



Juin 2017

Enjeux stratégiques du Big Data pour la Défense

*Possibilités offertes aux Armées
et enjeux de l'exploitation des données*

Par Axel Dyèvre, Martin de Maupeou
et Fleur Richard-Tixier

Les notes stratégiques

L'intelligence
de la décision



Les notes stratégiques

Notes d'étude et d'analyse

*Les auteurs souhaitent remercier l'ensemble des experts
rencontrés au cours de cette étude.*

*Les idées et opinions exprimées dans ce document n'engagent que les auteurs
et ne reflètent pas nécessairement la position de CEIS ou des experts rencontrés.*



CEIS est une société d'études et de conseil en stratégie. Sa vocation est d'assister ses clients dans leur développement en France et à l'international et de contribuer à la protection de leurs intérêts. Pour cela, les 80 consultants de CEIS associent systématiquement vision prospective et approche opérationnelle, maîtrise des informations utiles à la décision et accompagnement dans l'action.

CEIS met en œuvre et anime le DGA Lab, le laboratoire d'innovation du ministère de la Défense www.defense.gouv.fr/dga/innovation2/dga-lab

La Direction Générale pour l'Armement a initié en 2013 un centre de réflexion sur l'innovation et un espace de démonstrations de technologies innovantes dans le domaine des systèmes d'informations, le SIA Lab (www.sia-lab.fr), créé et animé par CEIS sous la responsabilité du Groupe Sopra Steria, architecte-intégrateur du programme SIA.

Forte du succès de cette initiative, la DGA a étendu en 2016 le périmètre de ce centre à l'ensemble des domaines technologiques d'intérêt pour la Défense et a lancé le DGA Lab. Espace de démonstration technologique mais aussi de réflexion collaborative sur les usages des nouvelles technologies, le DGA Lab est ouvert à l'ensemble des acteurs de l'innovation de défense, au premier rang desquels les opérationnels du ministère de la Défense.

Le DGA Lab est mis en œuvre et animé par la DGA avec le concours des sociétés CEIS et Sopra Steria.

Dans le cadre de cette activité, les consultants de CEIS publient des Notes Stratégiques portant notamment sur les questions relatives aux impacts de la transformation numérique pour la Défense.

Travaillant étroitement avec les équipes de CEIS et le DGA Lab, le **Bureau Européen de CEIS** à Bruxelles, conseille et assiste les acteurs publics, européens ou nationaux, ainsi que les acteurs privés dans l'élaboration de leur stratégie européenne, notamment sur les problématiques de défense, sécurité, transport, énergie et affaires maritimes. CEIS - Bureau Européen participe également à des projets de recherche européens dans ces domaines. Pour mener à bien l'ensemble de ses missions, l'équipe s'appuie sur un réseau européen de contacts, d'experts et de partenaires.

Contact :

Axel Dyèvre

adyevre@ceis.eu

CEIS

Tour Montparnasse
33 avenue du Maine
75755 Paris Cedex 15
+33 1 45 55 00 20

CEIS - Bureau Européen

Boulevard Charlemagne, 42
B-1000 Bruxelles
+32 2 646 70 43

DGA Lab

40, rue d'Oradour-sur-Glane
F-75015 Paris
+33 1 84 17 82 77

www.ceis.eu

Retrouvez toutes les Notes Stratégiques sur www.ceis.eu et www.sia-lab.fr

Sommaire

SOMMAIRE.....	6
SYNTHÈSE	7
DE LA DATA AU BIG DATA ANALYTICS.....	9
DES BASES DE DONNÉES AU BIG DATA.....	10
PRINCIPAUX ENJEUX DU BIG DATA.....	13
DE LA DONNÉE COLLECTÉE À L'INFORMATION UTILE.....	16
EXEMPLES D'APPLICATIONS DU BIG DATA ANALYTICS DANS LE CIVL	18
Finance et banque	18
Santé	18
Industrie	19
Commerce et grande distribution	19
BIG DATA ANALYTICS POUR LA DÉFENSE - OPPORTUNITÉS ET RISQUES	20
UN ENVIRONNEMENT DE PLUS EN PLUS « DATA CENTRÉ »	20
UNE NÉCESSAIRE RÉFLEXION SUR LES BESOINS ET LES USAGES	22
UNE APPROCHE « BIG DATA » DANS LA DÉFENSE : ACCORDER BESOINS, MOYENS ET IMPACTS.....	24
COLLECTER, STOCKER ET TRAITER LES DONNÉES : LE DÉFI DES MOYENS.....	25
COMPÉTENCES, FORMATION : LE DÉFI RH.....	28
METTRE EN OEUVRE DES OUTILS DANS UN CADRE ADAPTÉ : LE DÉFI ORGANISATIONNEL.....	29
RÉPONDRE À L'AUGMENTATION DE LA SURFACE DE VULNÉRABILITÉ : LE DÉFI SÉCURITAIRE	29
PUBLICATIONS RÉCENTES.....	32

Synthèse

La numérisation croissante de l'ensemble des secteurs de la société a entraîné une explosion du volume de données produites et donc potentiellement exploitables. On estime que les données numériques créées en 2020 devraient représenter, au niveau mondial, un volume de 44 000 milliards de giga-octets, soit dix fois plus qu'en 2013¹.

