

Defence Digital: Dashboard for interoperability

Problem statement and Executive Summary

Given data collected during planning and execution of CWIX, how can we measure levels of interoperability?

Addressed through:

Careful crafting of KPIs and their limitations. Providing recommendations for future KPIs and their associated data that could be collected.

What Key Performance Indicators (KPIs) can be developed from the data provided to measure the level of Multi-Domain participation of Nations and Capabilities at CWIX?

Addressed through:

Primary dashboard KPI

- 1) Number of capabilities tested by operational domain (Capability x Operational domain). This measures the number of capabilities that have been tested by one or more operational domains.
- 2) Percent variation of test cases over CWIX exercise cycles.
- 3) NATO planned projects against Capabilities at CWIX.

Secondary dashboard KPI

- 4) Percent of standard shortfall against the total test cases.
- 5) Number of capabilities from each nation.
- 6) Percentage of Not-tested with Overall-result.

How can such indicators be presented visually to key stakeholders such as NATO Senior management and participating Nations?

Addressed through:

This will be primarily addressed using the primary dashboard as this is aimed at getting the main KPIs to key stakeholders. If more detail is needed, they can click through individual graphs to get a more detailed data representation of the chosen theme.

Given any two CWIX events, how can we say whether we made more progress in one over the other?

Addressed through:

Many of the dashboard metrics have CWIX exercise cycles (time domain) mapped to a certain data field. The dashboard becomes increasingly useful to key stakeholders as more exercise cycle data is input. Trends in the KPI graphs can facilitate discussions on how well interoperability is being achieved.

Table of Contents

Problem statement and Executive Summary	1
Table of Contents	2
Level of innovation of proposed solution	2
Assessment of viability of solution	3
Ideation process of deriving KPI	3
Key Performance Indicators	4
Plan for demonstration of solution	6
Future KPIs recommendations	6
New information we need	7
Recommendations	7
Appendix	8

Ideation process of deriving KPI

Understand the Problem

The core of the problem is to quantify interoperability. The corollary to this is how best to present the quantification of interoperability to key stakeholders.

Develop possible solutions

We started by brainstorming KPIs using the data available. We came up with one sentence definition with an underlying narrative which served to provide context of the importance of the KPI. From the multitude of KPIs we developed, we narrowed down the metrics via choosing ones that added the most value. This was judged through discussions around which were the best indicators to measure interoperability whilst simultaneously being easily consumed by senior stakeholders.

Prototype test and refine

A lot of KPIs were scrapped due to lack of data or limitations on the effective measuring capability of the abstract concept of interoperability.

We extracted the data from Azure as a PostgreSQL database. We used pgadmin to access and assess the data. We tried to export into looker studio but the PostgreSQL connector would not work. We decided to move forward with using a free trial version of Power BI.

We first created a data model in PowerBI using the reference data model that was given. In this way, the data tables are coherent throughout the process of creating all the metrics required for the dashboard. The data model can be utilised by NATO engineers so that they can easily explain the derivation of the KPIs to key stakeholders.

We kept in mind ideas for measuring interoperability that we could deploy in the future with additional data points. We understand that some of these are *future-facing* metrics that cannot be quantified at present but will in time add value if we can measure the data.

We were given a new set of NDPP operational data on Tuesday morning. This data was evaluated as a possible additional KPI as it provided a way of comparing the strategic imperatives (which NATO operational domains are the priority for Operational deployment vs the actual domains that are tested at CWIX)

Implement

Once the prototype has demonstrated that it works for two exercise cycles we would want to extrapolate to include more CWIX data

Critical Success Factors

- a. How to measure levels of interoperability.
- b. Articulate effectively to key stakeholders.

Key Performance Indicators

Primary Dashboard

1) * Number of capabilities tested by operational domain (Capability x Operational domain) ****

This measures the number of capabilities that have been tested by one or more operational domains.

