Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Data Driven decision support during COVID

Michael Street, Ivana Ilic Mestic, Adelica Ndoni, Peter Lenk, John Teufert, Nuno Figueiredo

NATO Communication and Information Agency, The Hague, Netherlands

Abstract

This paper outlines the development and use of a tool suite developed by the NCI Agency to provide situational awareness and decision support during the current Covid-19. The tool suite was developed to understand how Covid-19 could impact the provision of communication and information services (CIS) to NATO, and so understand where risks to NATO operational functions might occur. The tool suite combines open source data on instances of Covid-19 globally along with internal information about the impact of Covid-19 on NCI Agency staff and the services they deliver to the NATO enterprise. It supports business impact assessments due to Covid-19; showing trends, age demographics, and providing early indications of critical services that may be affected, sites that may be affected, etc. The tool suite is an example of data science techniques supporting data driven decision making within a military organization.

© 2022 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the International Conference on Military Communications and Information Systems

Keywords: data driven decision making, advanced data analytics, named entity recognition, entity extraction, data visualisation, Covid

1. Introduction

During the early stages of Covid-19, the NATO communications and information agency (NCIA) developed a situational awareness and decision aid tool suite, to support management understanding and decision making within NCIA. The tool provides both global and local views, as well as external and internal perspectives. It has been developed to understand the current and potential impact of Covid-19 on delivery of the Agency’s services. This is based on the current impact on Agency staff, which services staff provide plus Covid situation in the vicinity of NATO locations and any limitations which this imposes.
The tool was created using existing data science tools and components. It harnessed three distinct types of data:
- Existing data within NCIA;
- New data collected within NCIA; and
- Publicly available data.
These data sources were combined in order to understand risks to staff, the services they provide and therefore potential impact on the operational functions those services support. They were also used to support effective functioning of the organisation while external factors changed rapidly. For example, during lock downs it was essential for some NATO travel to continue to support ongoing operations, this tool was used to manage travel more effectively in the very dynamic situation.

Section 2 outlines the data sources used for this work, section 3 describes the approach taken for data analysis and visualisation, section 4 highlights main results and section 5 concludes.

2. Data Sources

At the outset of the Covid-19 pandemic, many publically available data sources, such as those from John Hopkins University [1], were available and used by a variety of organisations to visualise the spread of the virus. Public data was used by media organisations for reporting and by government organisations to inform responses.

2.1. Public data for situational awareness

The tool suite developed by NCIA also used such common public data to give a global geographical view of the pandemic status but which could zoom in on the national or local situation. Data was also scraped from relevant news feeds and world health organization bulletins. Additional data relating to travel restrictions was also sourced from government and commercial sources, regarding airport status, travel restrictions etc.

During the early stages of the pandemic there were no accurate sources for such information on a NATO-wide level, and often even national data was not comprehensive, leading to the need to scrape multiple sources e.g. national health and transport ministries, regional government sources, transportation companies; to fully understand whether a journey would be possible and what additional restrictions were in place.

2.2. Internal data sources

Available data within NCIA was also used by the tool suite. This included:
- human resources data e.g. staff location, age, etc;
- service management data on ICT services;
- time management data e.g. who contributes to which ICT services.

In addition to this existing management data, at the start of the pandemic additional data was generated as staff reported on Covid-19 tests, self-isolation, working onsite, vaccination progress, etc.

3. Tools

A set of tools is being used for this work to combine different data sources, preprocess, analyze them and visualize the results. These tools are described in the below paragraphs, one of each tool.

3.1. KNIME Analytics Platform

KNIME analytics platform [5] used for much of this work is one of the leading open source data mining tools available; it provides a graphical workbench for visual assembly and interactive execution of data pipelines. It
features a powerful and intuitive user interface, enables easy integration of new modules or nodes, and allows for interactive exploration of analysis results or trained models. KNIME offers the possibility to retrieve and combine data from any source, like simple text formats (csv, pdf, xls, json, xml, etc), unstructured data types (images, documents, etc) or time series data. It is also possible to access and retrieve data from sources such as SharePoint, Azure, Oracle or Microsoft SQL databases which was also one of the main source for our analysis in this work.

KNIME provides an easy integration with some other leading tools used in data science such as Python and R when needed. This makes the data scientist’s job flexible, allowing the use of the best combination of these tools and possibility to approach the problem from different perspectives.