Les technologies classiques ne suffisent plus pour traiter de tels volumes avec, par ailleurs, des données de plus en plus hétérogènes dans leurs formats. IDC, dans son « Observatoire de la donnée » 2014 estimait ainsi que à peine 5% des données générées faisaient l'objet d'un traitement et étaient réellement analysées².

Le « Big Data », la science de l'analyse des données (data science) et l'usage de l'intelligence artificielle (ou augmentée) – également appelés Big Data Analytics – deviennent donc les éléments essentiels de la transformation des organisations. Ils sont placés au centre de l'innovation et de l'activité économique pour faire face aux enjeux « des 3V » présentés par les données : volume, variété et vitesse.

Avec le recours à ces outils analytiques, la part des données exploitables devrait largement augmenter : l'« Observatoire de la donnée » 2014 estimait ainsi qu'à l'horizon 2020, 35% des données produites pourraient être analysées.

Pour le ministère des armées, le Big Data traduit également l'accroissement du nombre et du volume des données non structurées et hétérogènes générées par la numérisation des équipements et des systèmes d'armes.

¹ Source IDC (International Data Corporation - IDC est un acteur mondial des études sur les marchés des Technologies de l'Information, et des Télécommunications)

² Observatoire de la donnée : étude IDC 2014, <https://france.emc.com/campaign/observatoire-de-la-donnee/index.htm?cmp=emc-fr-blog>

Des systèmes de géolocalisation à la logistique en passant la gestion du trafic aérien ou maritime, les Armées génèrent et exploitent, dès à présent, une importante quantité de données. Demain, objets connectés, capteurs embarqués, messageries et réseaux informatiques accroîtront encore ces volumes dans tous les systèmes de forces, posant ainsi l'opportunité et le défi de leur exploitation.

Précurseur dans l'utilisation de l'intelligence algorithmique depuis les premiers ordinateurs (chiffrement, guerre électronique, calculateurs de tir, radars, missiles, etc.), la Défense peut tirer profit des technologies émergentes liées au traitement et à l'analyse des données pour améliorer non seulement l'efficacité de son organisation mais également ses performances opérationnelles dans un environnement de plus en plus « data centré ».

L'exploitation de cette masse de données dépendra aussi et surtout de la capacité du ministère des armées à :

- **Définir des usages concrets du Big Data répondant aux besoins des Armées**, du soutien jusqu'à la conduite des opérations ;
- **Rendre disponibles les données nécessaires à ces usages** par des compétences et des outils adaptés aux spécificités de l'environnement militaire ;
- **Evaluer les impacts du Big Data** sur le fonctionnement, les missions et les ressources humaines et matérielles du ministère.

De la Data au Big Data Analytics

“The ability to take data - to be able to understand it, to process it, to extract value from it, to visualise it, to communicate it - that’s going to be a hugely important skill in the next decades.”³

- Hal Varian, Chief Economist, Google

La data ou donnée, est définie comme « l’enregistrement d’une observation, d’un objet, d’un fait, destiné à être interprété, traité par l’homme »⁴. On considère que la donnée est a priori objective par rapport à l’information qui correspond au signifiant, au sens attaché à la donnée, par nature subjectif.

Par exemple :

Donnée : variable température = 35°

Information : "La température est de 35°"

Information interprétée : "Il fait chaud"⁵

³Cité dans « Big Data et Analyse des données » - Yves Gueniffey

⁴<http://www.connex.lip6.fr/~soulier/data/TechnoWeb/introNoSQL.pdf>

⁵ Ibid

Des bases de données au Big Data

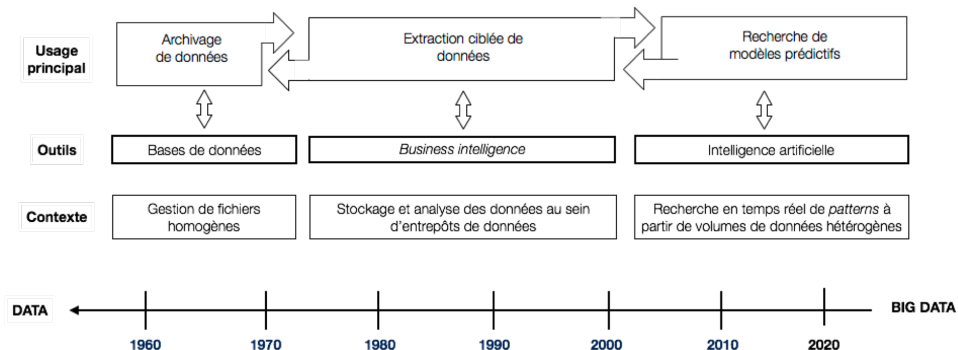


Figure 1 - Des premières bases de données au Big Data (Source : CEIS)

L'exploitation des données informatiques date de la création des premières bases de données dans les années 1960-70. Une base de données est un logiciel qui permet de contenir « *un ensemble de données modélisant les objets d'une partie du monde réel et servant de support à une application informatique* »⁶. De manière simplifiée, le logiciel de base de données permet de créer un énorme fichier (base de données) qui est une sorte d'entrepôt contenant des centaines de références (appelées « champs »), stockées dans des réservoirs appelés « tables ». Des logiciels permettent de manipuler ces bases de données (accès, modification, partage, etc.).

