Meeting the Unmet Data Collection & Management Requirements of Big Data Analytics and AI for Military Decision Making

A. Bell, J. Eaton, K. Peeper, R. Kolhoff, D. Richardson, G. Onslow, T.H. Heger, M. Raap, M. Krueger, M. Mirsberger, M. Gaul, O. Ozkan, N. Senay

SAS-111 Research Task Group

eaton.jacqueline@hq.nato.int, katrinpeeper@bundeswehr.org, a.bell@mc.nato.int

ABSTRACT

Commanders in NATO operations routinely rely on analytic support to enhance their decision making processes. Reliable and accessible data is an essential component of this analytical work. However, analysts frequently struggle to find the necessary data, partly because the data has not yet been collected and partly because existing data has not been well managed.

The private sector is rapidly recognizing the challenges and potential benefits associated with analytics of 'big data' and has already begun adapting its working practices to capitalize on these benefits for business success. NATO also needs to consider how it will adapt its collection and management of data to the future information environment. In order to adapt to this challenge, like the private sector, NATO will need to consider not just new procedures or databases, but also how to implement systemic changes across its data-related Doctrine, Organization, Training, Materiel, Leadership, Personnel, Facilities and Interoperability.

The SAS-111 Task Group undertook an activity titled "Collection and Management of Data for Analysis Support to Operations", in order to produce a guide aimed at military HQs to assist them in understanding the challenges and recommending the enhancements required for improved data collection and management. This paper presents a precis of that work.

1.0 INTRODUCTION

The quantity and variety of data available is growing exponentially as military operating environments become increasingly technologically complex. Not only are military systems collecting, storing and transmitting ever more data, but the civilian world is awash with data with potential relevance to military operations. There is no doubt that success in future conflicts will be heavily dependent on which side is able to harness the value of data to their advantage most effectively.

The foundation for an organization's ability to harness the value in their own and others' data is the organization's approach to Data Collection & Management (DC&M). DC&M done properly enables the provision of timely, comprehensive and accurate decision support products to military commanders'. However, poor DC&M means analysis takes longer, costs more, is less comprehensive and results are less reliable. Simply put, military commanders with poor DC&M are at an information disadvantage.

The three-year SAS-111 study on DC&M for analysis support to operations found that there are many DC&M challenges in NATO and national militaries today, not only in the networking and database solutions being used, but also across the dimensions of Doctrine, Organization, Training, Materiel, Infrastructure, Leadership, Personnel, Facilities And Interoperability (DOTMILPF-I). The output of our study is a guide, summarised in this paper, aimed at the Commanders of NATO military HQs who face the issue of managing data collection within their organisations. It recommends steps to make military organizations more capable to collect and manage the data they need to achieve decision superiority.

2.0 RATIONALE FOR THE STUDY

In any competitive environment, be it military, commercial, or sporting for example, the key to gaining the advantage and controlling the situation is to lead the decision making cycle, forcing your adversary into a reactive posture. This requires Decision Superiority, where confident, accurate decisions can be made quickly, driving your adversary's timelines. The desired confidence and accuracy in those decisions will be derived from knowledge of the situation, which in turn will be created from data and information collected, collated and analysed to feed understanding to the decision maker. Military structures, particularly HQ structures, exist to feed this data and information collection, collation and analysis process, thus enabling the decision maker as much as possible. The fundamental question behind the guide is how to enable these HQ structures to effectively handle increasing amounts of data that is more diverse than ever before.

In today's interconnected internet dependent world the quantity and variety of data that a military HQ must effectively and efficiently collect and process is, like the rest of the world, growing rapidly, partly driven by modernization of the battlefield. Modern military operations are also seen as part of a comprehensive approach across the Political-Military-Economic-Social-Infrastructure-Information (PMESII) domains and this requires commanders to understand and monitor the extent to which military operations influence and are influenced by the other domains. Much of the data necessary to generate this wider understanding is being created outside of the military. In some cases, such as with social media sites, the velocity, volume and variety of data being produced is proving too much for traditional database tools to manage (i.e. it is Big Data).

