

Using Data Analytics and Machine Learning to Assess NATO's Information Environment

Col Richard Blunt, CapDev JISR, SACT HQ

Allied Command Transformation
Blandy Road, Norfolk, VA
UNITED STATES

Richard.blunt@act.nato.int T: +1 757 747-4214

Chris Riley, Stratcom, Public Diplomacy Division, NATO HQ

NATO HQ
Boulevard Leopold III, Brussels
BELGIUM

riley.chris@hq.nato.int T: +32 2 707 1009

Marc Richter, Michael Street, Daniel Drabkin¹

NATO Communications and Information Agency

Marc.Richter; Michael.Street; Daniel.Drabkin @ncia.nato.int

T: +31 70 374 3444 / +32 65 44 9465

ABSTRACT

The digital world means that NATO's strategic communications operate in a complex information environment where assessment and effect are operating continuously in a competitive space. Many Publicly Available Information (PAI) sources - including traditional and social media, originating within or beyond the Alliance - impact the success of NATO's mission and effects. The volume, variety, velocity and veracity of PAI, which may provide insights, makes this a "big data" problem. While the complex relationships between information streams are difficult to make sense of using only natural assessment and intelligence processes. To provide commanders, Intel analysts and strategic communicators with clear insights requires an information environment assessment (IEA) which cannot be provided in its entirety with traditional analytical approaches. This paper gives an overview of IEA activities across NATO, exploiting data analytics, visualisation and artificial intelligence to bring meaningful insight from the information environment to the operational space. The paper describes the evolution of the existing Alliance Open Source System (AOSS), to operate on a greater range of input data sources, how this open source data is analysed using big data tools which exploit artificial intelligence, and visualised to produce information of operational value. The paper will address technological and non-technical challenges.

1.0 THE CHALLENGE

Many news reports in recent years have drawn attention to attempts to interfere with public opinion in several NATO nations. Such influencing operations have the potential to undermine NATO's strategic communications and therefore the mission of the Alliance. Publicly available information (PAI) – be it

¹ Daniel Drabkin is located at NCIA Mons, BELGIUM; Michael Street at NCIA The Hague, NETHERLANDS; Marc Richter has recently left NCIA The Hague for EUROPOL, The Hague, NETHERLANDS.

Using data analytics and machine learning to assess NATO's information environment

mainstream media, social media or other information which influences populations and affect their support (or otherwise) for NATO and the Alliance – is termed the *information environment*.

The information environment assessment (IEA) is a first attempt to apply data-centric technologies of data science, data analytics and machine learning, to understand hostile attempts to influence public opinion about NATO activities.

1.1 Current assessment of the information environment

Currently strategic communications (StratCom) across NATO, Commands and National elements of the NATO force structure assess the information environment using predominantly manual methods. These methods ensure significant domain knowledge from subject matter experts is used in the assessment. However the scale of the information environment (the volume of data) to be assessed, the range of information sources and the languages used (variety) all present a challenge to human analysts. Accurate assessments are made more challenging by the potential for rapidly changing messages and trends (velocity) and the potential for much misleading information (veracity).

Datasets which exhibit one of more properties of volume, variety, velocity and veracity are exactly those on which data-centric technologies are designed to operate. In this paper we refer to data-centric technologies of data science, data analytics, machine learning and artificial intelligence. Thus these technologies have the potential to make a significant impact when assessing the information environment.

2.0 INFORMATION ENVIRONMENT ASSESSMENT CONCEPT

The IEA, as seen in Figure 1, involves connecting sources of information on the left of the figure with users on the right. Information is a joint function in terms of 'Sensor', 'Assessor', 'Effector' and 'C2'; operating continuously for multiple communities of interest.

The IEA recognises that given the volume, variety and velocity of data which comprise it, it must harness data centric technologies e.g. big data analytics and artificial intelligence, and any capability development in this field will require a data centric methodology.

There is a recognition that requirements amongst disparate communities of interest (COIs) at the source and content aggregation function on the left hand side have a lot of commonality. Likewise, much of the central data science and analytics needed to aggregate content from data sources. Once the source information has been aggregated and prepared for analysis (or *extract, transform and load* of the data) a range of common tools and technologies will be used to analyse and present the data to various COIs. While the data lake and tools will be similar, the analysis and visualisation is likely to be specific for each different COIs.

2.1 Resourcing the IEA

Data Literacy and Data Readiness within NATO is identified as a key capability gap. Steps are being taken to adjust NATO information systems (and system managers) to a data-centric paradigm. While a blend of contracted services, industry expertise and a core of in-house skills will be needed to deliver an effective balance of leading-edge skills and contextual understanding needed to assess information for NATO.

