OECD Asia-Pacific Academy for Tax and Financial Crime Investigation

Managing Financial Investigations

OECD

OECD ASIA-PACIFIC ACADEMY FOR TAX AND FINANCIAL CRIME INVESTIGATION

Big Data Analytics and Financial Investigative

Kong Yew Hon
May 2024

INLAND REVENUE AUTHORITY OF SINGAPORE

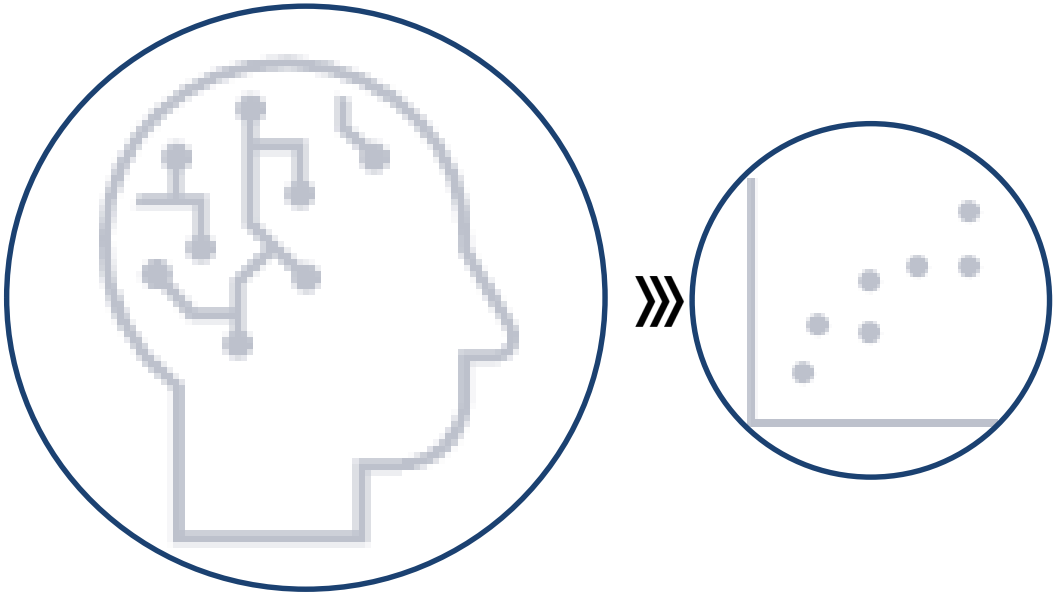
Learning Objectives

At the end of the course, you should be able to understand and appreciate:

(1) Overview of IRAS AI* Strategy

(2) All About Data

- ✓The different sources of information;
- ✓Data governance and limitation of certain data;
- ✓Effective management of information;
- ✓Application of concept : Case Studies;
- ✓Group sharing



(3) Big Data Analytics

- ✓Leveraging on technology & analytics in financial investigation
  - o AI as a gatekeeper;
  - o AI for Tax Crime;
  - o AI for investigative efficiency;
- ✓Application of concept : Case Studies;
- ✓Group sharing
- ✓What's next

** AI = Artificial Intelligence*

Survey

www.menti.com

Access via QR Code or Use Code : 77264430

(1) Overview of IRAS AI Strategy

AI as a Strategic Capability

AI is identified as a strategic capability that enables IRAS to pursue desired outcomes in the organisation

- The objectives of AI are aligned with those under the IRAS Leveraging Analytics, Design and Digitalisation (LEA:D) transformation movement
  - To improve the **efficiency** and **effectiveness** of tax administration and enterprise grant disbursements
  - To deliver **anticipatory** and **integrated** services to taxpayers
  - To use **rich, entity-centric data** to develop intelligent applications
  - To build organizational and individual **capabilities** to leverage big data and AI

AI as a Strategic Capability

LEA:D Strategic Priorities	How to achieve it	What this means for IRAS' Data?
Anticipate Needs, Co-Create & Customise Solutions	Customise information, service delivery & compliance treatment Collaborate & Co-create with community by default	• Right Data for the Right Subject at the Right Time • Common, consistent definition and understanding of Data to facilitate collaborations and co-creations
Connect Digitally	Be 100% digital Integrate tax seamlessly Build smart & agile IT systems	• Increased opportunities to capture all data used in our work • Appropriate data procurement process in place to ensure data captured can be integrated to our existing data pool and capable of supporting future analysis and operations.
Use Data Intelligently	Gain insights for smarter decision making Think data first Embed analytics in processes & systems	• Data must be accurate, timely and inter-operable across tax types, systems, and processes • Required data must be accessible with the appropriate safeguards to maintain public confidence in our digital push
Build an Adaptable & High Performing Workforce	Build capabilities to excel in a digital workplace Inculcate culture of innovation & experimentation Collaborate in cross-functional teams	• Staff have ready access to resources to help them learn and understand the data available e.g meta-data, data dictionaries, ETL rules • Staff is able to explore and use data freely in a secure environment. • Staff is aware of its duty in ensuring data confidentiality

IRAS AI Strategy

In FY2019, the **IRAS AI Strategy** was established and approved. It identified the strategic use-cases and roadmap to extend our analytics capabilities and to scale up the use of AI in IRAS.

Identify and prioritise Use Cases
(Problems that demand AI solutions)

Deliver Quick Wins

Develop AI Capabilities
(Structure, Skills, Technology/Tools, Data)

- Identify quick-wins and medium-term (or "invest") use cases to kickstart AI roadmap development
- Build and scale AI capabilities for prioritised and future use cases

Scale (FY20 – FY23)

- Identify other use cases and prepare for development
- Scale up use cases, including those identified above, and use of AI in IRAS

Key Pillars of IRAS AI Strategy

The **IRAS AI Strategy** has 4 key pillars:

Applications

Deliver high impact AI use-cases in an agile and continuous manner

Data

Acquire and utilize data effectively to meet AI needs

Infrastructure / Tools

Provide tools and acquire up-to-date AI capabilities in the long-term

People / Organisation

Develop central expertise and in-house AI capabilities

Getting the Investigation & Forensic Division (IFD) ready for the future

*IFD's Data Journey**

Before 2018 : Early Adoption of Analytics in IFD

- ❑ IRIN system containing tax data and key data obtained from other agencies
- ❑ Computer Forensics
- ❑ E-Discovery Tool
- ❑ Network analysis tool ("SNA")
- ❑ Digitising bank statements

2019 to 2022: Enhancing IFD's Analytics Capability

- ❑ Optical character recognition ("OCR")
- ❑ Forensics Lite
- ❑ Automated Data Extraction (standard request) / Ad hoc Data Extraction (non-standard request)
- ❑ Transaction Tracing Tool
- ❑ Mobile Evidence Analytics Tool

