Role of Big Data in Resilience Assessment in Military Context

Ivana Ilic Mestric
Service Strategy
NATO CI Agency
THE NETHERLANDS

ivana.ilicmestric@ncia.nato.int

Giavid Valiyev
Service Strategy
NATO CI Agency
THE NETHERLANDS

giavid.valiyev@ncia.nato.int

Michael Street
Service Strategy
NATO CI Agency
THE NETHERLANDS

michael.street@ncia.nato.int

Charalampos Sarantopoulos
Service Strategy
NATO CI Agency
THE NETHERLANDS

charalampos.sarantopoulos@ncia.nato.int

Riccardo D’Ercole
Service Strategy
NATO CI Agency
THE NETHERLANDS

riccardo.D’Ercole@ncia.nato.int

ABSTRACT

This paper is intended to demonstrate the potential of open source data, coupled with big data analytics and data visualization, to indicate levels of resiliency in specific domains which cover the baseline requirements (BLRs) of NATO resilience assessment.

The proof of concept described in the paper extracts relevant resilience indicators in a given region, across selected baseline requirements that include energy and transportation. The proof of concept uses interactive dashboards that allow end-users to explore the available public data from many perspectives, as well as results of advanced analysis and machine learning models conducted on these data.

Keywords—big data analysis, machine learning, resilience, energy, transportation, media

1.0 INTRODUCTION

The military community has become more and more aware of the importance and the role of big data analysis for their operational and strategic decisions. Access to the relevant information in the right time was always the key factor to make the best decisions. Today this impact is even greater as data and information can be collected on a large scale and are available to everyone. Technology and AI approach became huge enablers in exploiting the data. [1]

Wide availability of open source data coming from media, scientific articles, relevant (expert) web portals covering areas such as economics, politics, society, energy, transportation etc. bring the possibility of creating more insightful context and provide valuable new dimension to any kind of assessment by analysis of varied sources and consolidation of results.

From a military perspective, we identified many important indicators across different areas coming from open source data that can be relevant for assessment of readiness and resilience across the alliance. It appears that many of the indicators coming from different areas have an influence on each other and can be related.

During the last year, NATO CI Agency Data Science team was involved in an innovative proof of concept with transformational and operational commands e.g. ACT, SHAPE and JFCBS; in order to identify, extract, calculate and present the most relevant indicators available from open source data to support assessment of resilience across the alliance. As resilience is a complex assessment, which relies on relationships of many different areas and events, the project defined smaller scope focusing on the following critical areas:

- Critical Infrastructure - hospitals, power plants, ports, LNG terminals and military facilities
- Energy – focus on electricity and gas
- Transportation – focus on air, road, maritime and near real time traffic indicators
- Media – situational awareness

The main objective was to identify the relevant indicators by using big data from publicly available datasets. Then creating useful curated data and Machine Learning (ML) models to identify relevant relationships and provide insight on current situations and the impact of disruptive events. To improve the accuracy of results, we initially focused on a specific geographical region.

2.0 ROLE OF BIG DATA FOR RESILIENCE ASSESSMENT

To increase Resilience, we try to identify and predict events that have low probability but a large impact. By definition, probability and impact are measures that are tidily connected to, and derived from, data.

In order to understand the role of Big Data in Resilience, we can consider the Resilience problem through five key components, also known as 5 Vs: Volume, Variety, Velocity, Veracity and Value.

2.1 Volume

Access to a huge amount of accessible data can be valuable for Resilience to monitor trends and predict future events. The huge amount of data help us to determine the baseline and the state from where more detailed analysis could start.

2.2 Variety

Variety refers to diversity of data types and data sources. Being able to process and exanimate unstructured data, including images and video streams, which comprise more than 80% of available world datasets presents great value for extracting resilience indicators. Applying advanced analytics and Machine Learning algorithms provide us unified view across all variety of data.