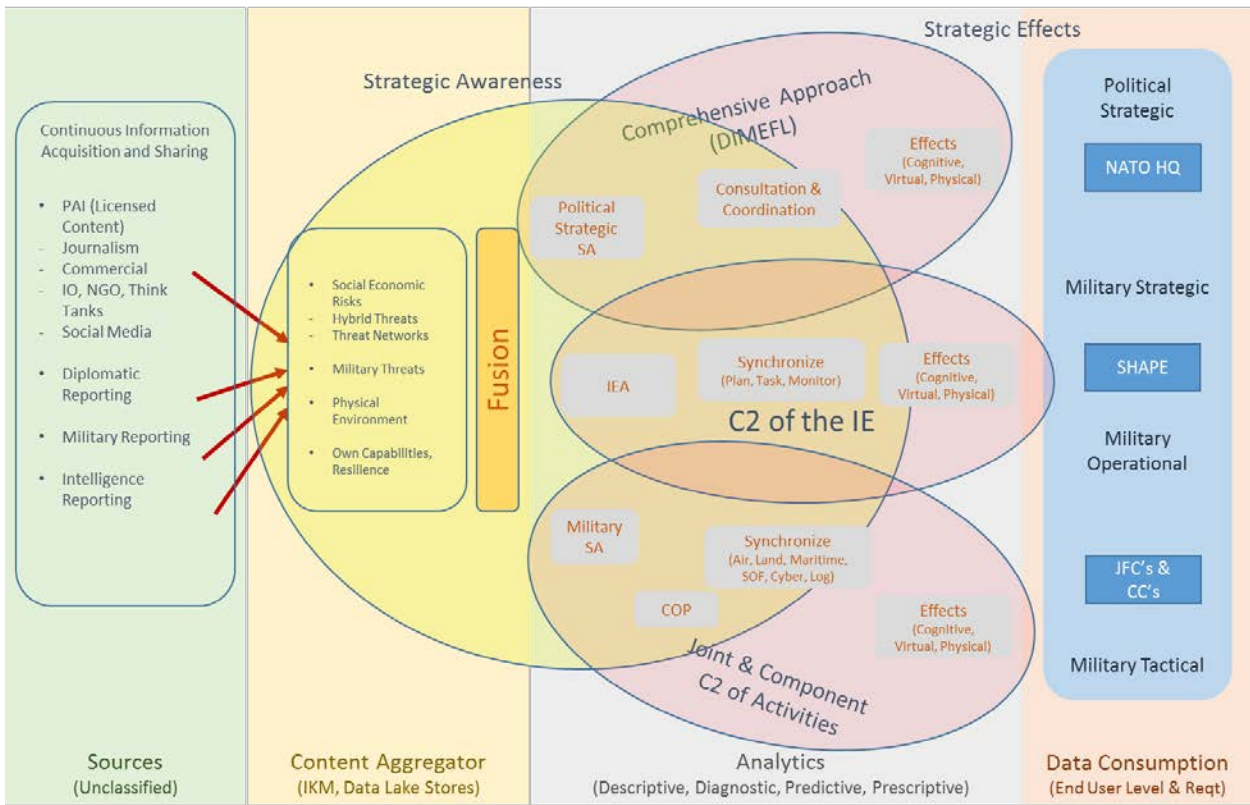


Figure 1: Information environment assessment.

2.2 IEA development areas

IEA development follows four lines of engagement. Industry input is being sought through the NATO Industry Advisory Group (NIAG). Recognising that industry possesses leading edge analytics technologies and expertise, NIAG study group 225, "Big Data in support of IEA" [1]. This provides a forum for industry to propose leading edge technology solutions. This study group is developing a proof of technology demonstration system, using public cloud services to work with large data sets and access sophisticated analytics tools and models. This IEA study builds on an earlier initial NIAG study on the use of big data by NATO [2].

A second line of engagement takes existing NATO capabilities and evolves them. Combining the existing Alliance Open Source System (AOSS) which is already used by analysts to access public information sources is being extended to incorporate sophisticated analytics and visualisation tools already used in NATO [3]. Applying some of NATO's existing data analytics and visualisation tools to AOSS, its existing open source information tool, will provide an initial capability as well as providing a testbed for experimentation and refinement of requirements, analytics processes and views.

Experimentation provides third line of engagement. Experimentation, led by ACT but involving many stakeholders across the Alliance, plays a key role in development, with major experimentation campaigns being planned around suitable events during 2018.