2023 onwards: Future of IFD's Analytics Capability

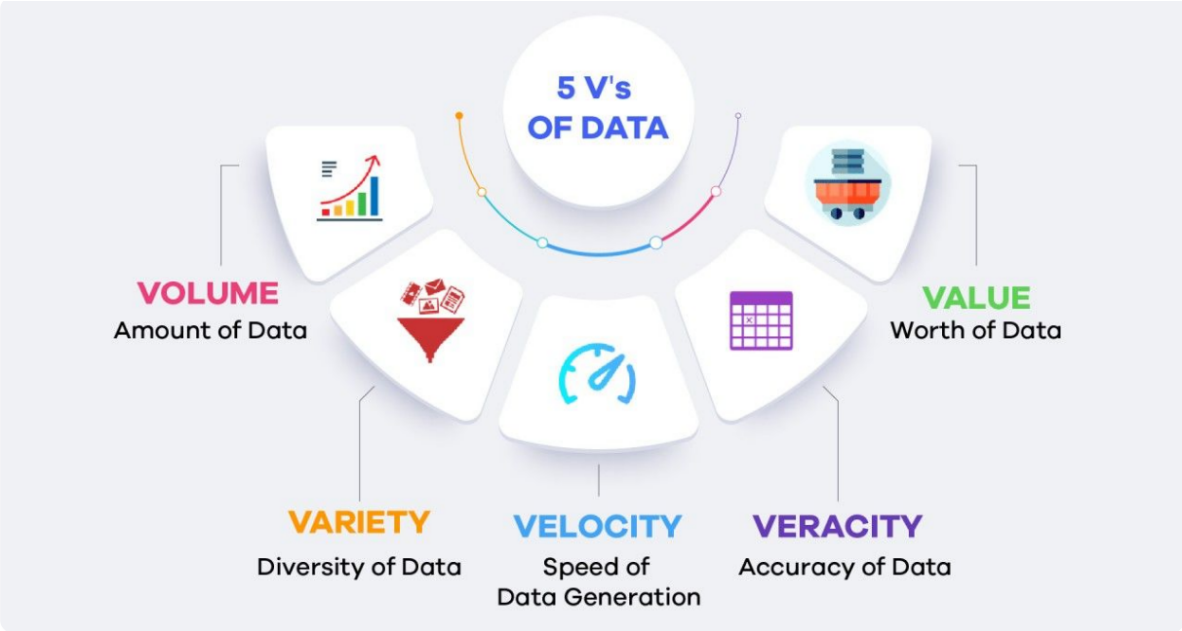
- ❑ Next generation network visualiser – Intelligent Network Analysis Tool ("iNAT")
- ❑ Cryptocurrencies

** Fully implemented for use by all officers including investigators, intel analysts, field officers and auditors*

(2) All About Data *[In the shoe of an Analyst / Investigator]*

5V’s of Big Data

- **Characteristics of Big data-** The five V's are volume, velocity, variety, veracity, and value*.



Video : https://youtu.be/bAyrObl7TYE?si=keT0GUC1UsAb5_D9

** Source : www.linkedin.com/pulse/big-data-explained-5v-excelsiorites*

The different sources of information

Sources of Data

Internal

Structured

Tax Returns e.g. Form B, GST F5

Application Forms e.g. GST registration application form (F1)

Property related – Tenancy / Ownership

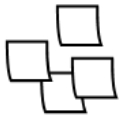
Tourist Refund Scheme

Unstructured

Audit / Investigation reports

Blacklist / Compliance Ratings

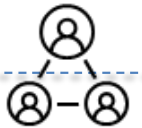
Correspondences with taxpayer



Intelligence

External

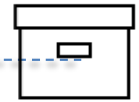
Semi-Structured



Other Government Agencies e.g. STRs[1]

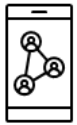
Other countries e.g. EOI[2]

3rd Parties – Banks / Telco / Employer - Auto Inclusion Scheme[3] etc



Devices / docs obtained from subjects

Open sources e.g. Internet



Social Media e.g. Facebook / Instagram etc

Informants

[1] *STRs = Suspicious Transaction Reports*

[2] *EOI = Exchange of Information arrangements with foreign jurisdictions*

[3] *Auto Inclusion Scheme ("AIS"): Under this scheme, employers submits the employment income information of their employees to IRAS electronically*

Sources of Data

Internal Structured Data	Internal Unstructured Data
Tends to be tax related data collected by IRAS from the various tax returns e.g. Income Tax returns - Form B / C, GST returns - GST Form 5 etc	Processed data after analysis performed on the data collected e.g. audit or investigation report / correspondences with the taxpayers
Most of IRAS' tax returns / applications are in electronic form where this facilitates data collection in structured format	Data are digitised and stored in IRAS Document Management System or shared folder
Data are stored in IRAS datawarehouses where it can be extracted for easy analysis and manipulation	

Sources of Data

External Structured Data	External Unstructured Data
Data provided by other government agencies e.g. Suspicious Transaction Report Office ("STRO") – STRs or Accounting and Corporate Regulatory Authority ("ACRA") – Singapore's equivalent of the Registrar of Business/ Companies	Subject's Facebook posts generally are collected on adhoc and need-to basis. IRAS also collect data from 3rd parties like banks for deposits / withdrawal data of our subjects during audit / investigation
Data from 3rd parties like employers under IRAS' Auto Inclusion Scheme – "AIS" e.g. Grab	Data would require considerable "cleaning" before use
Data are provided at a fixed intervals (e.g. annually for Central Provident Fund - "CPF"[1] data)	Users need to ascertain accuracy of data especially if sourced from the Internet due to the prevalence of fake news
	For investigation purpose, we have also employed OCR[2] technology to convert paper document seized from our subjects to digital format for ease of case analysis

[1] *Central Provident Fund or CPF is a mandatory social security savings scheme funded by contributions from employers and employees*

[2] *Optical character recognition or optical character reader I.e. OCR is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene photo or from subtitle text superimposed on an image (Source : Wikipedia)*

Internal Sources of Information

Tax Returns e.g. Form B, GST F5

Tax returns submitted by taxpayers. Examples of tax returns submitted to IRAS:

- Income Tax Form (Individuals / Partnership / Company)
- GST Form (Form 5 / Form 7 / Form 8)
- Withholding Tax

Application Forms – GST F1

Application forms submitted by taxpayers to apply for tax status or IRAS' schemes / incentives

- Income Tax Schemes / Incentives (Job Support Scheme etc)
- GST Registration application (Form 1) or scheme (Major Exporter Scheme etc)
- Overseas entities information (with effect from 1 Jan 2020 - under the Overseas Vendor Registration Regime – "OVR") *NEW*
- Withholding Tax

Property related – Tenancy / Ownership

Property related transactions

- Ownership of property (Sales / purchase price of property, name of buyer / seller etc)
- Tenancy related data (Amount of rental, period of tenancy, name of landlord / tenant etc)

Internal Sources of Information

Tourist Refund Scheme

Tourist / retailers for the Tourist Refund Scheme ("eTRS")
- Tourist spending information
- Retailer sales figures to tourist

Audit / Investigation reports

Processed data in the forms of reports
- Audit / investigation findings of taxpayer
- Provides insights to the taxpayer's attitude towards tax

Blacklist / Compliance Ratings

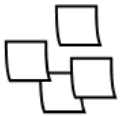
Internal assessment of the taxpayer propensity to commit fraud (blacklist) / non-compliance (compliance rating)

Internal Sources of Information

Correspondences with taxpayer

Past correspondences / enquires from the taxpayers

- Email
- Phone call (Telememo)
- Letter



Intelligence

Information concerning the subject of investigation obtained through intelligence channel

- Source can be from internal (IRAS officers) / external parties (e.g. Informant)
- Relationship linkage of subject
- Location of subject
- Modus operandi

External Sources of Data

Other Government Agencies

Other government agencies. Some examples:

- Business related data (E.g. Name of company director / shareholders / company registered address)
- Housing data (E.g. Public housing ownership information)
- Vehicle ownership data (E.g. Name of vehicle owner etc)
- Licensees data (E.g. Singapore Food Agency – Hawker licensee information)
- CPF data (E.g. Details of taxpayer's employment record)

Other countries e.g. EOI

Other countries:

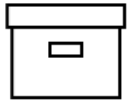
- Exchange of Information through Avoidance of Double Taxation Agreements (DTAs) & Multilateral Convention on Mutual Administrative Assistance in Tax Matters
- Financial Intelligence Unit ("FIU")

3rd Parties – Banks / Telco etc

3rd parties. Some examples:

- Financial (E.g. Banks / Shares / Insurance)
- Subscriber (E.g. Telco / Internet Service Provider)
- Utilities (E.g. SP Powers / PUB etc)

External Sources of Data

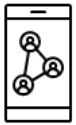


Devices / docs obtained from subjects

Obtained from our subjects:
- Physical document e.g. invoices, Purchase Orders etc
- Digital devices e.g. Cellphone, laptop etc

Open sources e.g. Internet

Open sources like:
- Internet
- News reports
- Research papers
- Published judgment in courts