La première génération de bases de données, appelées « hiérarchiques », est développée en majorité par des groupes américains tels qu'Oracle et IBM. À cette époque, les premières applications de centralisation et de partage de données, particulièrement dans le domaine commercial, permettent aux entreprises de rassembler et d'archiver les informations de leurs clients ou de

⁶http://georges.gardarin.free.fr/Livre_BD_Contentu/1-Introduction.pdf

leurs produits relatives par exemple aux factures, aux stocks ou encore à la comptabilité.

À partir des années 80, et bien qu'elles n'aient pas été conçues à cette fin, la disponibilité de toutes ces données centralisées et stockées en un lieu unique génère un besoin d'exploitation : la business intelligence ou informatique décisionnelle organise, analyse et interprète les données disponibles pour fournir une information utile à la décision. Des entrepôts de données sont créés pour stocker les données dans des unités spécifiques et se consacrer spécifiquement à leur analyse.

Cet intérêt pour les données ne concerne pas seulement les grands groupes commerciaux mais présente aussi des applications dans d'autres branches comme le management. Ainsi, au début des années 90, la société française BusinessObjects (rachetée par l'allemand SAP en 2008) développe des logiciels de *business intelligence* pour des entreprises dans la plupart des secteurs d'activité : services financiers, biens de grande consommation, santé, administration publique, etc.

Au tournant des années 2000, la croissance du volume des données mais surtout la diversification de leurs natures et donc de leurs formats bouleversent le paradigme des bases de données et soulèvent des enjeux et défis techniques pour leur exploitation.

Dans le même temps, l'augmentation de la puissance de stockage et de calcul des ordinateurs entraîne une augmentation de la rapidité du traitement des données, réalisé désormais en temps réel. Cette amélioration des capacités et des logiciels d'analyse permet l'exécution de traitements plus complexes, notamment par la construction de modèles prédictifs mais également par la découverte et l'exploitation de corrélations entre des données (ou *pattern matching*).



Dans un contexte où les volumes de données disponibles sont exponentiels, ces défis et progrès de l'analyse de données convergent vers le terme de Big Data. Désormais, l'ensemble des secteurs d'activité ont réalisé l'intérêt de l'interprétation des données transitant sur leurs réseaux informatiques et décident d'investir dans le Big Data pour développer des algorithmes et outils permettant d'établir des corrélations, de classer puis d'extraire de l'information à partir des données.

Principaux enjeux du Big Data

La multiplication des objets connectés, des ordinateurs, des smartphones, des tablettes et autres appareils se traduit par la création en masse d'une « matière première » de données. Se pose dès lors la question des enjeux techniques auxquels répondre pour la valoriser et en tirer des usages.

Aujourd'hui, les logiciels de traitement et d'analyse automatisés exploitent principalement :

- Des métadonnées qui, par opposition aux données, sont des « données sur les données ». Ces métadonnées sont structurées et répondent même souvent à des normes. Dans le cas d'une photo, les métadonnées sont par exemple l'heure et la date de prise de la photo, les caractéristiques de l'appareil, les réglages choisis, le modèle de l'appareil, de la distance de l'objet voire même de la localisation de l'objet de la photographie comme le montre l'illustration ci-dessous. L'existence de ces champs normés permet de traiter et d'analyser les métadonnées de milliers de photographies pour, par exemple, établir des corrélations (géolocalisation) ou les classer automatiquement en fonction de différents paramètres.



Figure 2 - Métadonnées d'une photographie de char Leclerc trouvée sur internet

- Des données hétérogènes qu'une phase d'Extraction, de Transformation et de Chargement (ETL) aura permis d'intégrer, de mettre en cohérence et de structurer. Au contraire des métadonnées et dans le cas simple de fichiers présents sur un ordinateur, la difficulté tient ici pour le logiciel à faire correspondre une photo avec un fichier Excel et un fichier Word. Le recueil de données à partir de plusieurs sources, leur consolidation en un seul format puis leur centralisation dans des entrepôts de données sont les principales étapes nécessaires à l'automatisation de leur traitement.

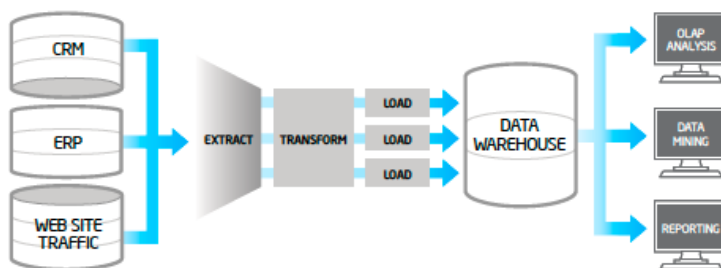


Figure 3 - Le processus « ETL » (Source : Intel)

La variété croissante des formats est le défi majeur de l'exploitation des données numériques - aujourd'hui, 80% des ressources en termes de développement dans un projet Big Data sont consacrées au processus ETL contre 20% pour la partie analyse des données⁷.