A final engagement strand is the collection of user requirements for IEA FOC, where all three lines above from NATO and industry will all assist stakeholders to define appropriate future requirements for a system which must employ data analytics and machine learning technologies which are comparatively new to

Using data analytics and machine learning to assess NATO's information environment

NATO. These requirements will not only cover aspects of technology and materiel, but the full DOTMPLFI capability landscape e.g. Doctrine, Organisation, Training, Materiel, Personnel, Leadership, Facilities and Interoperability.

2.3 IEA integration into operations

Figure 2 shows how the IEA capability will develop within the wider NATO landscape. Incorporating a functional IEA capability into NATO operations with *effects synchronisation* and *assessment of effects* (friendly and adversary). All of this will include a feedback loop into the IEA, including into the data analytics.

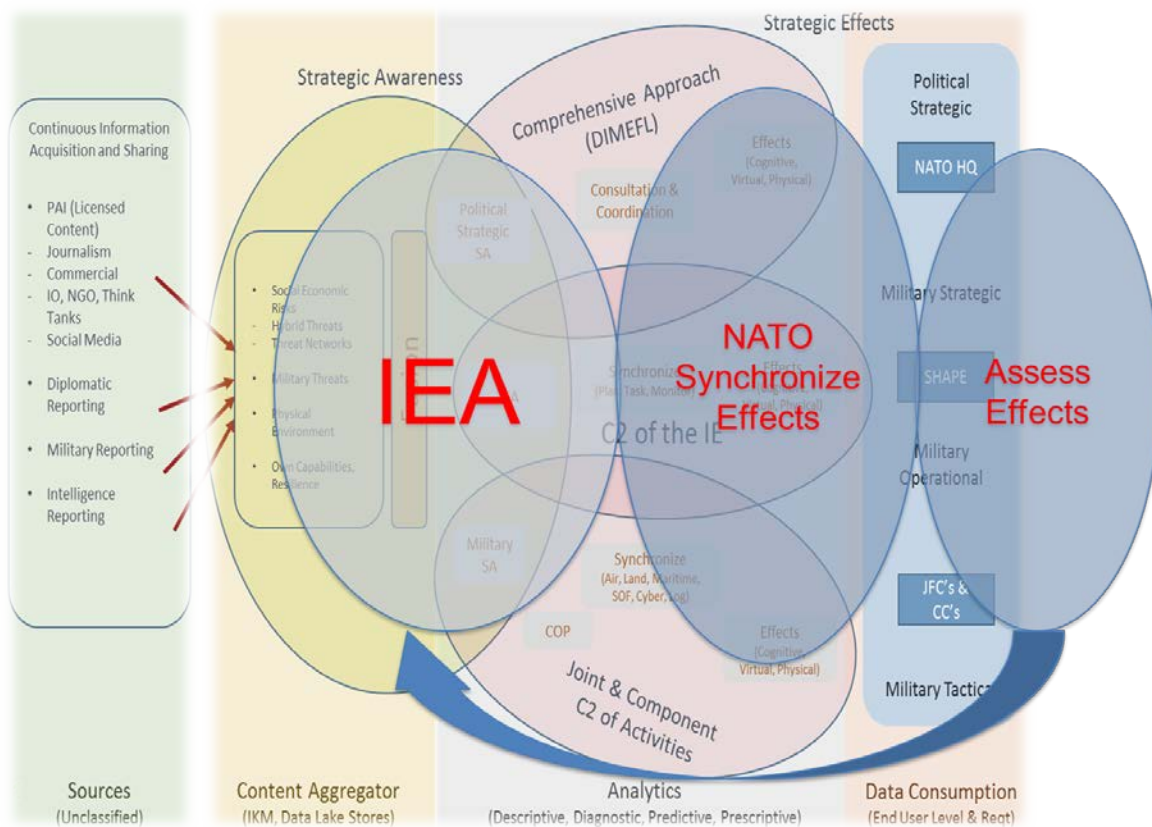


Figure 2: IEA interfaces to NATO operations and effects.

3.0 ALLIED OPEN SOURCE SYSTEM

AOSS is a functional service developed for NATO intelligence analysts. It provides access to a range of open source information sources including mainstream news sources and specialist open source providers such as Janes IHS etc. 90% of the content is subscription based [4]. AOSS currently (in version 5.5.) supports some analysis tools for intel analysts, as shown in Figure 3.