Social Media e.g. Facebook / Blogs

Social media:
- Facebook
- X (former known as “Twitter”)
- Instagram

Informants

Whistleblowers on tax offence committed by subject

Typical Data Used for Syndicated Case Analysis

In-house data available to IRAS:

- IRIN: Various tax returns data from Income Tax / Goods and Services Tax / Property Tax etc, assets, business ownerships
- Parent / Child / Spouse / Employer/ Employee / Tax Agent relationships, Bank accounts, Contact details, Special Attention cases
- Stamp Duty System: Tenancy, Property mortgage, Shares transfer

Open Source Intelligence and Investigative Tools:

- Social Media like Facebook, X etc
- Internet searches

Obtained by IRAS using powers under our tax legislations:

- Bank statements, Account opening, signatory, phone number used for 2 Factor Authentication (2FA)
- CPF Employer/Employee Contributions,
- Utilities bills

Use of Data (Syndicated Case)

Pre-Audit / Investigation

Subject / Case Profiling

Analytics Models

Subject / Case Identification

Audit / Investigation

Background of Subject – Location / Income

Relationship of subject – Family members, known associates

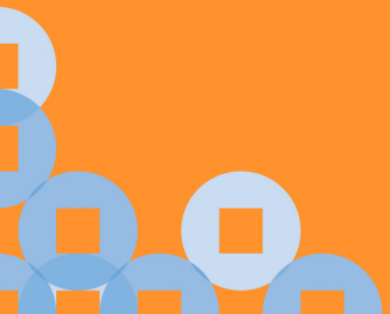
Unique identifiers of subject – Address / Email / IP address

Modus Operandi of fraud

Case Closure

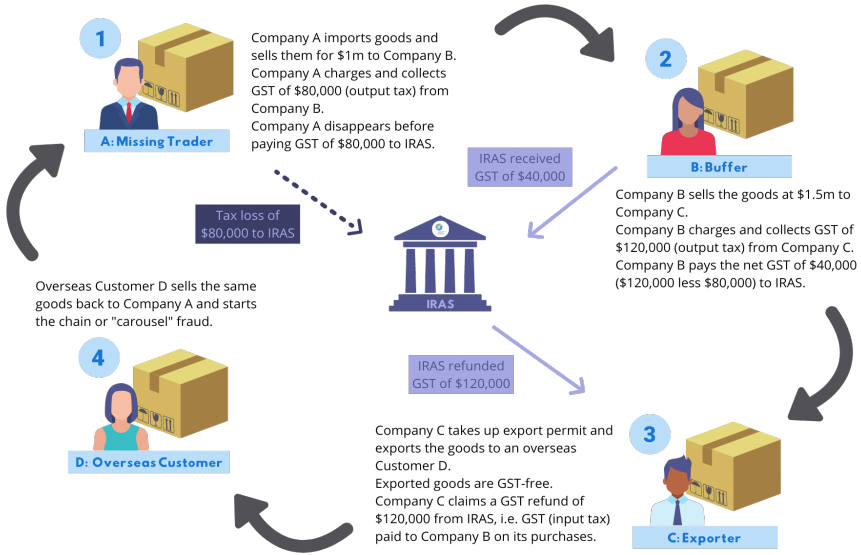
Audit / Investigation report - Feedback loop to refine Analytics Models / risk indicators

Case Studies



What is Missing Trader Fraud?

MTF* occurs when organised criminal groups abuse the GST/VAT refund system for fraudulent export arrangements and exploit the asymmetry of information in different jurisdictions to avoid being identified/traced.



** Missing Trader Fraud or MTF is also commonly known as Carousel Fraud in the other parts of the world like the European Union ("EU")*

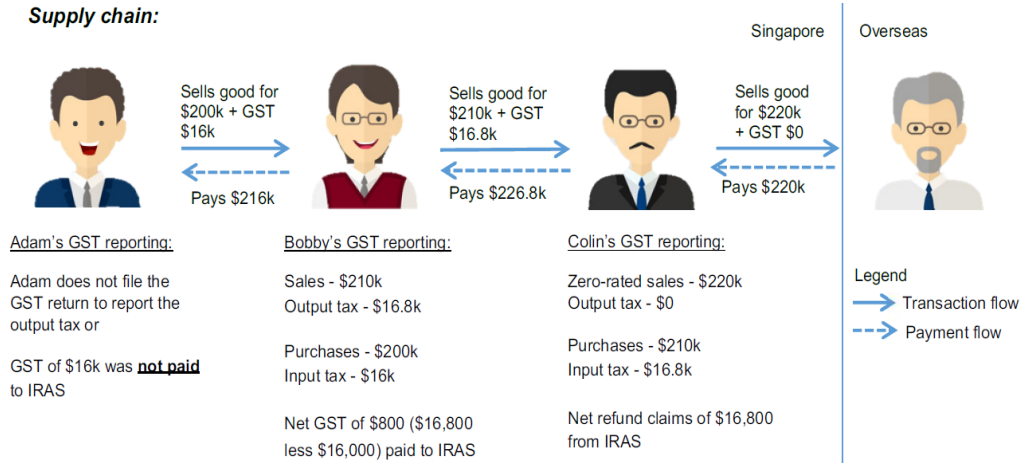
What is a Missing Trader Fraud (MTF):

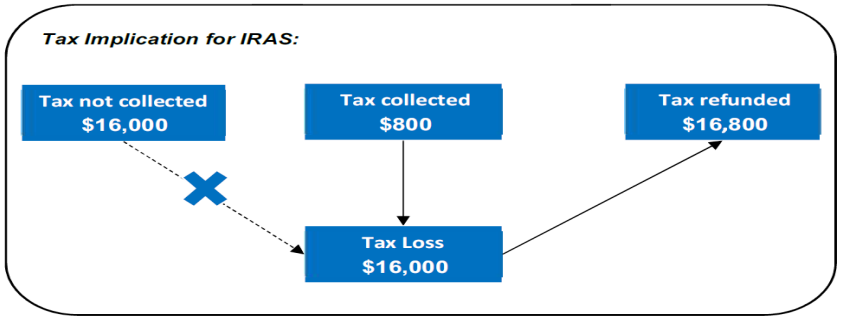
"Under a typical MTF arrangement, a group of businesses would form a supply chain and the same goods/services would be supplied through the chain. To ensure that the final sales of the goods/services are not subjected to GST, the goods/services would ultimately be exported to an overseas customer. A seller upstream in the supply chain would charge GST on the sale of goods to businesses downstream and instead of paying the GST to IRAS, the upstream supplier would ***fail to account*** *in its GST return the GST it had collected.*

This is termed ***"Missing Trader" fraud*** *as the seller disappears with the GST."*

What is Missing Trader Fraud?

