Fraud Data Capability Assessment – v1.0

User Guide (Accessible Version)

Copyright Disclaimer

Introduction

As we move towards a more digital society, economy and government, our traditional responses to countering fraud will less effective. Furthermore, with organised and opportunistic fraudsters targeting government programs every day, the need to share and make better use of data and information across government and the private sector has become imperative for combatting fraud. Digital transformation will create new opportunities to use data to combat fraud, and by investing effort and resources to increase our fraud data capability now, entities can better support the Australian Government's Digital Government Strategy through the design and delivery of more secure, more sustainable and more effective services.

An effective fraud operating model is built on a foundation of fraud data capability that operates together to identify, investigate, treat and prevent fraudulent activity. A data driven fraud operating model enables entities to examine more information and identify hidden patterns much faster than humans can. This can help entities proactively identify irregularities or indicators of fraud, which can be used to better target resources to investigate cases with the highest risk of fraud. Data analytics can also help entities prevent or disrupt fraud earlier and reduce the need for long and expensive investigations.

The Commonwealth Fraud Prevention Centre has created the Fraud Data Capability Assessment (the Assessment) to help Australian Government entities:

- identify their fraud data analytics strengths and areas for improvement
- enhance their fraud data analytics capability by identifying activities and training opportunities
- develop business cases for investment in their fraud data capability
- understand and evaluate the benefits of improving their capability to find and prevent fraud

What is the Fraud Data Capability Assessment?

The Assessment aligns to the Australian Public Service Commission's Data Capability Framework. This Framework contains 26 data-specific capability areas related to using data in the Australian Public Service and is designed to work alongside existing frameworks, including the Integrated Leadership System.

The Assessment includes a series of questions designed to measure and identify your entity's current fraud data analytics proficiency against each of the capabilities in the Data Capability Framework. Each capability contains three proficiency levels of foundation, intermediate and advanced.

The questions have been divided into eight categories, aligning with the Centre's Fraud Data Analytics Framework:

The questions have been divided into eight categories, aligning with the Centre's Fraud Data Analytics Framework:

- Risk Assessment/Requirements
- Data Acquisition
- Analysis
- Visualisation and Consumption
- Feedback Loop and Improvement
- People
- Technology
- Data Management and Governance

Who is the Assessment for?

The Assessment has been designed to be applied at the organisational level for Australian Government entities looking to identify their fraud data capability strengths and areas for improvement.

Please note, the Assessment is not designed to determine an individual's fraud data capability.

How to complete the Assessment

Who needs to be involved?

How to undertake the Assessment is completely up to your entity, and will be influenced by how your fraud data analytics function is structured. We recommend your entity complete the Assessment in a workshop environment where all relevant stakeholders were present to allow the answers to be discussed/debated, and help to reduce the potential for unconscious bias to impact the results of the Assessment.

The Centre is more than happy to assist throughout the Assessment – please contact us to discuss options.

Completing the Assessment tool

The Assessment consists of 43 multiple choice questions in two formats:

Tick all that apply – select all answers that apply to your entity. Select your answer/s by clicking the relevant check box. A second click will de-select the box. If you select 'none of the above', you will be unable to select multiple responses. These will appear as square radio buttons that are able to be selected.

One selection only – these answers will build on each other. I.e. it is assumed your entity also meets the requirements for the preceding answer options. Select your answer by clicking the relevant radio button. These will appear as round radio buttons that are able to be selected.

The Assessment also contains text fields. To insert your answer, click the text field and begin typing. These will appear as blue free text fields where you are able to input required information.

To ensure that the Assessment is as accurate as possible:

- Include all impacted areas/teams (depending on the size of your entity, it is possible that some data functions are not located within the counter fraud team, or are completed at an organisational level, e.g. data governance, IT).
- Answer the survey honestly and to the best of your ability.
- Refer to the Glossary of Terms at the end of this document to confirm understanding of terminology used throughout the Assessment.

For optimal results, we recommend completing the survey in a workshop environment where all relevant stakeholders are present. This will allow the answers to be discussed/debated, and help to reduce the potential for unconscious bias to impact the results of the survey.
The Assessment Tool allows answers to be changed if required prior to submission:

At the bottom of each page, there is a 'Clear page' button

This will clear the answers on this page only, allowing the answers to be re-entered.

Submitting your Assessment to the Centre

At the end of the Assessment Tool, there is a 'Submit via email' button.

Clicking this button will submit your completed Assessment to the Centre via email. Instructions are provided below:

Step 1.
Once you hit the submit button two sending options will appear. The first option will be 'Default email application (Microsoft Outlook)', the second will be 'use webmail'. You will select 'Default email application (Microsoft Outlook)'.