2.3 Velocity

Considering data of interest is generated very rapidly and more and more data sources are available in near-, or even real-time, it provides an opportunity to continuously monitor the impact and extract signals and warnings with great precision. For Resilience it is also valuable from the aspect of anomaly detection and to give early warning predictions.

2.4 Veracity (authenticity and credibility of the data)

This has a big reflection on Resilience indicators since data that are analysed are coming from all different sources. Some of them come from reliable and authoritative sites whereas other can include biased or untrusted data.

2.5 Value

Although today's technology enables data exploitation on a huge scale, the important aspect is to recognize and focus on the benefit for users and decision makers as the main result coming from the process. This Resilience experiment, the main indicators for the value is the end user's feedback and usage of the final dashboard.

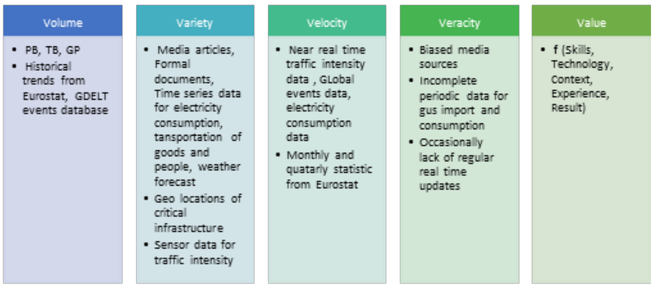

Figure 1: Role of Big Data for Resilience assessment – mapping

There have been several researches that tackle similar approach to Resilience assessment. The most detailed research is Big Data for Resilience Story book [2] that describes big data analytical model and humanitarian scenarios relevant for resilience. The other relevant research more focused on disaster resilience is described in [3].

3.0 PROBLEM STATEMENT

The key question for the experiment was how to use emerging technologies and Big Data, to assess Resilience indicators over publicly available data relevant for military operations?

3.1 Data sources

The beginning of the research was focused on identifying relevant data sources that are related to the areas of interest. Although there are a numerous publicly available data sources, only a limited number of them can be considered reliable and come with open APIs that properly reflect historical and current situations. In our experiment we mostly focus on European and national sources for relevant Energy, Transportation and Infrastructure indicators, which is listed in Table 1.

3.2 Model

Finding the relevant datasets in the public domain and their collection represents a huge effort, but defining the model that can provide answers and insights relevant for Resilience is the main challenge that will keep our attention most of the time.

To be able to identify events with low probability but huge impact we started with following approach:

- Define Indicators baseline using time series analysis that will reflect historical and current trends
- Train Machine Learning models and Time series models (ARIMA) based on historical data and relevant external indicators in order to have short and long term prediction for selected indicators
- Create what if analysis to address sensitivity of certain indicators
- Analyse media to identify trends and current narrative concerning the baseline requirements in a

given geographical area

- Identify cause and effect using media trends and indicators across the baselines.