Il existe cependant d'autres défis et notamment:

- **Le volume** : selon le cabinet IDC, les données numériques créées en 2020 devraient représenter, au niveau mondial, un volume de 44,000 milliards de giga-octets, soit 10 fois plus qu'en 2013. Autre indicateur pour mesurer l'ampleur du phénomène : en 2011, que 90% des données produites depuis le début de l'humanité l'avaient été dans les

⁷<https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf>

deux dernières années. D'autres estimations avancent que le volume de données produit dans le monde double tous les deux ans : entre 2013 et 2020 on passerait ainsi de 4,4 trillions (1 milliard de milliards) de Go⁸ à 44 trillions de Go de données. La croissance exponentielle du nombre de terminaux et senseurs explique pour une part cette explosion. A titre d'exemple, les 300,000 capteurs embarqués d'un Airbus A380 génèrent ainsi, en moyenne, 16 To de données par vol⁹.

- **La vitesse de captation, de diffusion et de traitement des données :** la multiplication des capteurs qui extraient des données en temps réel conduit à la génération de flux continus et de plus en plus rapides, ce qui constitue un défi technique. De plus, beaucoup de secteurs d'application du Big Data comme le marketing, la sécurité, l'ingénierie ou la maintenance nécessitent une analyse rapide voire instantanée de ces flux. Pour tous ces usages, la vitesse de captation, de diffusion et de traitement est indispensable à une exploitation pertinente en garantissant la disponibilité des données les plus à jour.

⁸[http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS_BRI\(2016\)589801_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS_BRI(2016)589801_EN.pdf)
⁹<http://www.journaldunet.com/solutions/cloud-computing/1170786-internet-of-things-time-series/>

De la donnée collectée à l'information utile

Les progrès dans la collecte et le stockage des données relancent ainsi plus particulièrement la question du traitement des données variées dites non-structurées : comment en effet traiter d'immenses volumes qui incluent des documents sous des formats très divers, des vidéos, des bandes sonores, des tweets, des commentaires, des informations de géolocalisation, de connexion, de mesures, des bases de données commerciales, des données scientifiques, etc. ?

Dans ce contexte, le terme « Big Data Analytics » s'est imposé au début des années 2010 pour désigner justement la capacité de traitement d'importants volumes de données variées et non structurées ainsi que l'analyse « intelligente » de ces données¹⁰. Le Big Data Analytics repose sur des outils et méthodes, au premier rang desquels figure l'Intelligence artificielle (IA), qui permettent d'analyser les données des systèmes d'information, à partir d'algorithmes de calculs définis à la mesure du besoin.

Depuis quelques années, les avancées dans le domaine de l'IA apportent notamment des réponses aux problématiques que posent la variété de la nature et du format des données. Définie comme « *une discipline scientifique recherchant des méthodes de résolution de problèmes à forte complexité logique ou algorithmique* »¹¹, l'intelligence artificielle permet le développement d'algorithmes de traitement complexes et de systèmes d'analyse de plus en plus fins dédiés à la valorisation des données.

Les techniques d'intelligence artificielle permettent notamment d'analyser les données non-structurées, de les organiser et d'en tirer des indicateurs.

¹⁰<https://www-01.ibm.com/software/data/infosphere/hadoop/what-is-big-data-analytics.html>

¹¹https://fr.wikipedia.org/wiki/Intelligence_artificielle

Jusqu' alors non exploitées, elles sont ainsi rendues exploitables par des outils informatiques traditionnels.

En plus de la définition des tendances d'un système, l'analyse et l'exploitation des données au moyen de l'intelligence artificielle et plus largement de la science des données sont porteuses de capacités d'anticipation : les tendances fortes remontées participent ainsi à la définition des modèles prédictifs.

Pour les établir, les algorithmes sont programmés sur mesure en choisissant les variables pertinentes : les outils de calculs fournissent ensuite des réponses automatisées. Les schémas de données établis aux moyens de ces algorithmes complexes peuvent être croisés avec des données externes, telles que la météo, les données de géolocalisation, ou pour des applications logistiques (consommation, remplacement de matériel).

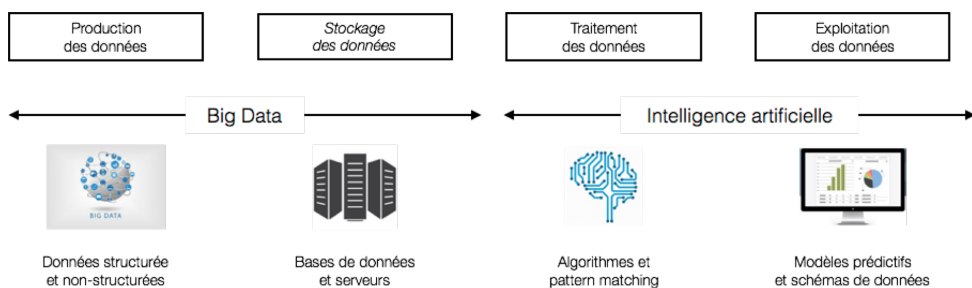


Figure 4 - Le Big Data Analytics (Source : CEIS)

Avec les progrès de l'intelligence artificielle, la part des données non-structurées exploitées devrait largement augmenter. À l'horizon 2020, la société d'études IDC estime ainsi que 35% des données produites seront analysées. Autre conséquence probable, certaines données considérées jusqu' alors comme éphémères car ne pouvant être analysées seront désormais stockées pour être exploitées.