Rapid decision making will depend on having the essential data immediately to hand, in an organised and accessible format. The implication of this is that data has to be collected, managed and validated in advance in order that it can become the knowledge base to feed the analysis, decision making and ultimately the orders of tomorrow.

2.1 So what do we mean by Data?

A popular way to understand data is as the bottom tier of the Data-Information-Knowledge-Wisdom (DIKW) pyramid where lots of data supports a smaller amount of information which supports a smaller amount of knowledge. For analysts in a military organization, anything that needs to be put in context before it is ready for the Commander's eyes is called data. Therefore, the guide uses the terms data and information interchangeably and defines **Operational Data** as "information created, used in, or in support of a combined/joint operation by the Combined/Joint Task Force headquarters, its components, and operating forces that support operational planning, analysis, and assessment of friendly and enemy activity to inform decisions and increase operational effectiveness." Operational data can be any combination of:

- **Military Data:** Data created by military forces and military systems.
- **Open Source Data:** Data available to the general public (often over the Internet) OR **Private Data:** Data not available to the general public (includes most military and commercial data).
- **Free Data:** Data that you can access without having to pay anything OR **Purchased Data:** Data that you can only access by paying a fee (such as academic journals).

Military organisations lack an enterprise level, integrated and standardized capability to collect, store, maintain, and provide access to operational data for analysis. Contributing factors include that data resides within various organizations and systems across the military organisations, that multiple data types and report formats are employed, that data is captured with varying frequencies, and that no standard procedures, applications or tools are used across the various (NATO) commands.

Understanding data as it is being collected, and storing it in a way that retains that understanding, is

essential to ensure that it is used accurately and safely. The implications of having accurate data on, for example, historic Improvised Explosive Device (IED) incidents which is then used to train and prepare own forces, or location and type of enemy activity to visualize risks in certain areas by representing incidents in a georeferenced format, should be obvious.

2.2 Data Collection and Management

DC&M is nothing new to military HQs in that the fundamental principles are already practised for certain specific functions, such as Intelligence Data Collection or Operations Assessment. The steps and processes required by each of these functions map directly onto those that describe a generic DC&M process. The result of our analysis is a process that is summarized in the central section of Figure 1 and is described in more detail in Figure 2. The white areas in Figure 2 are those that are covered in more detail in the chapters of the full DC&M guide.