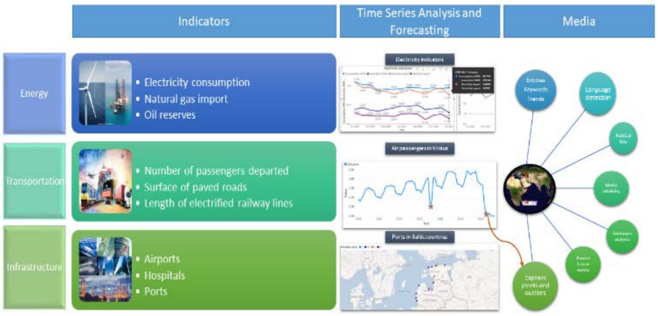

Figure 2: Big Data model for Resilience

Table 1: Data sources used for proof of concept

Source	Description	Frequency / granularity	Area
European Network of Transmission System Operators for Electricity (ENTSO-E) transparency platform **Error! Reference source not found.**	Electricity generation, capacity, consumption, flows (imports/exports)	Hourly, monthly	Energy
European Network of Transmission System Operators for Gas (ENTSO-G) transparency platform **Error! Reference source not found.**	Gas capacity, generation, flow (imports/exports), operators, connection points	Hourly, monthly	Energy
Gas Infrastructure Europe (LNG and STORAGE database)	LNG terminals and amount stored data	Daily	Energy
Aggregated LNG storage inventory	Gas storage data	Daily	Energy
EUROSTAT – European Statistics **Error! Reference source not found.**	Data sources for number of indicators	Monthly, Quarterly, yearly	Transportation, Energy
GDELT – global database of events **Error! Reference source not found.**	Global events database, Scraped public media articles from web	Hourly	Media
Media APIs: Current News API, News API, GNEWS API, The Guardian API, NY Times API	New feeds from publicly available APIs	Daily	Media
Road monitoring in Lithuania	Governmental sources from transportation department (intensity, restrictions and cameras)	Hourly, daily	Transportation
Trans-European Transport Network TENTEC - European Commission (europa.eu)	List of European transportation projects.	Static	Transportation
World Port Index	Data sources for Ports	Static	Transportation, Infrastructure

4.0 BASELINE REQUIREMENT FOR PROOF OF CONCEPT

The NATO Resilience Guidelines and evaluation criteria [4], have seven baseline requirements to assists nations in conducting national self-assessments of their resilience. As already mentioned above, for the proof of concept we focused on two baseline requirements (BLRs), energy and transportation, with the addition of media analysis to provide better understanding across different indicators. Also, we used guidelines for extraction of relevant indicators per each BLRs.

4.1 Energy

Open data on energy generation, consumption and distribution are sourced from European and national sources, as mentioned in Table 1.These cover a range of energy related categories such as Electricity, Natural gas and Oil products. This data is analysed and modelled to show dependencies, past trends and future forecasts. In this paper we will only mention a smaller subset of analysed indicators that represent certain sensitivity to the military context.

Figure 3 shows factors impacting gas supply resilience in a given nation, from demand patterns (left) and the predicted number of days' supply available in reserves (right).

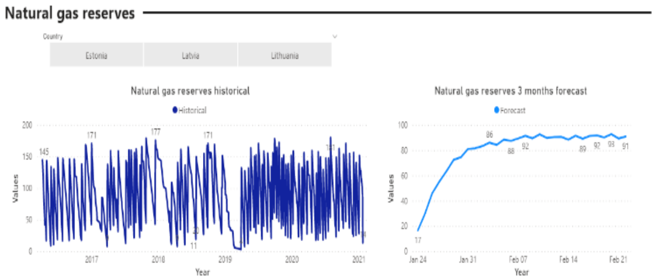

Figure 3: Natural Gas reserves

The Figure 4 shows locations of critical gas infrastructures (storages, LNG terminals, connection points) with details and gas flow (bottom) between different connection points. This view provide us clear picture of dependencies and critical points for gas distribution.