To be considered when consuming this KPI:

- This KPI has an advantage of incentivising each capability to be applicable as broadly as possible. However, this may not make sense for some forms of capabilities. This point is expanded below.
- Theoretical maximal of interoperability of a capability can be used and compared against the current domain capability. If it is a capability of just one domain (e.g. Air). It doesn't make sense to measure how effective it is across domains. This KPI will negatively impact this capability even though it should not.

Data attributes needed: Capability_Operation_Domain, Exercise_cycle (this is located under testcase)

Measured as: a cumulative bar chart comparison between years as a percentage increase from CWIX 2021 and 2022 with number of capabilities on the Y axis

Narrative: If the number of capabilities that are across multiple domains have increased this is a good measure of improved interoperability.

2) Percent variation of test cases over CWIX exercise cycles.

This can be measured as a bar chart of CWIX 20 and CWIX 21 overlaid on each other.

Narrative: It is important to understand the results of testing at CWIX as a trend. This data can be used to drive appropriate actions depending upon these results.

Success: Process flow to drive forward the implementation to all nations can be created for each capability that falls into this category.

Limited Success:

Not tested:

Interoperability issue: Another dashboard can be used to drill down to the reason why there was the test result. Is it Political/Engineering/Standards/other. Process flow to steps needed to resolve the issue.

Data Attributes: Testcase, Exercise Cycle.

3) NATO planned projects against Capabilities at CWIX.

Narrative: This KPI is used for comparing the strategic imperatives (The priority NATO operational areas needed for Operational deployment vs the actual capability domains that are tested at CWIX).

Limitation of data: There is not a clear one to one relationship between the operational areas and the capability operational domain so we have had to assume a relationship between data fields "Joint/Ending", "Stability/Reconstruction" and "Special Operations Forces" and "Other support services"

Data attributes needed: NDPP_Operational_area, capability_operation_domain

Secondary dashboard [Partially built]

4) Percent of standard shortfall against the total test cases.

Narrative: % of test cases that fail as a result of a badly defined standard that NATO has to go and fix.

This should be decreasing as time goes by and improvements to standards are made.

Data attributes needed: lo_shortfall_ind, testcases,

How it might be represented: pie chart, exercise cycle

Limitations:

- a) The same test in different CWIX exercises has a different test ID. This then fails to capture the possible improvements of a specific capability from changes in specific standard made by NATO over different exercises when Io shortfall is recognised.
- b) Currently, the two data points we have on our graph will not show a trend to make reliable conclusions. However, this metric becomes important when data over many CWIX exercises are collated.
- c) Note to NATO: There is a data quality management issue where the name row of "standards" (id 9) in "issue category" may not match the "io shortfall." Hence, we have proceeded with the metric given as above.

Possible error found in the data collected. Io shortfall when flagged are labelled as software issues under the interoperability issue when they are really meant to be standards issues.

5) Number of capabilities from each nation.

Narrative: This will allow us to show how much each nation contributes to the CWIX event. If certain nations dominate the CWIX this could mean that interoperability is 'not baked in' and that the other nations have to interface with their systems.

Due to time restrictions, we have not been able to implement this to our dashboard.

Data Attributes: Nation_id (part of nations), capability

6) Percent of Not-tested against Overall-result:

Narrative: If x amount of tests were not tested, it shows that the CWIX event itself was not organised sufficiently well enough where all planned tests could be performed.

Thereby delaying changes that could be made which has an impact on how quickly better interoperability can be achieved. It could also mean the tests brought forward by nations may be too long for the CWIX environment. Hence, this has implications on limiting the scope of tests brought forward.

Due to time restrictions, we have not been able to implement this to our dashboard.

Data Attributes needed: Test case: overall result, exercise cycle

How it might be represented: Points on a scatter graph over exercise cycles.