3.2. **Microsoft Power BI**

The NCIA’s Data science team has adopted Microsoft Power BI as a visualization tool to present the results in a dynamic and friendly-use dashboard to the end users. Although KNIME provides the possibility to explore the results by plotting them in specific data visualization nodes, the presentation is not as compelling as it could be with more specialized data visualization tools.

Given the availability in NATO domains and ease of use of Microsoft Power BI we have opted to create interactive dashboards allowing domain experts to visually explore and gain insights into the reasons behind the scores.

3.3. **ArcGIS Enterprise**

As individuals, we make decisions based on geography every day. The COVID-19 crisis is certainly an example of that. NCIA has used the Esri technology [7] to map and analyze the public available data in the context of location, providing web based geospatial tools such as Geo oriented dashboards and 2 and 3-D viewers to provide situation awareness. This tools makes it very intuitive and helpful to the users for explaining and exploring the available data in an understandable way, in a map! Putting the powerful GIS tools in the hands of people who might not otherwise have any experience with it.

4. **Data Analytics**

Extracting and combining the available data was performed using structured data science services already established within NCIA. These are available in the same security domain as the majority of our business systems. The tool suite used a structured set of services, shown in Figure 1, used to develop data science and AI solutions [2]. Data is always the foundation and the sources described above needed some ‘curation’ to clean, prepare and structure it before applying data science techniques.

The infrastructure layer provides the fast storage and graphical processing units (GPUs) needed for handling huge data sets and for rapid training of neural networks. Although in this project neither were used heavily, for most data science applications they are essential.

![Fig. 1. Structured Data Science environment](image)
Sitting on this infrastructure we have the Platform – a set of tools to allow data scientists to develop their solutions. KNIME is the open source data analytics tool that is used for data pre-processing (ETL, Extract-Transform-Load) phase and more advanced analytics of the available data. Specifically in this project, descriptive statistics are used the most by quantitatively summarizing the features from the collection of data. Power BI is the tool used for visualizations of the results in an interactive dashboard and ArcGIS for geographic representation of them. Existing KNIME workflows were used as the basis for the entity extraction techniques used to identify names and services, and to consolidate data from different sources [3][4].

Using the existing structured services allowed rapid development of specific Data Science Software to address specific questions described here. This was created by in-house staff with the Data Science Expertise needed both to put those blocks in place, and to pull them together into solutions that can provide ‘better decisions, faster’.

5. Results

5.1 External data to assist internal decision making

Dashboards were developed to assist with business functions, for example, to assist in planning and approving the limited, essential travel undertaken by staff, the dashboard below presented health risks and travel restrictions in a single, searchable dashboard, fed by many publically available data sources. This view shows travel restrictions in place at airports and ports, based on publically available information.

![Fig. 2. Travel Impact Dashboard](image)

This dashboard (see Fig 2) could be explored to see the current local situation on a local level, for example to focus on specific NATO locations (see Fig 3).

Pooling dozens of external data sources into a single dashboard which covered all 30 NATO Nations and other countries of interest provided a one stop shop for all decisions regarding travel and transport for NCIA. Use of this dashboard was incorporated into the rapidly revised business processes for travel during Covid.
5.2. Internal data for situational awareness

Interactive dashboards were built to present the pandemic situation of the NCIA staff members and locations. These dynamic dashboards were very helpful to Agency management team and health advisers which led to more informative and faster decisions taken during Covid.

Figure 4 shows an interactive dashboard shows the number of NCIA staff self-isolating or tested for COVID, allowing Agency management to see overall numbers and trends. The top left of Figure 4 shows the numbers taking Covid-19 related action each day while on the top right we see the overall number of staff affected, at this point less than 1% of all staff. On the bottom is the overall number of self-isolating staff, with and without symptoms and when they are planned to return to work.

Related dashboards based on this data also showed the daily number of self-isolating staff, with or without symptoms as well as the number of staff tested for Covid-19 and their results.
These results were also used to predict when staff who had tested positive or were self-isolating would be returning to normal duties, aiding management decision making.

Figure 5 shows the age range of staff who are self-isolating on the left. This was as useful parameter given the extreme age-dependence of Covid-19 risk. While below shows the location of self-isolating staff, useful to understand how individual locations were affected.