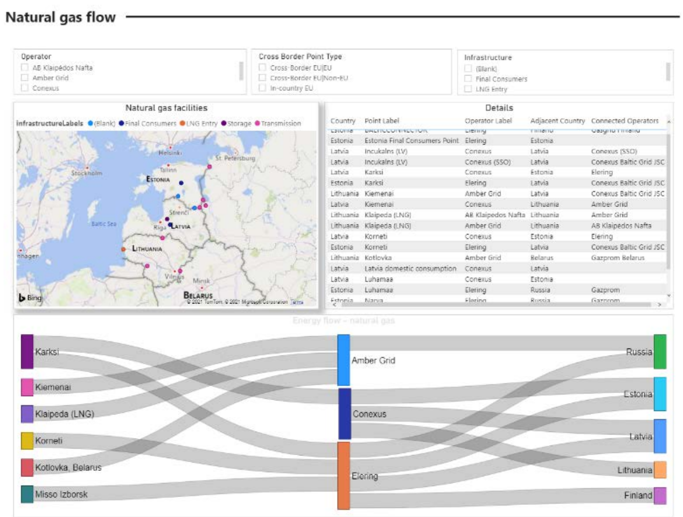

Figure 4: Natural Gas flow

Short-term prediction of electricity consumption and feature importance can be seen in Figure 5. The illustrated example shows how to anticipate electricity demand depending on external parameters such as weather conditions or holidays. We treated the above prediction as a supervised Machine Learning problem where we used historical time-series electricity data in conjunction with external factors for generation of relevant features that we used for training the model. The importance of the features is determined by the model itself; the non-significant features are disregarded by the model whereas the most significant are the ones assigned the largest weights. The use of this analysis in a military context can correspond to assessment of "electricity gap" if there is a sudden steep rise in electricity consumption due to a military exercise in the area (without providing its own power supplies) or due to failure in the electrical infrastructure.

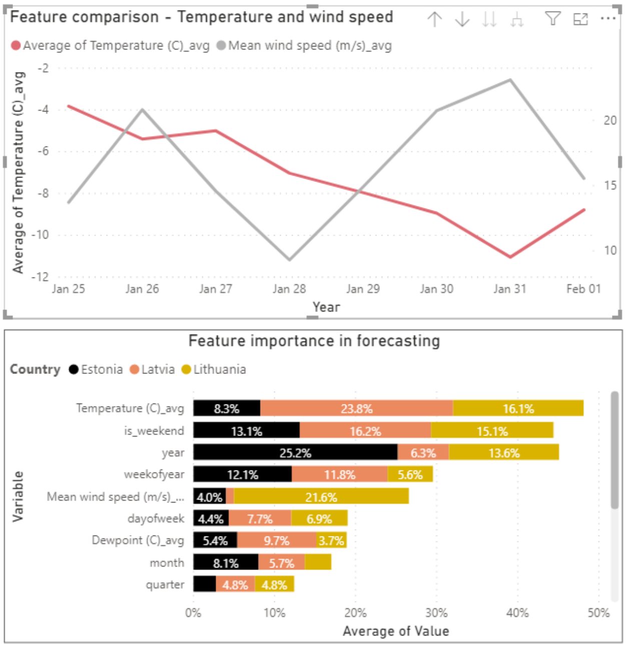

Figure 5: Electricity consumption prediction and Feature importance

Electricity Import dependency with partner countries, Figure 6, and what if analysis of available electricity, Figure 7, demonstrate the sensitivity component of Resilience that can be highly impacted by political atmosphere in the region.