Level of innovation of proposed solution

We proposed the limitations and considerations that one has to give for each of the KPIs during consumption. We considered a range of possible solutions to come up with the simplest metrics that provide the most insight into the interoperability. We narrowed down the metrics by considering:

- 1) Considering the holistic way of how a key performance metric can incentivise certain behaviours and matching it to NATO goal of improved interoperability.
- 2) From the high level stakeholders point of view, what is easily consumable.

With more time, we would use a systematic weighted survey filled out by engineers at CWIX over defined metrics of judgement to determine the importance of various metrics as a filtering mechanism to get main KPIs.

We used the latest analytics software as the problem was centred around KPI as opposed to designing an analytics platform.

Assessment of viability of solution

We used PowerBI (a mature low code analytics platform) which allowed the focus on the creation of the KPIs and how to present them to key stakeholders. As opposed to spending time on coding a dashboard and analysing data from a lower level platform/coding base such as SQL. This meant our solution was extremely viable. A data model was produced which means that data engineers at NATO can quickly reference where each data field is derived from and their relationship with other tables.

Plan for demonstration of solution

Introduce the team. Show the powerpoint slide. Explain the problem statement. Explain the Critical Success Factors. Explain why the main KPI was chosen and their narratives. Explain how it was chosen. (Through the design thinking process.) Explain the limitation to each KPI. Show them the dashboard and the drill through feature.

Future KPIs recommendations

1) Number of Service Desk calls received on CWIX applications from other nations

Narrative: If this decreases then we know that interoperability is getting better

Limitation: This is questionable as there could be other issues that affect services but it does show a trend of overall usage

2) Previous capability that failed due to standard issue needs to be captured and stored with their unique ID and carried over to the next exercise cycle.

Narrative: When standards are eventually changed, that same test could be used to validate that the change of standard was successful. This is a very important measure for increased interoperability.

Data Attributes: Exercise date linked to interoperability issues linked to standards.

3) Number of Capabilities used by one or more countries.

Narrative: This is a measure of utility of the capabilities by the countries

Why it cannot be done at present: This data is not collected by NATO from partner nations

Data attributes used: Capabilities, Nations,

4) Number of capabilities tested by one or nations

Narrative: This is an indicator of the number of nations that have participated in the testing of the capabilities

Indicator to measure rate of adoption of capability in a nation.

5) Distribution of the number of capabilities that achieved each stage of maturity throughout the CWIX years.

Useful information to collect at future exercises

- 1) Number of service desk calls
- 2) Date and time
- 3) Has the standard been updated, has the test been rerun as a result?
- 4) When has the standard been updated in relation to the test
- 5) How long a test takes (poorly planned test/uncompleted tests leading to slower implementation of interoperability.)
- 6) Planned time for test vs actual time for test

Recommendations

Service Desk data: To better understand the trends of implementation in the nations using the capabilities it would be good to request *service desk call information* this would be helpful in two ways:

- A. To understand which nation domains are consuming the capabilities (if the data allows us to understand the sub-organisations within the nation defence organisation)
- B. To understand if the nation concerned are receiving inquiries from other nations in NATO about their use of the capability

Feedback: There is limited time to complete all tests at CWIX.

However, capturing qualitative data by semi-structured interview questions will allow greater insight into the problem of coming up with useful metrics to measure interoperability. In order not to lose the momentum and ideas it may be a suggestion to carry out surveys by random sampling a cross section of participants at a post discussion forum incentivised by reward or recognition.

Time of test: While we have data on the failures of tests and the issues involved, it would be good to have a measure which compares the *planned time of the test* vs the *actual time of the test*. While it is understood that each test will vary due to the number of issues and variables involved, it would be good to quantify the variation between planned and expected as this would provide the organisers of CWIX with valuable information. The information if interpreted correctly may indicate how effectively the nations/organisers are planning the

tests against the time available. This may mean that future tests run at the CWIX amend the scope of the test or change the governance/organisation of the tests to increase the throughput of the tests at CWIX. By increasing throughput of testing there is more chance of interoperability increasing as the more testing is done, more capabilities may be introduced into operations.

Appendix