Illustration of a MTF arrangement

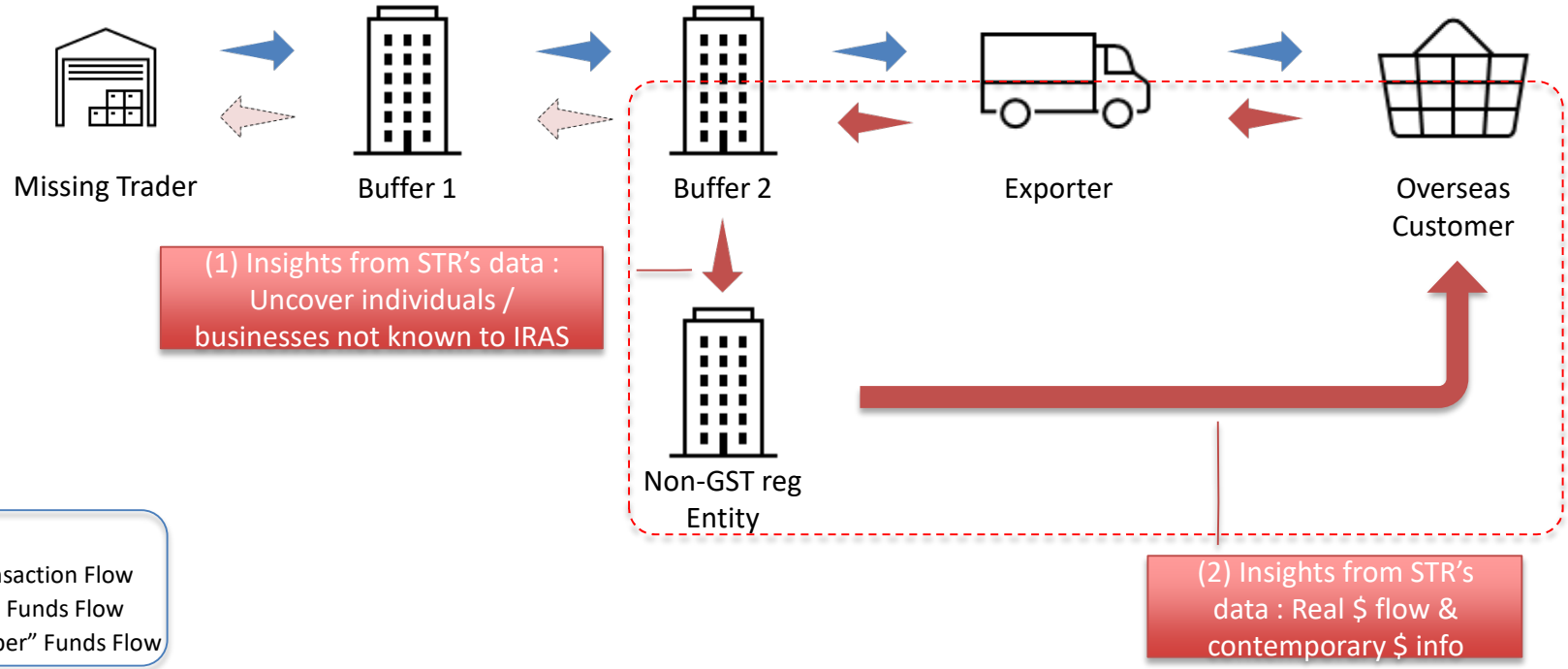




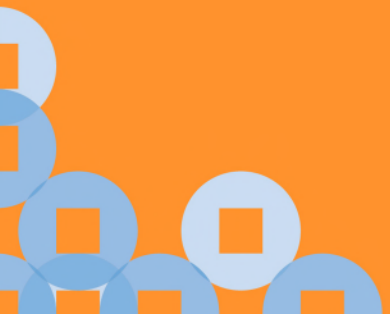
Question : Can share your agency's experience with MTF?

Operation Wand 2 – Missing Trader Fraud

Example : Use of Suspicious Transaction Report (“STR”) in Operation Wand 2

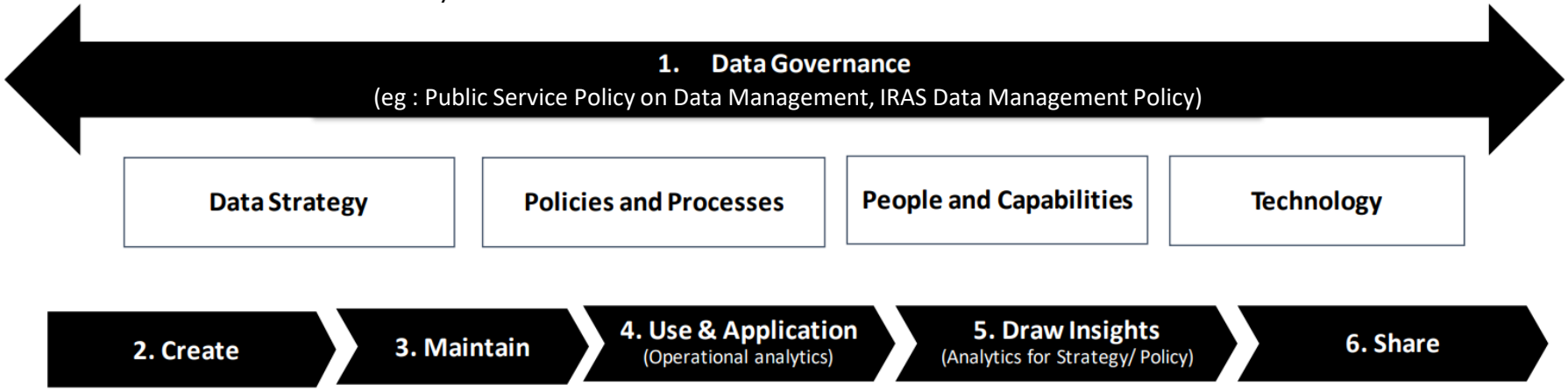


Data governance and limitations of data



Data Governance

- Overview of Data Lifecycle*



Data Governance is the exercise of decision making and authority for data-related matters.

The purpose of data governance is to ensure that data is managed properly, according to policies and best practices. This enables IRAS and the Government to make the best use of data for decision-making. It allows IRAS to achieve its goal of cultivating data- and insights-driven culture.

** Source : Data Steward Reference Materials*

Limitation of Data

Completeness : Data completeness describes whether the data you've collected reasonably covers the full scope of the question you're trying to answer, and if there are any gaps, missing values, or biases introduced that will impact your results*. Example, the data in the employment data table in IRAS system may not be complete as IRAS does not mandate the submission of employment data for businesses below certain staff strength.

*Source : https://www.montecarlodata.com/blog-what-is-data-completeness/

Relevance : Data relevance is the degree to which data provides insight into the real-world problem or purpose being addressed and contributes to the overall understanding of the business^. For example, IRAS conducted audit on subject from 2014 to 2015. Due to the long passage of time, the audit findings for the subject may not be relevant for the subsequent investigation in 2024.

^Source : www.metaplane.dev/blog/data-relevance-definition-examples

Timeliness : The degree to which data represent reality from the required point in time#. In IRAS' example, certain information are provided at fixed interval e.g. CPF data is provided to IRAS annually. Hence, to get the most up-to-date info, officer would need to write to the relevant authority for the latest information.

#Source : https://dsstream.com/introduction-to-data-quality-terms-definitions-examples-of-use/

Limitation of Data

Reliability and authenticity : Data reliability is the degree to which data, and the insights gleaned from it can be trusted and used for effective decision-making*. For example, certain data provided / obtained by IRAS will need to be assessed for its reliability like financial data of Entity A shared by whistleblower. IRAS would not know if the data provided by the whistleblower is authentic.

Source : www.thoughtspot.com/data-trends/analytics/data-reliability

Accuracy : Data is considered accurate if it describes the real world^. One example will be for data obtained from the Internet; it would need to be validated against other sources to confirm its accuracy.

^Source : www.metaplane.dev/blog/data-accuracy-definition-examples

Evaluation Matrix of Data / Information Provided to IRAS (e.g. Whistleblower)

TABLE 1 - SOURCE RELIABILITY

	RATING	DESCRIPTION
A	Reliable	No doubt about the source's authenticity, trustworthiness, or competency. History of complete reliability.
B	Usually reliable	Minor doubts. History of mostly valid information.
C	Fairly reliable	Doubts. Provided valid information in the past.
D	Not usually reliable	Significant doubts. Provided valid information in the past.
E	Unreliable	Lacks authenticity, trustworthiness, and competency. History of invalid information
F	Cannot be judged	Insufficient information to evaluate reliability. May or may not be reliable.

TABLE 2 - INFORMATION RELIABILITY

	RATING	DESCRIPTION
1	Confirmed	Logical, consistent with other relevant information, confirmed by independent sources.
2	Probably true	Logical, consistent with other relevant information, not confirmed.
3	Possibly true	Reasonably logical, agrees with some relevant information, not confirmed.
4	Doubtfully true	Not logical but possible, no other information on the subject, not confirmed.
5	Improbable	Not logical, contradicted by other relevant information.
6	Cannot be judged	The validity of the information can not be determined.