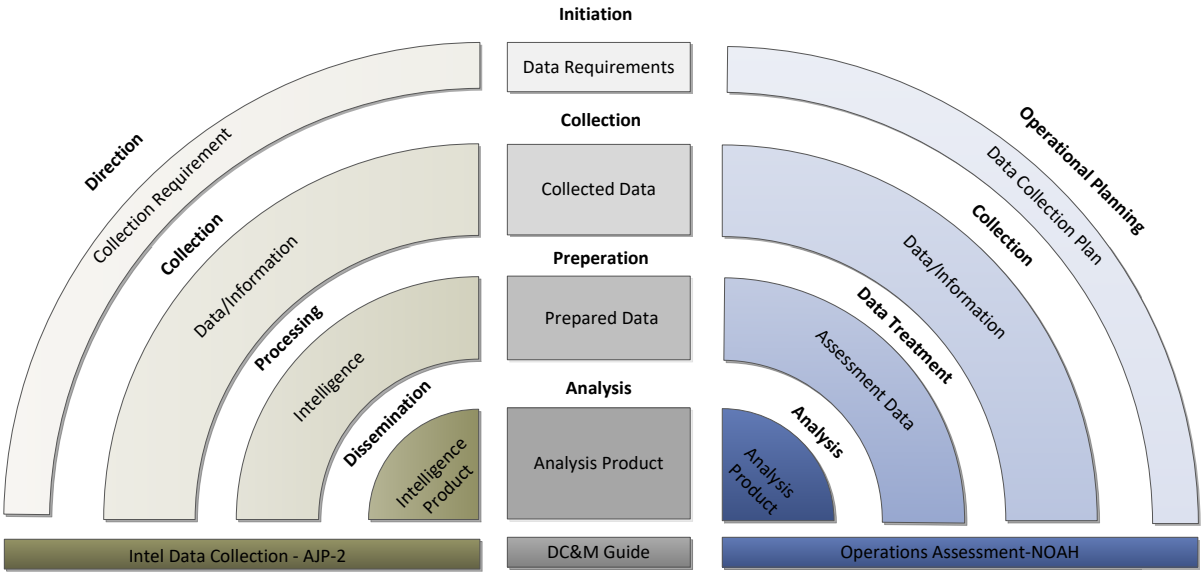

Figure 1: Comparison with Existing Data Collection Processes

Steps	Process	Sub-process	Data State	
RFI/Commander's Information Requirements Operations Assessment Question etc.				
Pre data-life	**Initiation**	- Identify data requirements - Select sources	Raw data	Data sharing
In use	**Collection**	- Create data collection environment - Populate data collection environment	Collected data	
	Preparation	- Cleansing - Verification - Validation	Prepared data	
	Analysis		Analysis product	
Post data-life	**Archiving**	- Archiving - Preserve long term	Archived data/ operational records	
Data available for post operation analysis and future requests (e.g. historic analysis etc.)				

Figure 2: The Recommended Data Collection and Management Process

The steps, processes and sub-processes in the DC&M process result in the different types of data being transformed into different states with respect to their utility for analysis.

- **Raw data** is data in the original form in which it was collected, and which has not been processed in the context in which the analysis will take place. This means raw data can be directly supplied, such as sensor data, or already partly processed data such as operational reporting of SIGACTs, threat reporting, atmospherics, perception data such as surveys or data from social media activity.
- **Collected data** is that which has been stored in a military system in a format that makes it accessible for analysis, together with the metadata that is required to describe the how, where, when and why it was collected in order that its context can be understood.
- **Prepared data** is that which has been checked, cleansed, validated and verified to the extent that it can be considered as 'trusted' data, and used openly to support analysis tasks.
- **Archived data** is that which has been preserved for the long term, but again in a way that its existence and its context is recorded and understood so that it can be located, accessed and analysed in the future as requirements arise.

Common to all of the steps in the DC&M process is the need for Data Sharing, shown in the final column of Figure 2. The sharing aspect of data and information management will become more important than what is achieved today due to the fact that collected information needs to be reusable.

2.3 The Big Data Elephant in the Room

Whilst there is an expectation that Big Data is everywhere, in reality in the military environment it is limited to certain specific systems. At the decision making level of command at which the guide is focused, that Big Data from military systems is unlikely or does not need to be available, with an HQ level organization only seeing the processed results from the sorts of systems that generate Big Data rather than the raw data feed. NATO may therefore have a 'bigger'[1] data problem in organizing and knowing its data, rather than a 'Big Data' problem, combined with the challenge of incomplete or inaccessible data sets for analysis.

External or non-NATO data is another matter, together with certain specific aspects of internal data. Big data techniques and software could be used by NATO to conduct internet and social media research and to search and organize its own network in order to know what NATO already knows and, more importantly, to know where this knowledge and data is located within the NATO CIS (Communication Information System)-architecture.

2.4 From Data Towards Knowledge

The ultimate goal of the DC&M process is to collect and collate the data required which, combined with subsequent analysis, will provide a military HQ with a pool of data, information and knowledge that can collectively be referred to as the military knowledge base. This is however an open and continuous process that will never achieve a final end state. In general, boundaries for the full range of potential analysis work that could be undertaken during an operation are difficult to predefine due to the fact that no one can foresee the full spectrum of questions which may arise from challenges caused by future developments. However the basis for any analysis can easily be defined: sufficient quantity and variety of raw data that is available at the right time. Sufficient raw data and timely access enables high quality analysis that enhances the decision making process. It is therefore important to collect and manage all relevant data, and potentially relevant data for future analysis needs, as required and as completely as possible.

