2020/02/the-army-needs-full-stack-data-scientists-and-analytics-translators/.

Ganger, Robert, John Coles, Jacob Ekstrum, Timothy Hanratty, Eric Heiman, Jason Boslaugh, and Zachary Kendrick. "Army Intelligence Data Science Workshop Summary Report." ARL-TN-0931, U.S. Army Research Laboratory, November 2018.

Knopp, Bradley M., Sina Beaghley, Aaron Frank, Rebeca Orrie, and Michael Watson. *Defining the Roles, Responsibilities, and Functions for Data Science within the Defense Intelligence Agency.* Santa Monica, CA: RAND Corporation, 2016.

Lee, Michael, Valisetty Ramakrishna, Alexander Breuer, Kelly Kirk, Brian Panneton, and Scott Brown. *Current and Future Applications of Machine Learning for the U.S. Army* (Adelphi, MD: U.S. Army Research Laboratory, April 2018).

Office of Business Transformation, U.S. Department of the Army. *The Enterprise Data Analytics Strategy for Army Business 2018-2022.* Washington, DC: Department of the Army, 2018.

Office of Force Transformation, Secretary of Defense. *The Implementation of Network Centric Warfare*. Washington, DC: Secretary of Defense, 2005.

Office of the Army Chief Information Officer/G-6. *Army Data Strategy.* Washington, DC: HQDA CIO/G-6, 2016.

U.S. Army Artificial Intelligence Task Force. "Areas of Interest." Army Futures Command. Accessed 20 April 2020. https://armyfuturescommand.com/wp-content/uploads/2019/11/AITF-Scope-Areas.pdf.

U.S. Department of Defense. *Summary of the 2018 Department of Defense Artificial Intelligence Strategy – Harnessing AI to Advance Our Security and Prosperity*. Washington, DC: Department of Defense, 2018.

U.S. Department of the Army. *The Enterprise Data Analytics Strategy for Army Business 2018-2022*. Washington, DC: Department of the Army, 2018.

U.S. Military Academy West Point. "Applied Statistics & Data Science Major." Accessed 21 February 2020. https://www.westpoint.edu/academics/academic-departments/mathematical-sciences/applied-statistics-and-data-science.

Review

CGSC Learning Resource Center, Combined Arms Research Library. E-mail submission. April 8, 2020. Reviewed for grammar, punctuation, and clarity of expression.