Question : Can you share how your agency deals with whistleblowers?

Effective Management of Information

IRAS Data Management Policy

Extract of IRAS Data Management Policy (dated 12 Apr 2019)

"Information is a vital organisation resource and it depends on the data assets that are being collected, created and managed. Data is a strategic asset that shall be managed securely, effectively and efficiently during its entire lifecycle to meet the business requirements of IRAS, to facilitate effective data sharing within the public sector and with the general public, to achieve quality policy formulation, catalyse innovations, facilitate analysis and research"

Data Governance : Set out the approving authority for data policies and role / responsibilities of data users

Authorities / Guidelines : Sets out the relevant legislations / guidelines / circular on data management

Assessment of Implementation Guidelines : Assessment of IRAS data management processes compliance with stated laws & guidelines

Information Management System – IFD

Component
- Hardware
- Software
- People

Process
- Data Collection
- Report Generation

Information System
- Report & Information

- Hardware (OCR machine to convert hardcopy document to digital document)
- Software (Data Digitisation Tool / Network Analysis Tool)

- Data are digitised for easy retrieval and analysis
- Data are shared via Collaboration Space
- Digital evidence are indexed and catalogued for easy search / review

- Investigation Paper (e-IP)
- Inhouse Case Management System
- Notes of Meeting from management meeting
- Legal opinion library for cases
- Others

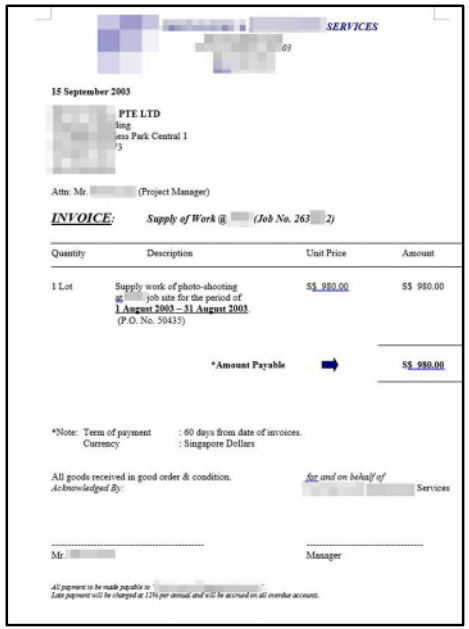
1 Access Control : Data is shared on need-to-know basis

Activity Logging: All activities are logged 2

Source : whatissixsigma.net

Information Management System – IFD

Some benefits of Data Digitisation Tool : Digitisation of hardcopy evidence



SERVICES

15 September 2003

PTE LTD

Attn: Mr. (Project Manager)

INVOICE: *Supply of Work @ (Job No. 263 2)*

Quantity	Description	Unit Price	Amount
1 Lot	Supply work of photo-shooting at job site for the period of 1 August 2003 – 31 August 2003 (P.O. No. 50435)	S$ 980.00	S$ 980.00

*Amount Payable S$ 980.00

*Note: Term of payment : 60 days from date of invoices.
Currency : Singapore Dollars

All goods received in good order & condition.
Acknowledged By:

for and on behalf of Services

Mr.

Manager

SERVICES

15 September 2004

Invoice No. 04/P 0915

MacPherson Road Post Office
PO Box
Singapo
Tel: 632
Fax: 63

Attn: Mr.

INVOICE: *Supply Labours and Tools to - Project*

Period: 01 September 2004 to 15 September 2004

1)	To supply workers to your above work site:	S$ 5,963.00

*Amount Payable S$ 5,963.00

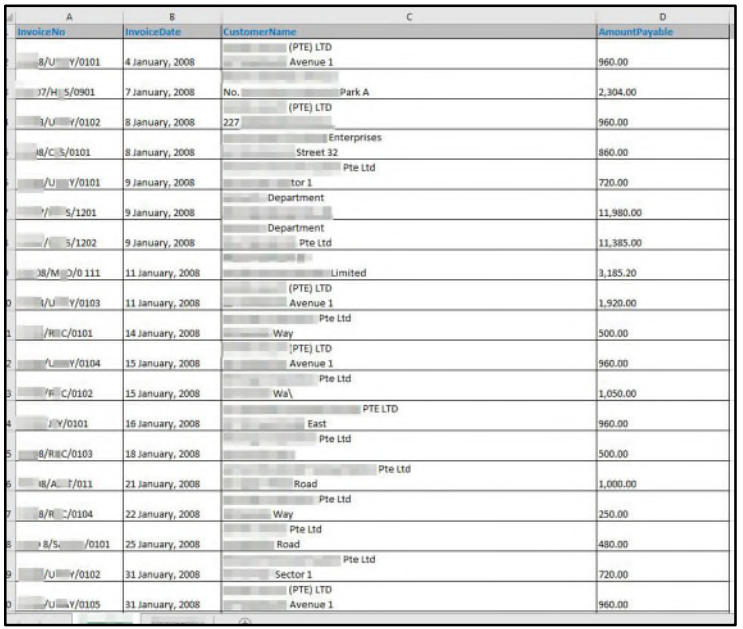
*Note: Term of payment : 30 days from the date of this invoice.
Currency : Singapore Dollars

All goods received in good order & condition.
Acknowledged By:

for and on behalf of Services

Mr. Yeo C T

Manager



A	B	C	D
InvoiceNo	InvoiceDate	CustomerName	AmountPayable
8/U Y/0101	4 January, 2008	(PTE) LTD Avenue 1	960.00
07/H S/0901	7 January, 2008	No. Park A	2,304.00
3/U Y/0102	8 January, 2008	(PTE) LTD 227	960.00
8/C S/0101	8 January, 2008	Enterprises Street 32	860.00
/U Y/0101	9 January, 2008	Pte Ltd tor 1	720.00
S/1201	9 January, 2008	Department	11,980.00
S/1202	9 January, 2008	Department Pte Ltd	11,385.00
8/M O/0 111	11 January, 2008	Limited	3,185.20
I/U Y/0103	11 January, 2008	(PTE) LTD Avenue 1	1,920.00
/W C/0101	14 January, 2008	Pte Ltd Way	500.00
/U Y/0104	15 January, 2008	(PTE) LTD Avenue 1	960.00
/6 C/0102	15 January, 2008	Pte Ltd Wa\	1,050.00
Y/0101	16 January, 2008	PTE LTD East	960.00
8/R C/0103	18 January, 2008	Pte Ltd	500.00
8/A /011	21 January, 2008	Pte Ltd Road	1,000.00
8/R /0104	22 January, 2008	Pte Ltd Way	250.00
8/S /0101	25 January, 2008	Pte Ltd Road	480.00
/U Y/0102	31 January, 2008	Pte Ltd Sector 1	720.00
/U Y/0105	31 January, 2008	(PTE) LTD Avenue 1	960.00

In the past, case officer had to manually key all these invoices into Microsoft Excel for analysis (can **take months** depending on volume of transactions)

Unstructured data is transformed to structured data for easy analysis (done in a matter of a **few days**)

Information Management System – IFD

IFD's Inhouse Case Management System (CMS) : Tracking of status and results of investigation cases