Exemples d'applications du Big Data Analytics dans le civil

Finance et banque

USAGES	EXEMPLES
<ul style="list-style-type: none">✓ Améliorer les temps de traitement des données des marchés, analyser les tendances en temps réel et effectuer des transactions à haute fréquence (d)✓ Lutter contre les fraudes en détectant les anomalies dans les opérations bancaires et en alertant les clients automatiquement	<p>NYSE Euronext, une compagnie de trading boursier a réduit de 99% le temps habituel pour identifier les pratiques frauduleuses en utilisant le Big Data Analytics (e).</p> <p>La banque OCBC (pour Singapour et la Malaisie) utilise les traitements du Big Data pour analyser les préférences de ses clients et coordonner une offre de services plus ciblée. En 18 mois cette campagne a permis une augmentation de ses ventes croisées de 60%.</p>

Santé

<ul style="list-style-type: none">✓ Analyser les informations issues des capteurs en temps réel pour les patients à risque (épileptique, cardiaque, etc.)✓ Passer d'une médecine post traumatique et chimique à une médecine préventive voir prédictive	<p>Des bases de données sont développées en France pour étudier au quotidien le mode de vie, la santé et les habitudes alimentaires de grands échantillons de patients. Ces données sont ensuite traitées pour déterminer les facteurs de risque, de protection ou pour faire de la surveillance épidémiologique (f).</p> <p>Les données générées par objets de santé connectés d'Apple et Google (capteurs de fréquence cardiaque, de taux de glycémie ou de pression artérielle) sont stockés et analysés dans le cadre d'études cliniques (g).</p>
--	---

Industrie

USAGES	EXEMPLES
✓ Gérer les besoins en approvisionnement des stocks	Amazon a créé un algorithme pour suivre l'acheminement d'1,5 Mds de produits dans près de 200 centres de stockage dans le monde.
✓ Améliorer l'allocation des ressources humaines et financières	Pour le transport aérien, Thales a produit un outil prédictif qui, en analysant l'historique des données des systèmes de mesure embarqués sur les 5 années précédentes, permet de réduire de moitié les coûts de maintenance (c).
✓ Optimiser la maintenance	

Commerce et grande distribution

✓ Analyser les comportements des clients pour proposer des publicités et des offres de produits ciblés	Amazon traite sa banque de 152 millions de clients pour personnaliser les contenus et proposer des chariots connectés ciblés (a). Facebook, Twitter et Yahoo utilisent et vendent les données publiques et parfois privées de leurs adhérents à des publicitaires (ex: Publicis) qui proposent des contenus « ultra-ciblés » (b).
--	--

Sources :

- (a) <https://www.technologyreview.com/s/509471/amazon-woos-advertisers-with-what-it-knows-about-consumers/>
- (b) http://www.lemonde.fr/technologies/article/2007/11/10/le-site-facebook-vend-le-profil-de-ses-internautes-aux-publicitaires_976841_651865.html
- (c) <https://events.thalesgroup.com/euronaval/fr/article/777225/Cyber-et-big-data-enjeux-de-developpement-majeurs-du-segment-naval>
- (d) http://fr.wikipedia.org/wiki/Transactions_à_haute_fréquence
- (e) https://www-935.ibm.com/services/multimedia/Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf
- (f) <https://www.inserm.fr/thematiques/technologies-pour-la-sante/dossiers-d-information/big-data-en-sante>
- (g) <http://www.lalibre.be/actu/sciences-sante/objets-connectes-base-de-donnees-et-reseaux-au-secours-de-la-sante-56cc86833570ebb7a8ab2956b>

Big Data Analytics pour la Défense - Opportunités et risques

Conséquence des progrès technologiques actuels, les Armées opèrent un très grand nombre de capteurs, réseaux, ordinateurs embarqués et systèmes d'information. Leurs nouveaux équipements et véhicules (blindés, drones, Scorpion, etc.) sont équipés de senseurs de plus en plus nombreux et avec des des fonctions variées. Ce phénomène contribue à l'explosion des volumes de données numériques et expose le ministère des armées, comme toute autre organisation, aux défis et potentialités du Big Data : comment rendre disponible et exploitable cette masse de données pour en tirer des usages répondant aux besoins de la Défense ?

Un environnement de plus en plus « data centré »

Des systèmes de géolocalisation, en passant par la logistique et la gestion du trafic aérien ou maritime, les Armées génèrent, dès à présent, une importante quantité de données du fait de la numérisation des équipements et des systèmes d'armes. Demain, objets connectés, capteurs, messageries et réseaux informatiques accroîtront encore ces volumes dans tous les systèmes de forces, ainsi que le défi de leur exploitation.

On estime ainsi que l'ARGUS-IS¹², module d'observation constitué de 368 capteurs mis au point pour l'armée américaine, collecte 6 millions de Go de données par jour, alors qu'un drone peut générer au cours d'une mission de 14 heures, environ 70 000 Go de données¹³.

¹²Autonomous Real-Time Ground Ubiquitous Surveillance Imaging System

¹³http://www.claws.in/images/journals_doc/1511401708_RedefiningMilitaryIntelligenceUsingBigDataAnalytics.pdf

Cette explosion des volumes de données présente des opportunités d'usages dans de nombreux secteurs d'intérêt pour la Défense : le renseignement, la santé, la simulation ou encore la maintenance.

Le programme SCORPION offre un exemple de ces opportunités pour la conduite des opérations comme pour l'amélioration de la maintenance des équipements.