[1] Unlike Big Data, Bigger Data is not a defined term. It is rather an allusion to taking all (available) data into account.

3.0 STUDY FINDINGS

The requirements for implementing DC&M in military HQs can be broken down into three categories: processes (as summarised in Figure 2), people, and tools.

3.1 Processes

3.1.1 Initiation: Identify Data Requirements, Select Sources

Planning and conducting an operation or mission is the core task of a military HQ. But not all analysis is urgent. A driving factor in the Initiation step of the DC&M process is the urgency of the Request For Information (RFI)/ Commander's Information Requirements or Operations Assessment Question, etc. that prompts the analysis. Urgent requests for analysis can only be met using Prepared Data, while less urgent requests allow the analyst time to work with Raw or Collected Data.

Any urgent operational request for analysis support typically has to be answered in accordance with the battle rhythm or the appropriate planning cycle in order to be beneficial to the decision making process. In the best case, the analyst has a few hours or days for researching data and analysing the problem. Such demands are often triggered by unforeseen events or an apparent mismatch in planning assumptions, meaning Prepared Data may not already be at hand. But, that is exactly what the analyst requires:

- the data needs to be accessible from the system where the analysis tools are installed and bandwidth must be adequate for the analysis to take place;
- data classification must allow the analyst access without restrictions;
- the origin and source of the data must be known so that credibility and reliability can be evaluated;
- the data must already be structured and in a common and usable format, or it must be possible to reliably automate structuring and format conversions;
- there must be an audit trail from prepared data to the raw data it came from in case the analyst needs to double check results;
- the data collection environment must be populated and in accordance with an agreed schedule so that analysts know how up-to-date their results are.

These requirements are only achievable when the data has been collected and prepared well in advance. It is therefore recommended to establish an operational data collection plan such as the one defined for the operations assessment process (reference: OPLAN ANNEX OO[2]), but considering all types of data that are known from prior experience to be important, or are easily or automatically collectable within the HQ/staff and attached or subordinated units during the operation.

On the other hand, longer-term analysis tasks arise from questions that have to be answered within weeks, rather than days or hours. This gives the analyst time to conduct a more thoroughly prepared analysis and also provides the opportunity for a deeper examination of the data. It is therefore possible to relax some of the above constraints:

- There is time to ask for further access permissions to overcome classification restrictions;
- Data transfer for reach back can be done slowly and incrementally over low bandwidth links;
- Data does not need to be in usable formats there is time available to process it;

[2] NATO Allied Command Operations Comprehensive Operations Planning Directive Interim V2.0, 04 Oct 2013.

- There is time to prepare the data. For example, structured data can be extracted from unstructured data manually or using data mining and big data tools and techniques, and there is time to validate that the results are appropriate for analysis;
- Reliability can be augmented by comparing multiple sources.

3.1.2 Data Collection

Efficient data collection is of immense benefit to the analyst in that:

- It saves analyst's time and effort and reduces the errors associated with manual input;
- Data processing and Big Data tools can be applied to reduce analyst's workloads;
- The analyst's focus can shift away from data preparation to data analysis;
- Analytical results will be derived from higher quality data;
- Analysts are able to answer more questions, more quickly and more accurately to enhance decision making.

Who collects the data is a complex question, and a data collection strategy will need to be developed. There are a variety of situations under which data will be gathered. The ideal is where the data is automatically collected, although where this is not possible then manual collection will be required. At best manual collection may be data captured digitally via a computer and entered by a subject matter expert: at worst, it will be a photocopy of a handwritten patrol diary or some other non-digital document.