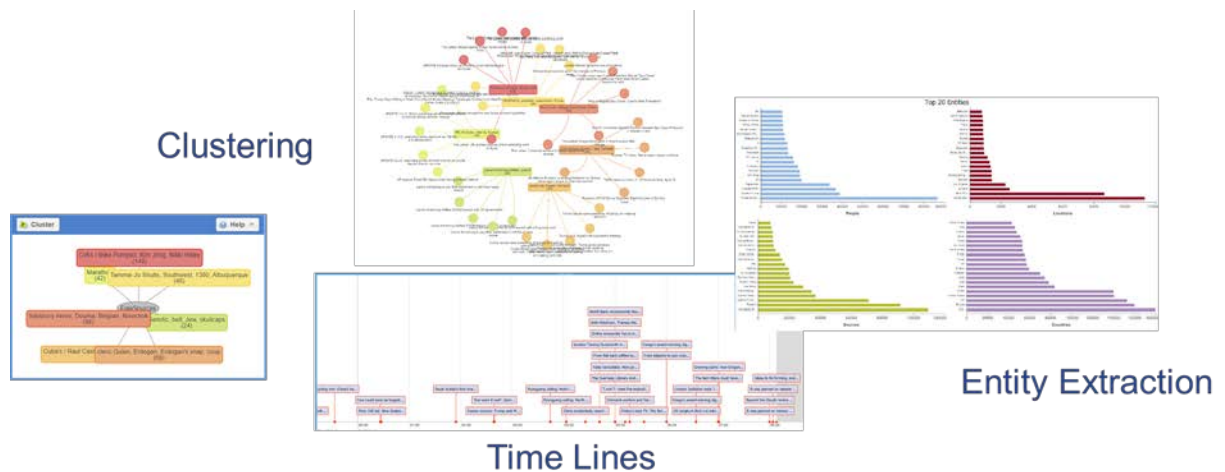


Figure 3: Data analytic functions currently supported in AOSS.

4.0 EXTENDING ANALYTICS CAPABILITY

An IEA initial operating capability (IOC) is in development, drawing on existing capabilities and expertise within NATO. This will take the existing sources of PAI and the intelligence analysis system around them, and add data analytics and data visualisation capabilities already in use in PMAR, supplemented by additional code development to provide initial data analytic services.

Since its formation in 2012, NCIA has used data analytics and a data-centric approach to harvesting data from a range of existing sources and systems into a *data mart* (or small data lake). This data was then analysed using a number of tools in order to glean meaningful data from these varied sources. Since this initial implementation the data extraction methods have been enhanced, the *data mart* has grown to a *data lake* (fed by over 50 tributary data sources) and the analytic toolset has been expanded and refined. This is described further in [3].

A key component of this existing toolset is the KNIME analytics platform. This is a leading open source data mining tool and provides a graphical framework to manage data preparation. The graphical framework significantly reduces the need for low-level programming, making the data mining process accessible to a larger group of data analysts. KNIME also supports PYTHON natively and allows leading open source deep learning libraries to be integrated e.g. TensorFlow. This provides access to a range of sophisticated deep learning resources which can be integrated with KNIME and fed by data sources prepared using conventional KNIME functionality for big data analytics.

The combination of KNIME, machine learning and deep learning models, applied to NATO data sets is addressed in depth in [5]. These functions are now being integrated into AOSS in order to extend the functionality towards an IEA initial operating capability.

The ability for machines to analyse and learn does not lead directly to insight and value from the data. It will be necessary to train and refine the machine learning / deep learning models. This training process is known to require large quantities of data, coupled with time and expertise from data scientists to refine the models, and domain experts to explain data structures, features and relationships.

5.0 RECOMMENDATIONS

It is clear that NATO is on a steep learning curve as it deploys data-centric technology to support its mission. However, it is not starting from scratch and has valuable experience in the analysis of open source intelligence (through AOSS), sophisticated tools for data analysis and visualisation (PMAR) and the application of data centric technology to support operations [3][5].

Industry will be a source of leading edge technology, and close cooperation exists between the work described in this paper, and the activities of NIAG study group 225 which is looking at architectures and technologies for IEA FOC.

REFERENCES

- [1] NATO Industrial Advisory Group Study Order 225, DI(2017)0261, September 2017.
- [2] NATO Industrial Advisory Group Study Group 208, final report “On adopting big data in NATO”, AC/322-N(2017)0085, June 2017.
- [3] M. Richter, M. Street, P. Lenk, “Lessons learned from initial exploitation of big data and AI to support NATO decision making”, STO IST-160, May 2018.
- [4] NATO open source intelligence (OSINT) policy 3.1.
- [5] M. Richter, M. Street & P. Lenk, “Deep Learning NATO document labels: a preliminary investigation”, Int. Conf. on Military CIS, Warsaw, May 2018.