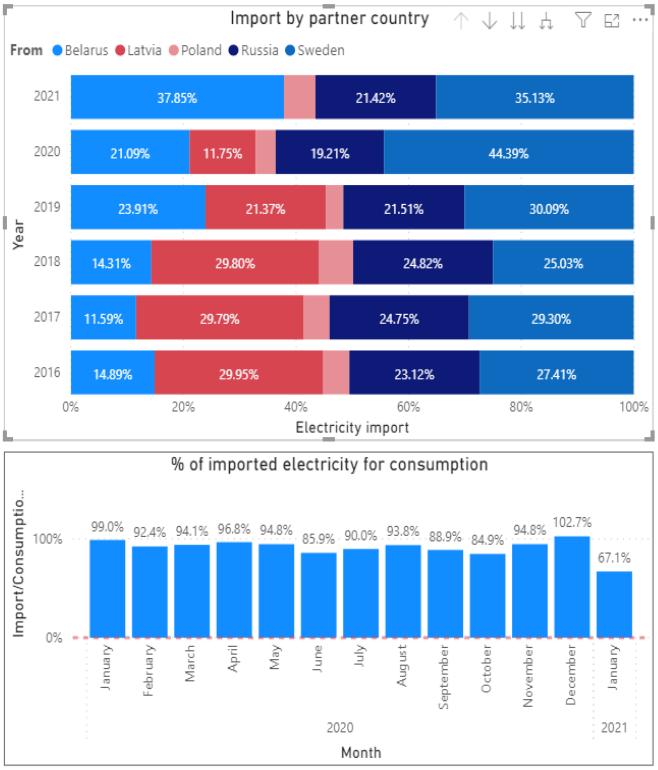

Figure 6: Electricity import dependency

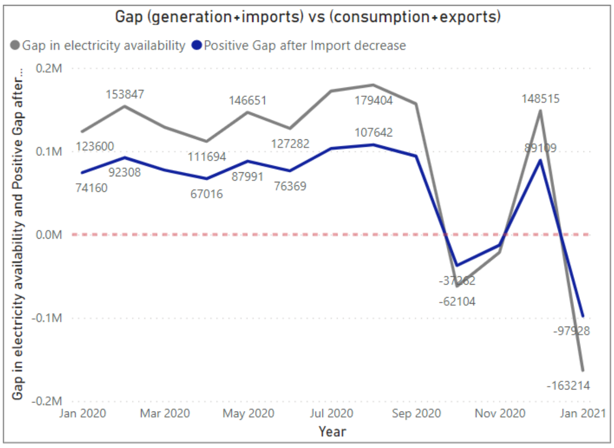

Figure 7: What if analysis

What if analysis enables military analysts to change a parameter and see what is the effect on overall availability of electricity capacity in certain period. This is the method to assess sensibility on external factors such as political or environmental (natural disaster) events.

4.2 Transportation

National and European data sources are again used to give an overview for each country of aspects of transportation. Both historical and near-real-time data can provide information and patterns about different transportation indicators related to road, rail, air, and maritime. This is supplemented with near-real-time data that gives updates on traffic restrictions and camera feeds that enable remote visual inspection of road and weather conditions in areas of interest. Resilience in transportation is described in terms of infrastructure capacity, actual usage and how fast it reacts when shocks occur.

Example of such a shock on the system can be seen in Figure 8 where we used the model to predict the movement of passengers based on historical data, but the situation caused by COVID-19 had a huge impact starting with early 2020 (black line in Figure 8). Unarguably, this was not anticipated, and like any other shock, the consequences should be measured.

Figure 8 Forecast of passengers arrived / departed

The other important transportation indicators refer to road accidents and traffic intensity. Knowing the coordinates, we have created a heat map indicating areas where the accidents took place, Figure 9. Additionally, there is information about circumstances under which the accidents occur such as type of the road, the weather conditions that prevailed at the moment, the lighting conditions but also statistics such as number of involved people, casualties, injuries, type of involved vehicles etc. Finally, relating all the parameters together we can identify clusters of road spots that are prone to accidents.

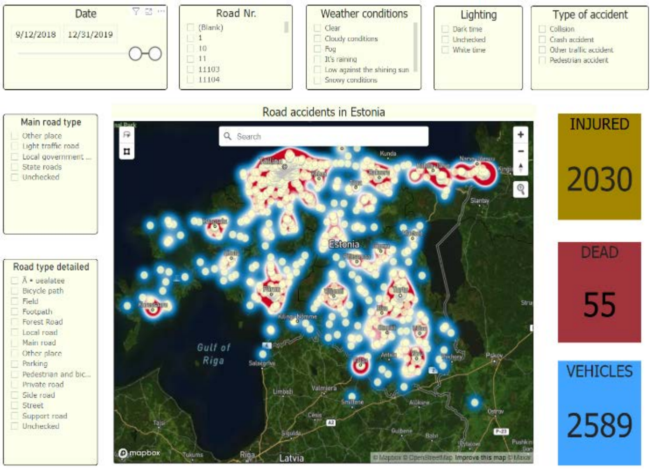

Figure 9: Roads prone to accidents

Using historical data, we can identify part of the roads that have high traffic density, Figure 10. Additionally, monitoring of traffic intensity can be near real time (every 15 min) via sensors installed in specific parts on the roads. The sensors are scattered throughout the country on highways and regional roads.