Au niveau opérationnel, son nouveau système d'information et de communication unique (SICS)¹⁴ lui permettra d'intégrer des données tactiques entre tout un ensemble de systèmes d'arme, et de partager ces données avec les différents échelons hiérarchiques.

Pour la maintenance, SCORPION permettra la collecte, l'analyse, l'exploitation et l'échange des données : le GRIFFON et le JAGUAR seront tous deux de véritables véhicules connectés dotés de capteurs recueillant automatiquement les données relatives à l'état de leurs composants critiques (moteur, transmission, système de freinage, etc.). Il suffira aux maintenanciers de brancher une « valise électronique » pour collecter ces informations, lesquelles seront ensuite transmises aux systèmes de gestion centralisée de la flotte¹⁵. En pratique, ces innovations technologiques devraient faciliter la maintenance prévisionnelle des matériels et améliorer très substantiellement leur taux de disponibilité, la cible de 80% étant parfois évoqué.

¹⁴Le système d'information du combat Scorpion (SICS)

¹⁵<http://www.defense.gouv.fr/terre/equipements/a-venir/scorpion/actualites/2017/scorpion-disponibilite-amelioree>

Une nécessaire réflexion sur les besoins et les usages

Il y a une dizaine d'années, la réflexion sur la donnée et l'info-valorisation s'est cristallisée sur la nécessité de mieux gérer l'information à caractère opérationnel afin d'accélérer le tempo des opérations¹⁶. Avec le Big Data, l'enjeu est désormais de définir et permettre des usages qui répondent aux besoins de l'ensemble des missions et fonctions des Armées.

Le numérique, présent dans tous les systèmes et outils des Armées, offre en effet un large spectre de domaines d'applications pour le Big Data Analytics, qu'elles soient, pour certaines, partagées avec le civil ou relèvent, pour d'autres, d'usages spécifiquement militaires :

- **L'entraînement des troupes** : le Big Data permettrait de réaliser un entraînement différencié et personnalisé grâce à un suivi en temps réel des performances physiques ou de tir des unités, mais également, au niveau collectif, la mise à disposition d'une cartographie immédiate du taux d'entraînement de l'ensemble des forces¹⁷.
- **La planification et la conduite des opérations** : l'analyse des données permettra de modéliser et simuler des situations au plus près de la réalité des théâtres d'opérations ; d'analyser les futures possibilités de manœuvre et de déterminer les meilleures options ; puis d'améliorer les analyses RETEX pour la manœuvre comme pour l'utilisation des équipements.
- **Le renseignement** : le croisement de données multiples et hétérogènes permettra d'améliorer l'analyse de l'information et d'obtenir une meilleure appréciation de situation.

¹⁶ Rythme des opérations et nouvelles technologies – Le processus décisionnel français à l'épreuve, Note Stratégique CEIS, juillet 2015. <https://ceis.eu/fr/note-strategique-rythme-des-operations-et-nouvelles-technologies-le-processus-decisionnel-francais-a-lepreuve/>

¹⁷Le numérique au quotidien du ministère de la Défense, Chef d'escadron Hugues HELIE. https://penseemiliterre.fr/le-numerique-au-quotidien-du-ministere-de-la-defense_2015297.html

- **La gestion des ressources humaines** : chaque année, le Ministère recrute près de 20,000 personnes dans plus de 400 spécialités, tout en étant à même de répondre à des aspirations individuelles plus importantes et à une évolution sociétale où la recherche de l'équilibre entre vie professionnelle et vie personnelle est au cœur des préoccupations. Dans ce contexte, les données disponibles, déjà nombreuses, pourront être réunies et croisées afin de mieux prévoir les effectifs futurs et les besoins en recrutement, en améliorant par exemple l'identification des recrues présentant un risque « démissionnaire ».
- **La gestion des flux logistiques et des stocks** : l'utilisation accrue d'outils analytiques exploitant les données des systèmes d'information logistiques permettra de mieux planifier les besoins, de calculer dynamiquement l'état des stocks, d'anticiper leur évolution et d'optimiser la gestion des flux en réduisant les coûts, et permettant au commandement de se focaliser sur la prise de décision.
- **La sécurité informatique** : les technologies du Big Data permettront par exemple une analyse automatisée et une exploitation plus fine des événements sécurité (logs) relatifs au trafic sur les réseaux des Armées.
- **La santé** : l'analyse approfondie des paramètres médicaux en temps réel mais également l'établissement automatique de liens avec l'historique des patients permettront un soutien médical amélioré.

Une approche « Big Data » dans la Défense : accorder besoins, moyens et impacts

Au regard, d'une part, du volume et de la taille des données générées et, d'autre part, des particularités des missions des Armées, l'exploitation croissante et systématique des données par et pour la Défense entrainera des changements en termes :

- de **moyens techniques** pour stocker, traiter, échanger et analyser les données,
- d'**organisation** pour s'adapter à cet environnement de plus en plus numérisé,
- de **ressources humaines** pour recruter et former les personnels.