5.3. Data driven decision making

Figure 6 shows how self-isolating staff were distributed across the organization, based on data which was updated several times per day. These figures are taken from an interactive dashboard, allowing management to drill down into details for any location, organizational element or service. All numbers can be filtered to specific locations, organizational elements or services to give a detailed picture of the impact, as shown in the figures below.
The Agency provides ICT services to NATO which depend on groups of key staff. Figure 7 shows the numbers of key staff (and age ranges) supporting some specific key services in absolute numbers (top) and as a percentage (bottom) to help identify risks to service provision, such as services provided by staff with a large proportion at greater risk if they contract Covid-19.

Figure 8 shows similar information, but specifically relating to the location and organization elements. This is arguably the most important dashboard as it focused on key staff who supported specific services or operational users at a given NATO location. The upper section of this dashboard shows the % of key personnel at each NCIA location who were self-isolating on a given day (top).
The lower half of Figure 8 shows which services are most impacted by key personnel in self-isolation. This view allowed NCIA management to identify where support to NATO users might be at risk because of staff at a given location, or delivering key services, might not be available to fully support their NATO users.

![Fig. 8. Impact and Risk to Services](image)

This allows the Agency to ensure it keeps essential ICT services available to NATO, allows management to plan support for functions which may be at risk if additional staff fall sick, and aids the protection of staff at greatest risk.

On-site tracker (see Fig 9) was another effort that NCIA took during the pandemic in order to facilitate the decision making process when it comes to ‘work from home’ advice. This trend and its correlation with other analysis, for example number of positive cases in a specific period, was found helpful for making informed decision based on the analysis results at that point of time.

![Fig. 9. On Site Tracker](image)

Another important trend to analyze was the vaccination program and how it progressed within the Agency. Analyzing the number of people being vaccinated by a specific duty location or service line, is, again, very helpful for the management team to be better informed in decision making.
Figure 10 shows this distribution of vaccinated people between different locations and service lines within The Agency.

6. Conclusions

This work demonstrates the value which can be brought by combining disparate data sources – both public and internal data - and from applying advanced data analytics and visualization. The tool suite that has been developed allowed data driven decision making to de-risk and ensure provision of ICT services for NATO’s operational community, particularly during the early stages of the pandemic when infection rates were high, treatment developing and medical aspects of the virus were still unknown.

The use of existing, structure data science services allowed this tool suite to be developed on an extremely short timescale. The first iteration of the tool was available during March 2020. Refinement of the tools continued with additional data and functionality being added into 2021.

The tool suite was accessible to all staff across NCIA, although some detailed views were limited to those responsible for service provision or health and safety.

This tool suite demonstrated the potential of data science to rapidly enable data-driven approaches to situational awareness and decision making by combining public and private data sources. This was done to improve decision making and increase resilience of NATO ICT service provision, in a very short timescale, in the face of an unexpected scenario.
Acknowledgements

The authors wish to thank colleagues across NCIA who enabled access to internal data sources, and to the many experts who made important information publically accessible.

References

[1] John Hokins University, https://coronavirus.jhu.edu/map.html, accessed 15 March 2022.
[2] Lenk, Peter, Michael Street, Ivana Ilic Mestric, Arvid Kok, Giavid Valiyev, Philippe Le Cerf, and Barbara Lorincz. "Data Science as a Service: The Data Range." Information & Security: An International Journal 47, no. 2 (2020): 157-171.
[3] Ivana Ilic Mestric, Arvid Kok, Michael Street, “Named Entity Extraction in a Military Context”, Int. Conf. on Military CIS, May 2019.
[4] Ivana Ilic Mestric, Arvid Kok, Giavid Valiyev, Michael Street, Peter Lenk, Mihaela Racovita, and Filipe Vicira, “Extracting Value from NATO Data Sets through Machine Learning and Advanced Data Analytics,” IST-178 Specialists meeting on Big data challenges: situational awareness and decision support, Budapest, October 2019.
[5] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kotter, T. Meinl, P. Ohl, C. Sieb, K. Thiel and B. Wiswedel, KNIME: The Konstanz Information Miner,” in Studies in Classification, Data Analysis, and Knowledge Organization, Springer: 2007.
[6] Microsoft Power BI, https://powerbi.microsoft.com, accessed 15 March 2022.
[7] ArcGIS, https://www.arcgis.com, accessed 15 March 2022.