These indicators can help military exercise planners to focus their attention while planning the routes of convoy movement. Particularly as previous exercises have shown that road accidents during an exercise are a significant concern for military and civilian authorities.

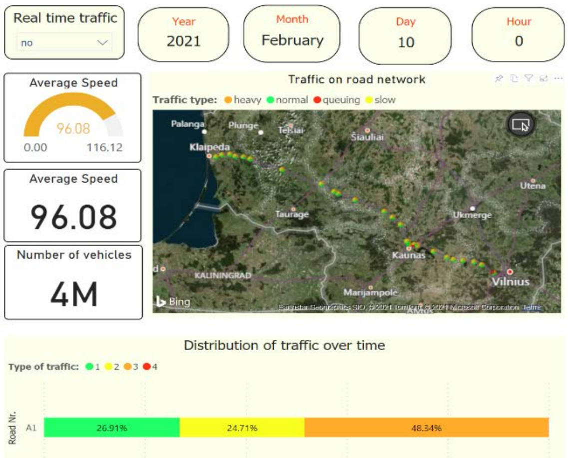

Figure 10: Real time traffic density

4.3 Critical Infrastructure

One of the key requirements for resilience is the ability to identify critical civilian and military infrastructure and facilities in the area of interest. The approach was to research the relevant open source data that can provide an overview of such cross-referenced visualization for the countries of interest. Additionally to data sources listed above, Google API provided valuable geolocation data for the relevant military and civilian facilities and points of interest. Critical infrastructure dashboard, Figure 11, represents the map with marked critical infrastructure facilities/locations including power plants, hospitals, military facilities, ports, LNG terminals and electrical transmission points. Additionally to geographical points, we have related details for each facilities such as number of beds in hospitals, size of ports, generation power of power plants with outage statistics etc. These details provide the ability to assess the impact of a significant event in the area. The dashboard view allows a military analyst to observe the area of interest an in selected radius (in km).

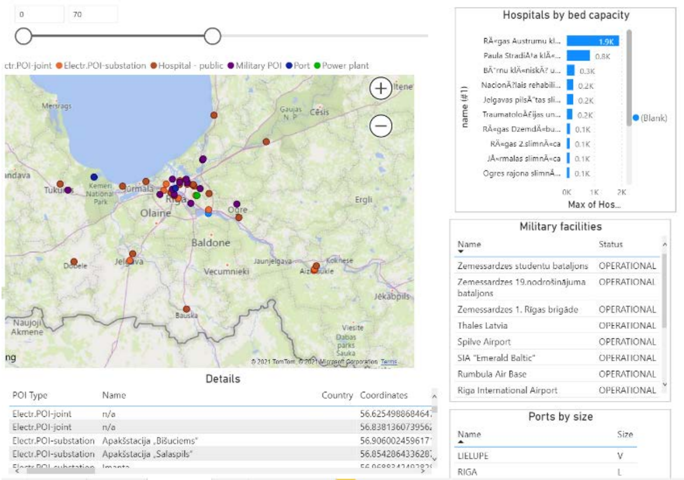

Figure 11: Critical infrastructure

4.4 Media analysis

During the research of possible relevant data sources for resilience baseline requirements, we identify media sources as important in a way to provide us valuable insight and situational awareness. For that purpose, we used International APIs and the GDELT database with events dating back to 2015, covering 44 different languages. Certainly, this can be considered as Big Data analysis from many aspects, especially considering the volume and veracity. These datasets were processed with machine learning techniques to perform named entity recognition, sentiment analysis, article categorization, and anomaly detection techniques. The results can be seen in Figure 12 where we can see an "Impact of relevant event's categories vs Sentiment analysis" for given period and geographical area. Where sentiment can be used as an indicator of civilian impact. Also, clustering of similar articles can help us to identify the main events and the most relevant topics.