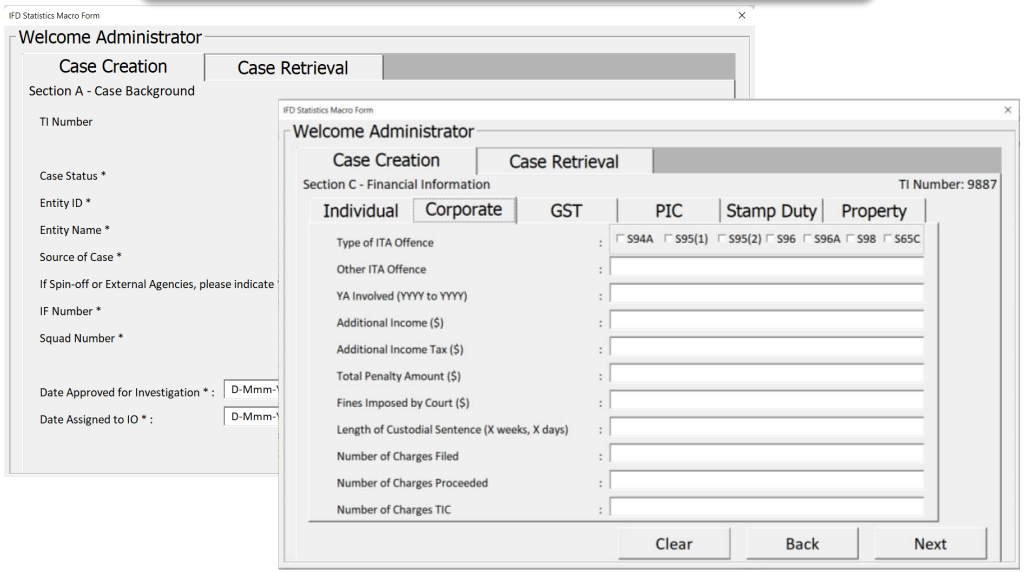
Tracking of Cases / Outcomes

Case Tracking File

Case Settlement Report

Others e.g. Performance Appraisal File

Case Management System (CMS)



Previously IOs need to update several files for reporting purpose. New CMS consolidates all reporting requirement in 1 platform - cut down the need for IOs to update different files

Country Sharing

Country Sharing (45 mins)

Any data that has been helpful to combat tax crime?

What data would you like to have to enhance your capability to combat tax crime?

- Group 1 : Internal Data (Structured)
- Group 2 : Internal Data (Unstructured)
- Group 3 : External Data (Structured)
- Group 4 : External Data (Unstructured)

(3) Big Data Analytics

What is Big Data Analytics?

“Big Data Analytics is the process of analyzing large and complex datasets using advanced technologies to uncover valuable insights and patterns. It enables organizations to make informed decisions, optimize processes, and gain a competitive edge by extracting meaningful information from massive and diverse data sources.”

Source : Big Data Analytics Overview (openai.com)

Quiz Time

www.kahoot.it

Game Pin : To Be Provided

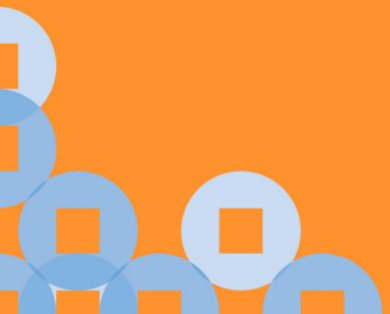
Country Sharing

Country Sharing (45 mins)

Any AI / analytics tool(s) that have been used by your country to combat tax crime?

- Group 1 : Enhance case detection (non-syndicated cases)
- Group 2 : Enhance case detection (syndicated cases)
- Group 3 : Improve productivity (non-syndicated cases)
- Group 4 : Improve productivity (syndicated cases)

Leveraging on technology & analytics in financial investigation



AI as a gatekeeper

Over the years, IRAS has deepened our AI capability where various AI tools were developed to detect crimes proactively

AI as a gatekeeper - Overview

Tax Fraud / Evasion

SMALL BUSINESS

Baseline Model (BLP) for Small Corporations

SMALL BUSINESS

Risk Profiling Model for self-employed

CORPORATE TAX

Baseline Model (BLP) for Corporations

CORPORATE AND SMALL BUSINESS

Anomaly Detection Model

ENTERPRISE GRANTS

One payout platform anti-gaming models

GOOD AND SERVICES TAX

GST Registration and Refund Social Network Analysis Model

GOOD AND SERVICES TAX

Embedding Analytics into GST refund processing

GOODS AND SERVICES TAX

Missing Trader Fraud Buffer Model

TAX CRIME

Missing Trader Fraud Syndicate Member Model

(In Progress)

TAX CRIME AND FINANCIAL INTELLIGENCE UNIT

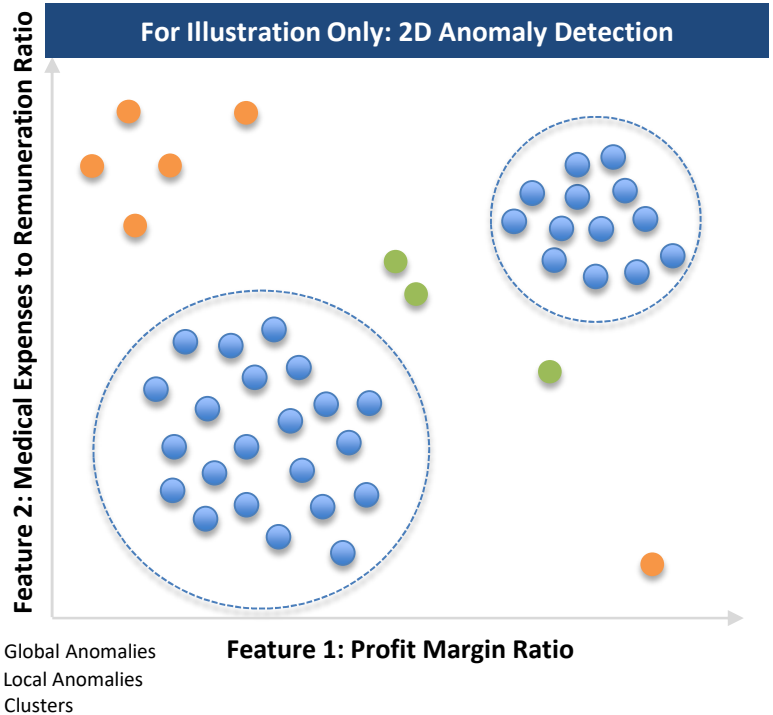
Missing Trader Fraud interagency collaboration

(In Progress)

Tax Non-Compliance

Anomaly Detection Models for Audit Case Selection

The Corporate Anomaly Detection Model aims to identify unknown risk areas and improve the current audit case selection process through the use of anomaly detection techniques



What Is an Anomaly?

An anomaly is a data point that is significantly different from the rest of the data points in the data. In IRAS, these would be companies that have different profiles from the rest of the population.

Benefits of Anomaly Detection

Anomaly detection can help identify companies with anomalous profiles over **one or more features.** This is a departure from the traditional approach where features are analysed independently from one another. This could uncover new insights or risk areas for IRAS.

Results

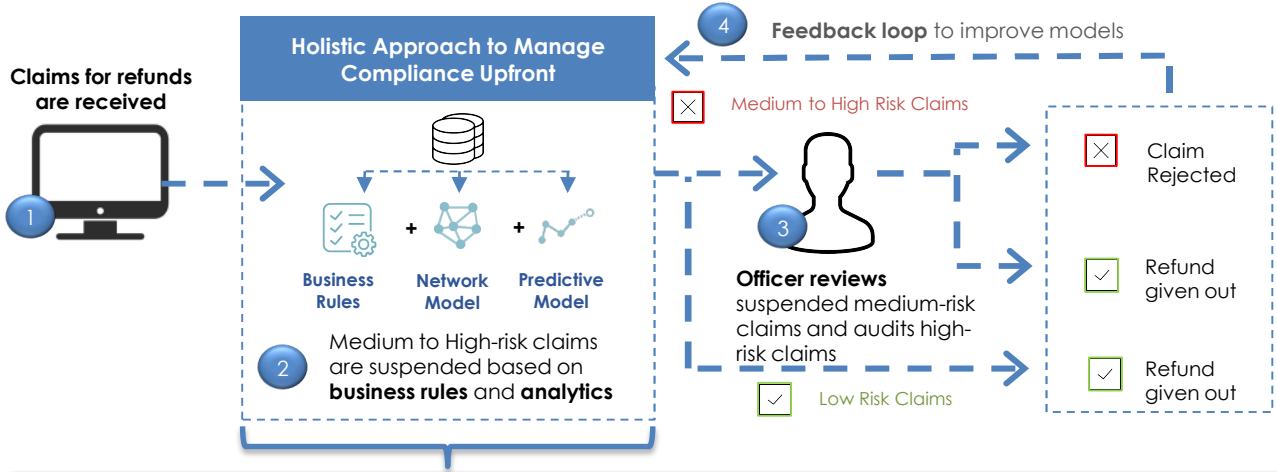
- Deployed in Mar 2021 for auditors' review
- Deployed the AI model and it identified a total of 92 anomalous Corporate Tax cases in the Year of Assessment ("YA") 2020 for auditors' review. As of Jul 2022, among the 44 closed cases, 10 cases have yielded a total recovery of $8.23 million

^ This includes a notional tax recovery of $7.67m and an actual tax recovery of $564k.