Data is usually stored in some kind of a database. This can be as simple as an Excel sheet, more advanced like a relational database such as SQL Server, or even a data lake. In any case, the data should make its way into that database. We refer to this combination of a database and its input environment as a data collection environment.

The collected data by itself is useless when it has no tag describing what it means. A straightforward example is a set of dates. The dates are data but have no meaning by themselves. By providing metadata to the dates, e.g. dates of IED incidents, the data becomes information and can be exploited for analysis. Also, once the data is stored in the system, the analyst must be able to find the relevant information for their needs. Metadata is: "Structured information that describes, explains, locates, and otherwise makes it easier to retrieve, use and understand an information resource. Metadata facilitates the association of records within the context of broader business activities and functions" (NATO, 2007)[3].

The NATO document "Development of a Data Collection Environment based on a Mission Thread Approach" describes the ideal data collection process in a standardized manner based on a Mission Thread Approach. It uses NATO architecture diagrams such as Activity Charts and Organizational Relationship Charts to give a structured overview and guideline of the process. As a result, the people responsible for designing the requirements for the data collection environment are aided in that process in order to find a balanced solution that ensures required quality of the collected data as well as the required ease and speed of insertion, while keeping the development costs proportionate.

3.1.3 Data Preparation

Data is considered to be ready for analysis, or prepared, when it is structured, complete, accurate, non-redundant, verified, and trustworthy. These are the dimension that makes up the concept of quality data. Each of these conditions is of absolute importance because the analytical results are at best as good as the data they are based upon.

[3] *C-M (2007)0118 The NATO Information Management Policy 11 Dec 2007*

Data Preparation consists of data cleansing, validation and verification, and finally classification. The ability to detect errors in the data may require specific knowledge of the domain, making a domain expert necessary within this process. However, due to the potentially huge amounts of data, the process should be as automated as possible. Error types include:

- Syntactical errors: violations of the specified format of the data;
- Semantical errors: integrity constraint violations, subset constraint violations, duplicates, invalidities, and redundancy;
- Coverage errors: missing values or entire entries.

Measures for data quality are completeness, validity, constraint violation, uniformity, density, and uniqueness (Müller and Freytag, 2005[4]). However, the required quality of the data depends on the application. In some cases, it is not necessary for the data to be perfect, where fixing the tiniest and most detailed errors is not worth it. For each application and type of analysis the necessary and sufficient level of quality must be addressed. From that point of view, the consequences and effects of the decisions made based upon analytical results must be carefully taken into account.

The results of the cleansing process should be subject to Validation and Verification. Validation is all about whether the process meets the need of the end user if successfully implemented and verification is more about whether the process meets its own aims.

3.1.4 Archiving

Data Archives are both critical and valuable assets within any organisation. In a 'Big' or 'Bigger' Data future, no data may be considered obsolete – what may not be judged important today will have intrinsic value that may ultimately become critical historic information for some future operation. However, such archives are complex to manage as they are often extremely large data repositories, usually containing historic data but also increasingly where current and future 'big data' will reside after it has been processed by core computer applications.

It is important to note that archiving is a process, currently and historically undertaken by archivist personnel, and should not be construed as simply a technical issue to be dealt with by separate automated server(s). Moreover, although desirable, not all information items and data are operational records and need to be archived. Currently, simply storing and archiving everything, for every operation or mission is beyond all practical possibilities of NATO in general and its entities in particular, especially considering the amount of data recorded during deployed operations.

Although the principle of data archiving is relatively straightforward, it can be problematic in practice.

- Data integrity / completeness is of utmost importance, because other sources may no longer be available and data cannot be reconstructed;
- Source and originator of data must be known, with background information and (operational) context on them available;
- Time and location must be known;
- Indication whether data is raw or manipulated/pre-processed data;
- Data sets must be searchable and accessible.