Step 2.
This will open a new email on Microsoft Outlook with pre-filled information. This prefilled information will be the email address it will be sent to (The Centre), the subject and the completed survey attached. Click 'Send' to email your survey. *If this does not work, please save the survey and manually email it to info@counterfraud.gov.au.

Once you have submitted your completed Assessment and submitted it to the Centre, we will process the results. The Centre will provide a report to your entity outlining your proficiency against the Data Capability Framework's 26 capabilities.

Glossary of terms

Artificial intelligence – mimicking or simulating human intelligence and actions in machines.

Bayesian statistics – Statistics calculated using Bayes theory of probability.

Business intelligence – use of strategies and technology to convert data into information to allow for data-driven decision making.

Control – individual measures, processes or functions that help entities prevent, detect and respond to fraud. An integrated assembly of controls make up a control environment.

Classification – A way to group a set of related categories in a meaningful, systematic, and standard format, e.g., country or region.

Clustering algorithms – Clustering refers to the process of dividing data points into groups (clusters), so that each data point in a cluster has more in common with their cluster than with other clusters.

Data analysis – the process of examining data for the purposes of extracting insights and supporting decision making. It is a necessary sub-component of the data analytics discipline.

Data analytics – Examining raw datasets in order to draw conclusions about the information they contain

Data collection – The process of gathering data.

Data editing – Correcting identified errors to improve the quality of the data.

Data governance – A collection of practices and processes which help to ensure the formal management of data assets within an organisation.

Data management systems – Systems to store, collect and analyse data.

Data matching – Comparison of two or more data points to identify inconsistencies or confirm compliance.

Data products – A tool or technique that processes data and produces results.

Decision trees – algorithms that map out logical decisions and their possible consequences/outcomes, creating a branching structure.

Descriptive statistics – The use of basic statistical calculations (such as mean, median, standard deviation) as well as visualisations (such as histograms) to provide insights around datasets and their distribution.

Enterprise Fraud Risk Assessment – Fraud Risk Assessments that look at the entity as a whole and how susceptible it is to fraud. This is the most general level of fraud risk assessment.

Entity – a department of state, a parliamentary department, a listed entity or a body corporate established by a law of the Commonwealth.

External data – Data acquired through the public domain, through agreement with the private sector, or licensed from specialist data vendors.

Fraud – dishonestly obtaining a benefit or causing a loss by deception or other means.

Fuzzy matching – Also known as approximate string matching, this technique attempts to connect text elements that are similar but not exactly the same.

Information management principles – The gathering data and then analysing, categorising, contextualising, and archiving (and in some cases, deleting) it, in order to support a business's needs.

Inter-agency data – Data shared between agencies to facilitate better fraud detection.

Internal data – Data generated internally within the agency's systems. Depending on that nature of fraud risk to be analysed, this can be sourced from many different types of systems including data from finance, HR, email, IT, and customer systems.

Organisational data – Data that results from the operation of administrative systems (e.g. data collected by government entities for registration, transactions and record keeping reasons).

Output – Analytical outputs include graphs, reports or infographics with analytical information.

Linked/integrated data – combining data from different sources, creating a unified view

Natural language processing – Algorithms that have been developed to examine and convert text data (unstructured) to structured data, to facilitate further analysis.

Network analytics – Establishes relationships between entities based on common data attributes to create a network of activity.

Neutral networks – algorithms that create a series of decision layers (nodes) that "weight" their inputs and "fire" if the result is above a threshold.

Machine learning – the use of computers to learn from data, without explicitly coding in rules of analysis

Metadata – Data that provides information about other data.

Processing Methodology – Statistical procedures used to deal with intermediate data and statistical outputs, e.g., weighting schemes, statistical adjustment, or methods for imputing missing values or source data.

Regression analysis – algorithms that attempt to understand if there is a relationship between a dependent variable (the data point that is being analysed) to independent variables (data points that are of unknown impact).

Support vector machines – The algorithmic equivalent of drawing a line (or plane) through data to separate it into different categories by identifying the plane which is the furthest from the datapoints in the training dataset.

Thematic Fraud Risk Assessment – Fraud Risk Assessments that look at major functions or activities (often identified by the enterprise-level risk assessment).

Time series Forecasting – Use of statistical methods to predict future behaviour based on historical data.

Trend analysis – Analysis of a series of data to identify patterns and abnormalities. In most cases, trend analysis is conducted over time-series data.