Embedding Analytics into GST Refund Process

Embedding of **2 analytics models** within the GST Refund Process to identify risky refunds:

- Pre-Refund Analytics (PRA) Model
- Social Network Analysis (SNA) Model



Returns are scored by a mixture of on-the-fly rules that score instantaneously and pre-computed scores

On-the fly scoring

- SNA transaction rules and PRA model score returns as they are received
- Historical data needed for scoring is aggregated beforehand

Ensures scoring accuracy

+

Pre-computed scores

- For SNA entity/network scores that are largely static
- Scores are refreshed in fixed intervals within SNA system

Minimises computing effort

=

$\sum$ *Analytics Scores*

Final Analytics scores

Missing Trader Fraud (“MTF”) Buffer Model

The MTF Buffer Model augments IRAS’ analytical capabilities with the use of an Auto Machine Learning solution to aid our fight against Missing Trader Fraud

- Enable IRAS to pre-emptively detect buffer entities engaging in MTF activity and **disrupt the supply chain early**
- **AutoML model** trained on known buffers and non-buffers
- Scoring is performed on GST returns to generate a score of **the likelihood that the return displays buffer-like attributes.**

Historical Data

Automatic Machine Learning

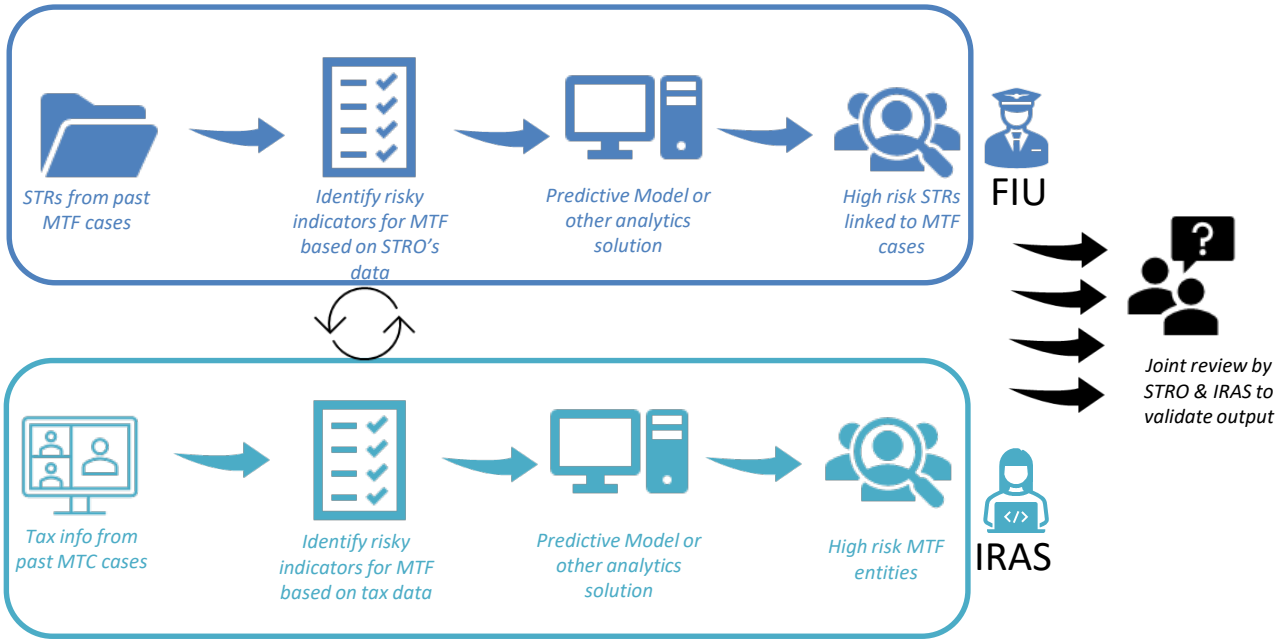
Risky GST returns

Results

Model was **3 times more effective** in detecting MTF buffer entities compared to traditional approaches, allowing IRAS to detect **>$30m of potential tax fraud**.

IRAS / FIU Interagency Analytics Collaboration

The IRAS-FIU interagency analytics collaboration adopts a federated data analytics approach to better identify Missing Trader Fraud ("MTF") in IRAS and STRO's respective systems



Progress to date

- Collaboration with our Financial Intelligence Unit ("FIU") from the Singapore Police Force to develop analytics models using past STRs filed / tax info from MTF entities to identify unknown MTF cases / clusters.
- Currently, project is at the model deployment stage where both agencies will analyse the cases generated to evaluate the model effectiveness.

AI for operational analysis

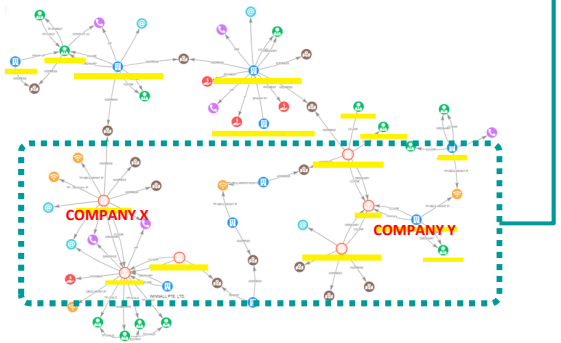
IRAS employs AI to amplify intelligence for better sense-making and targeting

Enterprise Network Visualising Tool - Intelligent Network Analysis Tool ("iNAT")

iNAT helps officers easily connect suspicious entities, even if they are linked by a complicated network of entity ownership, business transactions, and personal relations over multiple hops.

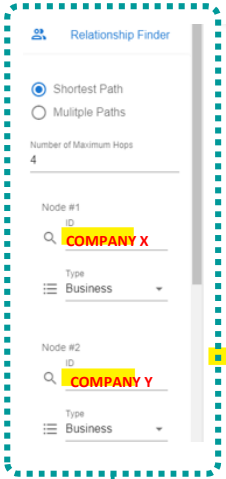
Example of iNAT applied to Missing Trader Fraud ("MTF") investigations

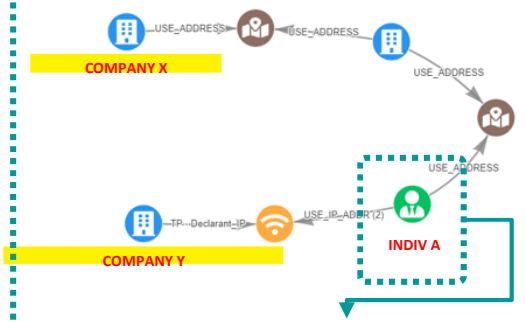
1 **X** and **Y** are missing traders in a Missing Trader Fraud scheme that involved **$114 Million in fictitious sales** and over **70 entities**.