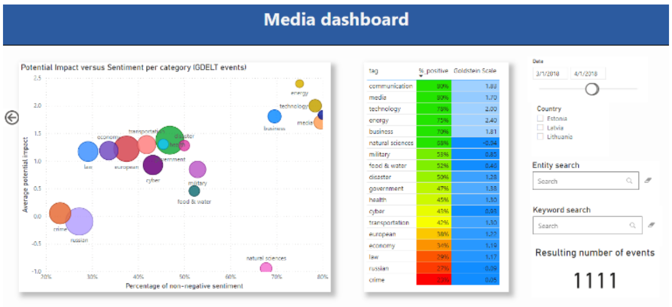

Figure 12: Impact of event's categories

The other relevant media analysis provides us topic trends over time and locations related to BLRs, Figure 13. This approach can help us to identify hot topics and relevant event during the time and in conjunction with other indicators provide us better understanding of a situation, cause and effect.

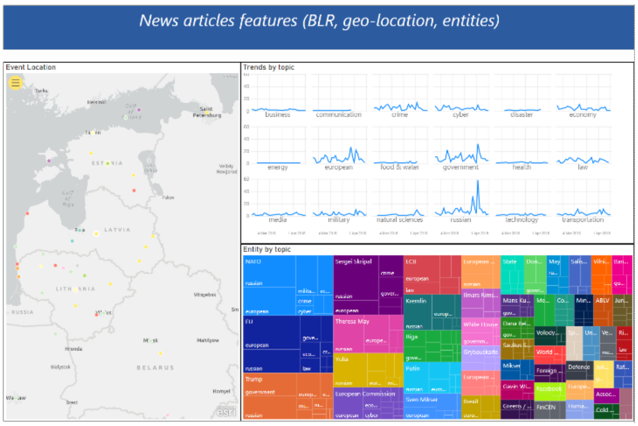

Figure 13: Topic trends over time and locations

5.0 CONCLUSION & RECOMMENDATIONS FOR FURTHER WORK

This work has been the first step towards identifying the military requirements to assess resilience using publicly available big data and machine learning. Feedback from end users shows that there is lots of potential and value in big data analysis for resilience assessment. This prototype has the potential to become a fully automated tool in order to provide the end user both the historical background needed for a high-level understanding of the resilience landscape, but also to provide real time information for tasks requiring updated information (exercise planning, recovery from a disaster etc.).

The next step is automation of collection, data parsing and cleansing processes, so that it continuously reflects the current situation and minimizes human interaction (that can possibly lead to mistakes or prolonged maintenance). Further work should be done on media analysis to better scope media articles based on keywords and topics relevant to resilience baseline requirements. This can help us to identify relevant historical events, from a resilience perspective, and extract the features related to them which can be used later in ML models to predict similar events. Finally, we want to combine the critical indicators from different baselines, including the media, and enable military analysts to easily identify anomalies and recognize the early signs of events could test resilience.

6.0 ACKNOWLEDGMENT

We would like to thank our colleagues in inputs, guidance and continuous support. Without their contribution, our work will be more difficult and less exciting.

7.0 REFERENCES

[1] NATO, 2019, EDT Strategy

[2] Ospina, A.V. (2018) Big Data for resilience storybook: Experiences integrating Big Data into resilience programming.

[3] Group of authors, 2020, Disaster resilience through big data: Way to environmental sustainability

[4] NATO, 2017, “EVALUATION CRITERIA ON RESILIENCE.

[5] European Network of Transmission System Operators for Electricity (ENTSO-E) transparency platform, https://transparency.entsoe.eu/

[6] European Network of Transmission System Operators for Gas (ENTSO-G) transparency platform, https://transparency.entsog.eu/

[7] EUROSTAT – European Statistics, https://ec.europa.eu/eurostat

[8] GDELT – global database of events, https://www.gdeltproject.org/