A major prerequisite for NATO is that archived records/information must endure for as long as needed to satisfy NATO operational and legal requirements (including relevant Data Protection Acts). With the

[4] Müller, Heiko and Johann-Christoph Freytag. *Problems, methods, and challenges in comprehensive data cleansing*. Professoren des Inst. Für Informatik, 2005.

prospect of increasing volumes of structured, semi-structured and unstructured data needing to be collected as part of current and future NATO operations, so it will become necessary for archival practices to adapt to the challenge, through the acquisition of data analytics tools and improved methods and processes.

3.1.5 Sharing Data

A guiding principle in the NATO Information Management Policy is that, "Information shall be managed with an emphasis on the 'responsibility-to-share' balanced by the security principle of 'need-to-know', and managed to facilitate access, optimise information sharing and re-use and reduce duplication, all in accordance with security, legal and privacy obligations." The same is true of datasets.

However, data reuse is rarely achieved either within an HQ or among the HQ and other entities. Despite the guidance on 'responsibility-to-share', the default position is not to share data. This is partly due to staff erring on the side of caution when there is lack of clarity regarding what data can be shared and under what circumstances it can or should be shared or reused, and partly due to NATO CIS that were designed and implemented independently and are inadequate to support proper data management and data sharing.

The fundamental basis of data sharing is an understanding that data is an asset. In that sense it can be acquired, shared, bought, sold and importantly has an owner. The military do not own all of the data they need to operate in a modern environment. Quality external data is worth paying for because it frees up military staff to focus on military tasks and is the fastest and most cost effective way of accessing certain types of information, but military HQs need allocated budgets for data procurement, need to be familiar with the associated licencing and legal requirements and NATO needs mechanisms to ensure that when it pays for data it only pays once, not multiple times from different entities.

Sharing should also not be limited to only prepared and archived data. Raw and collected data should also be made available, at least to analysts and Artificial Intelligence and Big Data applications, in order for them to be able to verify results or discover new insights from the data.

3.2 People

3.2.1 Roles and Skills

Data Collection and Management is not always the responsibility of the analyst; practically all members of a military HQ will have some role in data collection and management. It is vital that roles and responsibilities are assigned and not just presumed. Data collection and management roles in support of HQ decision making can be generic, incidental to the staffs' primary role, or specialized, where there is particular expertise in the development and management of data collection systems and processes. Key specialist roles, visualised in Figure 3, include:

- **Data Manager:** A small subset of the HQ staff that will have particular expertise in the development and management of data collection environments, systems and processes.
- **Data Analyst:** Responsible for providing quantitative assessments such as descriptive statistics and probability models of source, collected, prepared or archived data.
- **Data Scientist:** A new type of analytical data expert who have the technical skills to solve complex problems. They possess a unique combination of data analysis experience, software development know how, and subject matter expertise.
- **Operational Analysts:** Use analytical methods and mathematically based procedures to enable leadership decisions in a constantly changing operational environment.
- **Knowledge Management Officer (KMO):** Responsible for integrating and synchronising

knowledge and information management.

- **Data Engineers:** Designs and develops customized data collection, management, and search-and retrieval systems to support collection, processing, exploitation, analysis, and dissemination of big and complex datasets.
- **Data Archivist:** Responsible for storing and cataloguing data at the end of an operation, and also after completion of projects by staff.