Cette transformation exige de définir les conditions d'une meilleure utilisation des données qui tiennent compte des spécificités d'un environnement militaire. Il s'agit ainsi de répondre à :

- Un défi technique en termes d'infrastructures et d'outils ;
- Un défi opérationnel pour transformer les données disponibles en information utile à la décision ;
- Un défi en matière de ressources humaines (RH) pour attirer des experts en *data science* et former les utilisateurs au sein des forces ;
- Un défi organisationnel pour intégrer et adapter l'analyse des données au fonctionnement global du Ministère ;
- Un défi sécuritaire pour assurer l'intégrité, la confidentialité et la validité des données produites, traitées et exploitées.

Collecter, stocker et traiter les données : le défi des moyens

Si certains usages des données apparaissent au sein des Armées pour répondre à des besoins concrets, ils restent encore limités en raison de la difficulté à exploiter des données issues de systèmes différents et qui sont donc le plus souvent non structurées et hétérogènes. Une récente étude du think tank britannique RUSI (Royal United Services Institute for Defence and Security Studies) voyait ainsi dans la variété des données le principal défi pour une organisation telle que le ministère de la Défense britannique.

On estime aujourd'hui par exemple que l'essentiel de l'analyse automatisée d'informations pour le renseignement porte sur les métadonnées : ce qui explique que, selon un ancien agent du renseignement britannique, 95% des données captées ne sont ni vues ni analysées par les analystes. La DRM analyserait pour sa part jusqu'à 20% des données dont elle dispose¹⁸.

Le premier enjeu du *Big Data* reste dès lors pour les Armées, comme pour toute organisation, l'acquisition des données — ou en d'autres termes de la « matière première ». Celle-ci nécessite une stratégie et des moyens techniques pour identifier les informations d'intérêt, les localiser, les extraire et les rendre exploitables. Ce qui renvoie aux enjeux déjà abordés du traitement des données structurées et non structurées.

Mais il faudra pour cela disposer de moyens de stocker d'importants volumes de données, un enjeu particulièrement complexe dans un environnement militaire.

¹⁸Le renseignement militaire, nerf de la guerre contre Daesh Le renseignement militaire, nerf de la guerre contre Daesh - <http://www.lefigaro.fr/international/2017/03/24/01003-20170324ARTFIG00336-le-renseignement-militaire-nerf-de-la-guerre-contre-daech.php>

Corollaire du Big Data, le *cloud computing* est l'accès via un réseau de télécommunications (généralement Internet), à la demande et en libre-service, à des ressources informatiques (capacités de stockage de données et/ou de traitement) partagées¹⁹. Le développement d'une stratégie « Big Data » au sein des Armées rendra nécessaire le recours au fonctionnement en cloud, qui n'apparaît désormais plus comme une limite compte tenu du progrès des technologies permettant l'hybridation entre infrastructures locales et à distance.

En effet, la spécificité des missions des Armées oblige à penser le déploiement et l'utilisation de ces moyens techniques dans un environnement particulier caractérisé par des réseaux contraints et des conditions d'emplois dégradées sur les théâtres d'opérations. Le stockage, l'échange et l'exploitation de gros volumes de données posent dès lors des enjeux de répartition des données rendant nécessaire l'hybridation des ressources entre des serveurs centraux et des ressources réparties du niveau opératif jusqu'aux véhicules.

En plus des moyens de stockage nécessaires à la mise à disposition et à l'accès des données, se pose le défi de la réduction du temps de l'analyse des données une fois disponibles. Dans le cas du renseignement, les satellites et autres capteurs produisent en temps réel des données croissantes alors que le nombre d'analystes du renseignement n'augmente pas.

Plus largement, les Armées dans leur ensemble risquent de se trouver confrontées à une masse de données avec un intérêt ou une validité difficile à identifier. Comme le souligne le document de réflexion prospective Action terrestre future (ATF), publié en septembre 2016 par l'Armée de Terre, le « vacarme informationnel » et « l'inflation de l'infomasse » compliquent la capacité de tri et de hiérarchisation de l'information.

¹⁹<http://www.nist.gov/itl/cloud/>

La surabondance de l'information disponible, du nombre de capteurs, de leur autonomie énergétique et des systèmes d'information et de communication fait peser un risque de saturation, voire de surcharge cognitive, et *in fine* de paralysie de la chaîne décisionnelle. En conséquence, il est essentiel d'établir les bonnes priorités de recherche et d'utiliser des solutions appropriées pour le classement, l'analyse et la valorisation des données.

Il faut donc que les outils soient plus performants pour permettre de produire des analyses opérationnelles dans le temps de l'action. Cette performance passe en premier lieu par la capacité des systèmes à éliminer l'information non pertinente pour focaliser les utilisateurs sur les données de qualité et ainsi, non seulement réduire le volume à analyser, mais aussi améliorer la qualité de l'information-résultat.

La réponse à ce défi reposera sur le recours à l'Intelligence Artificielle qui dotera le système d'une capacité à trier voire évaluer de façon autonome la qualité des données. L'enjeu sera également, lors de l'analyse, de croiser les multiples sources d'informations pour produire des éléments utiles à la prise de décision.

Pour garder l'exemple du renseignement, il s'agira de flux d'images, fournies par des satellites, avions de reconnaissance ou drones, de données électromagnétiques issues des écoutes, d'informations d'origine humaine recueillies sur le terrain et de données collectées sur les réseaux numériques.

Compétences, formation : le défi RH

La réponse à l'ensemble de ces défis technique et humain passe par le développement de compétences spécifiques au sein du ministère des armées pour, d'une part, disposer des profils techniques nécessaires à l'exploitation des données et, d'autre part, former les utilisateurs au sein des forces armées au recours de plus en plus systématique aux outils et données numériques dans la réalisation de leurs missions.

Le *data scientist* transforme les données en information et ainsi crée de la valeur en développant des algorithmes pour analyser les données massives. Son profil est multidisciplinaire : mathématiques, informatique, algorithmique et statistique. Il doit de surcroît disposer d'une expérience militaire ou collaborer de manière étroite avec des opérationnels qui exprimeront des besoins que le *data scientist* doit être à même de comprendre.

Avec l'intérêt croissant des entreprises pour le *Big Data*, il y a pénurie de ces profils. En 2018, il devrait ainsi manquer 190 000 *data scientists* aux Etats-Unis. Aussi, la concurrence pour les attirer est forte ce qui conduit les entreprises à leur proposer des rémunérations attractives - le salaire annuel moyen d'embauche d'un jeune diplômé dans ce secteur est actuellement de l'ordre de 60,000 €. Une difficulté supplémentaire est de convaincre ce type de profils - davantage attiré par les start-ups - de rejoindre les grands groupes ou les administrations.

En plus de ce besoin en compétences techniques, le recours et l'utilisation de plus en plus systématiques des données tendent à rendre certains métiers de la Défense très techniques. Si l'ergonomie des interfaces permettra aux utilisateurs de s'affranchir de certaines exigences techniques, le *Big Data* créera un besoin de formation de ces utilisateurs sur l'ensemble de la chaîne opérationnelle, depuis le logisticien gérant les stocks dans les entrepôts jusqu'aux soldats échangeant des informations sur les théâtres d'opération.

Mettre en oeuvre des outils dans un cadre adapté : le défi organisationnel

Le recours aux outils du Big Data s'opèrera dans un contexte où la numérisation des Armées est déjà largement entamée : de nombreux systèmes d'information opérationnels, logistiques, de communication, administratifs coexistent au sein du ministère. À cela s'ajoute les très nombreux équipements (infrastructures, véhicules et soldats) aujourd'hui connectés et remontant les données depuis le terrain.


Dans une logique de cohérence globale, la cohabitation entre systèmes anciens et nouveaux permettra la multiplication des sources et des possibilités de traitement des données.

Répondre à l'augmentation de la surface de vulnérabilité : le défi sécuritaire

La systématisation de la collecte, de l'analyse et de l'exploitation de données numériques, ainsi que leur échange et transfert sur des réseaux, exposent les Armées à un large spectre de risques, que ceux-ci soient la conséquence d'actes intentionnels ou de vulnérabilités accidentelles.

Un impératif de sécurité permettra de garantir la disponibilité, la confidentialité, l'intégrité et la traçabilité des données exploitées, ainsi que la sécurité des informations communiquées et celle, physique, des capteurs faisant remonter les données. Le tout en prenant en compte les spécificités d'un emploi dans un contexte militaire (réseaux et connectivité contraints).

Exemple concret de vulnérabilité, la cybersécurité des données collectées par les capteurs militaires est un défi important de la généralisation des objets



connectés dans la défense. La multiplication des senseurs élargit les surfaces et donc les points d'entrées pour les cyber-attaques visant à extraire et siphonner des données critiques durant les opérations. Ces capteurs doivent donc faire l'objet d'une protection renforcée et d'un enseignement d'usage adapté par et pour les militaires.

Une politique de gestion des données devra également être mise en place, prenant notamment en compte les aspects légaux ainsi que les procédures actuelles du ministère pour, le cas échéant, les adapter de façon adéquate. Des données pourront être exploitées sans conservation ultérieure, d'autres auront vocation à être stockées de manière temporaire ou pérenne.



ceis

Société Anonyme au capital de 150 510 €

SIREN : 414 881 821 – APE : 7022 Z

Tour Montparnasse - 33, avenue du Maine - BP 36

75 755 Paris Cedex 15

Tél. +33 1 45 55 00 20 / Fax +33 1 45 55 00 60 / contact@ceis.eu

Publications récentes

A télécharger sur www.sia-lab.fr ou www.ceis.eu

[Internet des Objets \(IoT\) - Nouvelle donne pour la Défense ?](#) - Juin 2017

[Enjeux stratégiques du Big Data pour la Défense](#) - Juin 2017

[Emploi du Cloud dans les Armées](#) - Juin 2016

[Impression 3D - Technologie de rupture au service des Armées](#) – Juin 2016

[Rattrapages technologiques et technologies de l'information](#) - Déc.2015

[Impact de la numérisation sur l'exercice du commandement](#) – Déc. 2015

[Les objets connectés et la Défense](#) – Déc. 2015

[Numérisation de l'outil de Défense](#) - Juin 2015

[Rythme des opérations et nouvelles technologies](#) - Juin 2015

[Le SIA Lab – Retour sur 2 ans d'activité](#) – Juin 2015 (English version available)

[Conditions d'utilisation des logiciels de l'OTAN par les Nations Alliées](#) – Juin 2014 (English version available)

[Mission des Armées et systèmes d'information](#) - Déc. 2013

[Le Système d'Information des Armées \(SIA\)](#) - Déc. 2013