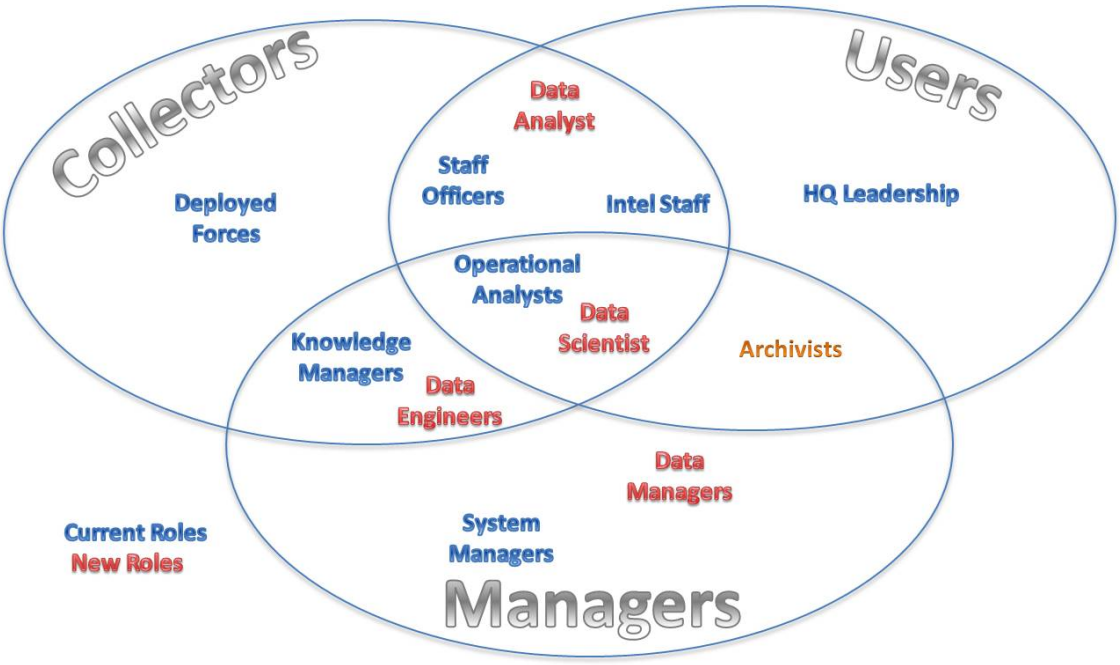

Figure 3. Generic and Specialist HQ Roles

3.2.2 Training

All staff will require the correct skills and access to appropriate training, and with element of a data collection capability within a HQ. Individual training can be sought from four principal sources:

- **HQ Internal Training:** Learning from colleagues, from material developed within the HQ that covers specific procedures or systems, and learning from on-the-job experience.
- **NATO/National Training:** Training from specialist military training organisations, such as the NATO Communications and Information Systems School or the NATO School in Oberammergau.
- **External Training:** Training from commercial training providers or from the training direct from the suppliers of the systems being used in the HQ. Some of these routes to training can result in formal qualifications. A well-known example is Learning Tree International.
- **E-Learning:** On line training which can vary between free courses to paid for training that results in a formal qualification. Examples include Lynda, Coursera, Udemy, or edX.

Whilst individual skills are required in order to understand how to identify, collect, manage and analyse data, the way in which information and knowledge are created from that data within an HQ and for the benefit of a whole HQ will need to be practised through collective training (exercises), where the sharing and collective exploitation of the data available can be practised.

The achievement of this type of training will require both the designation of data collection and management as an exercise training objective, and exercise control that is designed to deliver a continuous flow of data during exercise execution. To achieve this on a NATO-wide basis will require the development of data generation tools to fit exercise scenarios by organisations such as the Joint Warfare Centre, who are responsible for the management and running of major exercises within the NATO Command Structure. This in itself will require the development of individual expertise in the fields of data generation and management in the bodies responsible for exercise control (EXCON).

3.3 Tools

Data science is a rapidly evolving discipline with many new techniques and solutions being implemented and available first as open-source/freeware tools. As a major advantage, these tools are freely available, some with commercial extensions. Unfortunately, it takes some time before any new tool becomes stable and tested software. In the early stage of open source software, necessary documentation can be insufficient but with increasing maturity it typically reaches the level which is standard for proprietary, commercial software.

As it stands there is no 'approved' set of specialist analysis tools issued by NATO and hosted on NATO CIS. Microsoft Office suite is the only constant companion while proper analytic tools that would prove useful to analysts rarely being available due to cost or security concerns, such as:

- Statistics packages such as R and SPSS (Statistical Package for the Social Sciences).
- Database engines, SQL or NoSQL based, such as Filemaker Pro.
- Simulation Engines such as SIMIO or a System Dynamics tool.
- Data analysis tools that facilitate the use of 'Big Data' tools.

Today, the use of Big(ger) Data tools seems much more common in commercial applications than in the military field. Data collection, management, and applications of Big(ger) Data in the NATO context should therefore make use of the experience gained in a non-military context and consider the variety of tools that have been developed by companies, universities, and the open-source and freeware community. But recognize that the NATO context is not necessarily equivalent to the commercial world (e.g., heterogeneous data vis-à-vis homogenous data). For use of these heterogeneous offers in military context, special environments for software and data access under given security regulations must be developed.

Federated Mission Networking[5] is one capability aiming to support command and control and decision-making in future operations through improved information sharing. It provides the agility, flexibility and scalability needed to manage the emerging requirements of any mission environment in future NATO operations.

Commercial cloud services typically deliver Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS)[6] and provide a framework for setup, usage, visualization, and monitoring of Big(ger) Data analytics' processes. Many of these frameworks also provide options and interfaces to use common open-source/freeware tools which can be embedded in their workflows.

However, using such all-in-one packages requires the analyst to give up control over process and results and could indirectly reveal the focus and current interests of the military commander to external entities. Therefore it is not only security concerns pertaining to handling of classified data that makes military application of these services potentially problematic. Nevertheless, some defence/security related

[5] Federated Mission Networking is a NATO initiative to connect the mission networks of NATO organisations, NATO Nations and Mission Partners. http://www.act.nato.int/fmn

[6] Wikipedia, the free encyclopedia (2018): Cloud Computing. (Retrieved from https://en.wikipedia.org/wiki/Cloud_computing)

institutions have already started to use commercial services, such as Amazon Web Services. In these cases, potential added value must be carefully balanced against the potential risks involved. For NATO use, development of a bespoke solution drawing from commercial solutions and also NATO data science developments might address this issue.

4.0 CONCLUSIONS

Military organisations lack an enterprise level, integrated and standardized capability to collect, store, maintain, and provide access to operational data for analysis. This needs to be fixed in order to derive benefits from Big Data Analysis and application of Machine Learning and Artificial Intelligence in a military context.

However we need to be careful about assuming that 'Big Data' is a challenge at operational or strategic level military HQs. Whilst individual military systems may generate what qualifies as Big Data, it is likely that the higher decision making levels will only see the processed output of these systems, not the raw data. On the other hand, military operations conducted in a comprehensive environment, where the actions and influence of non-military actors must also be taken into account, also generate a need to analyse 'Big Data' sources such as social media.

Essentially there needs to be better planning for DC&M, such that data is routinely collected and available in order to enable rapid, short term analysis in the dynamic environment of a military operation. Accurate structuring, cleansing, validation and verification of that data will enhance not only the speed but the quality and confidence of decision making that it enables.

Data archives are critical and valuable assets within an organisation. In a 'Big' or 'Bigger' Data future, no data may be considered obsolete – that which may not be judged as important today will have intrinsic value that may ultimately become critical historic information for some future operation.

Data collection and management roles in support of HQ decision making can be generic, incidental to the staffs' primary role, or specialized, where there is particular expertise in the development and management of data collection systems and processes. It is vital that roles and responsibilities are assigned and not just presumed.

Commercial applications of Big(ger) Data tools are widespread today, much more so than in the military environment. The variety of tools developed and experience gained by companies, universities, and the open-source and freeware community should therefore be exploited, bearing in mind the military need for security.

Ultimately the implementation of a coherent Data Collection and Management Strategy within NATO military HQs based on the principles outlined in the SAS-111 Guide is required in order not only to enhance current processes, but to enable the implementation of modern capabilities such as Big Data Analytics, Machine Learning, or the application of Artificial Intelligence to feed military decision making.