Part of MTF network involving X and Y

2 With iNAT's Relationship Finder feature, IOs can access rich **relational databases** and visualize the links between X and Y





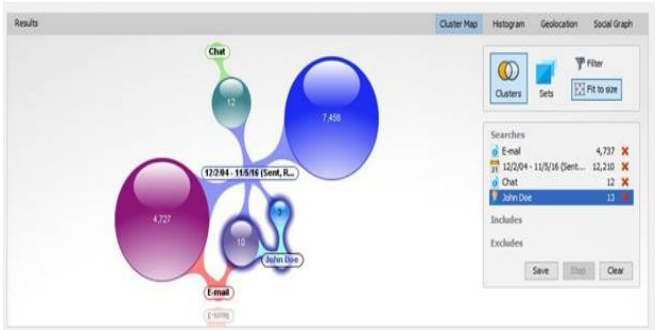
3 **A** is quickly uncovered as a suspicious individual who shares location and IP address with X and Y. **IOs can expand targeted investigations on A.**

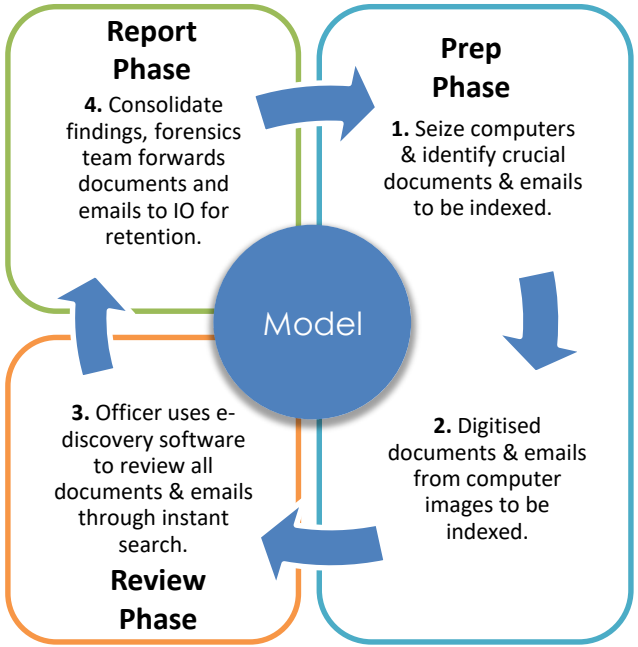
AI for investigative efficiency

Democratising AI to all officers for greater speed and effectiveness of casework during investigation

e-Discovery : Self-review Workstation

Investigation and eDiscovery software tool for single users who need to process, search, filter and produce amounts of digital evidence.



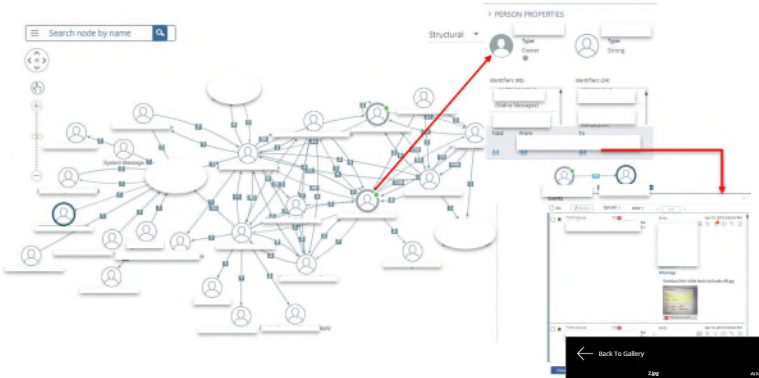


Benefits

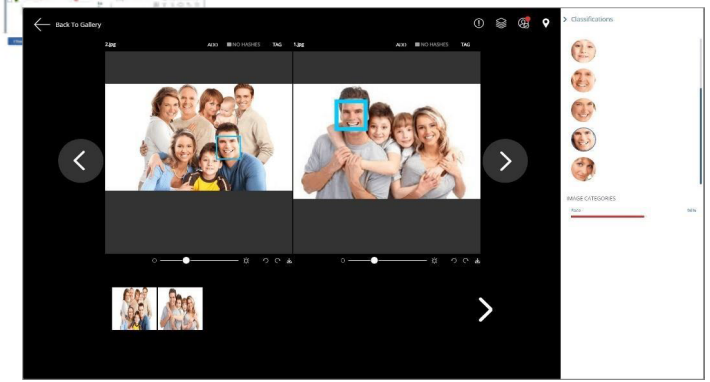
Users can quickly and easily process, search, review, and analyze digital evidence obtained. The easy-to-use interface and Cluster Map allows analysts to quickly find critical data, visualise relevant relationships, and drill down to the most pertinent data.

Mobile Evidence Analytics

Allows multiple mobile devices to be reviewed simultaneously and provides a contextual view of the relationships between devices



Relationship/Graph View



Face/Image Similarity Recognition

Benefits

- Enable multiple users to access at once to look at different facets of a syndicated case
- Other capabilities like facial recognition also allowed users to identify new subjects / relationships

Automated Data Extraction project

Improve productivity of users when reviewing case by automating some of the data extraction process from IRAS' various databases and charts / tables are provided for ease of analysis

Close to **300** data fields were extracted from various databases

*(> **150** data fields for business & > **100** data fields for individual)*

> **10 Key financial ratios** (e.g. Industry Gross Profit Margin) & network analysis are incorporated to supplement background check

Up to **30%** of First Level Report ("FLR" – report submitted by Intel Analyst) is pre-filled with requisite information from extracted data

> 10 charts and tables are created using POWER BI to facilitate user's review and analysis

Automated Data Extraction project

Key Benefits

Data from IRAS database is extracted automatically by running a code

Report template is pre-filled with information extracted via DW

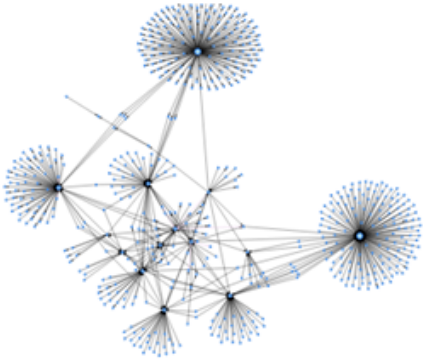
The report template is linked to POWER BI for data visualization

Productivity gained from manual to automated data extraction process (1-3 mandays to 1-2 hours)

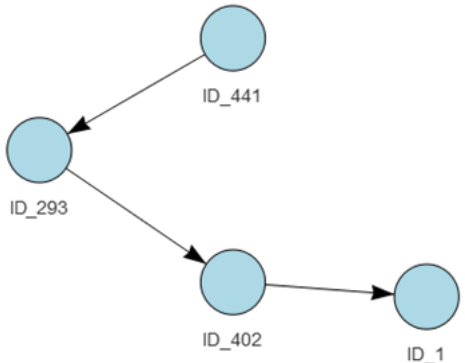
Transaction Tracing Tool (“TTT”)

Facilitate mapping of the supply chains for syndicated cases by providing officers with the entire flow of transactions based on a selected transaction

- Some Missing Trader Fraud ("MTF") clusters involved **large number of entities** (up to 50 entities per cluster) with **high transactional volume** (may be more than 10,000 transactions per cluster).
- After ingesting the transactional data, TTT will provide the supply chain by mapping out the predicted transaction flow for the users



Network Visualisation of transactions



Predicted Transaction Flow in Supply Chain

Transaction Tracing Tool (“TTT”)

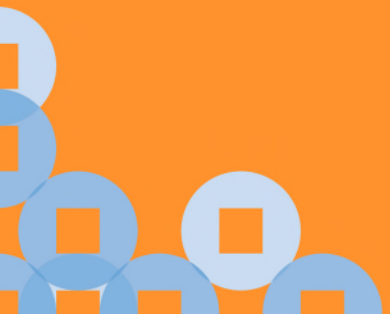
Key Benefits

Productivity gained by analysing large amount of transactions effectively – can quickly identify the fraudulent supply chains

- Saved effort of reading and matching invoice no & date, long text descriptions, quantity of goods
- Saved time from manually drawing supply chains

Officers can complete the tracing within 6.5 to 9.5 manhours by using TTT compared to 9.5 to 12.5 manhours when tracing transactions manually

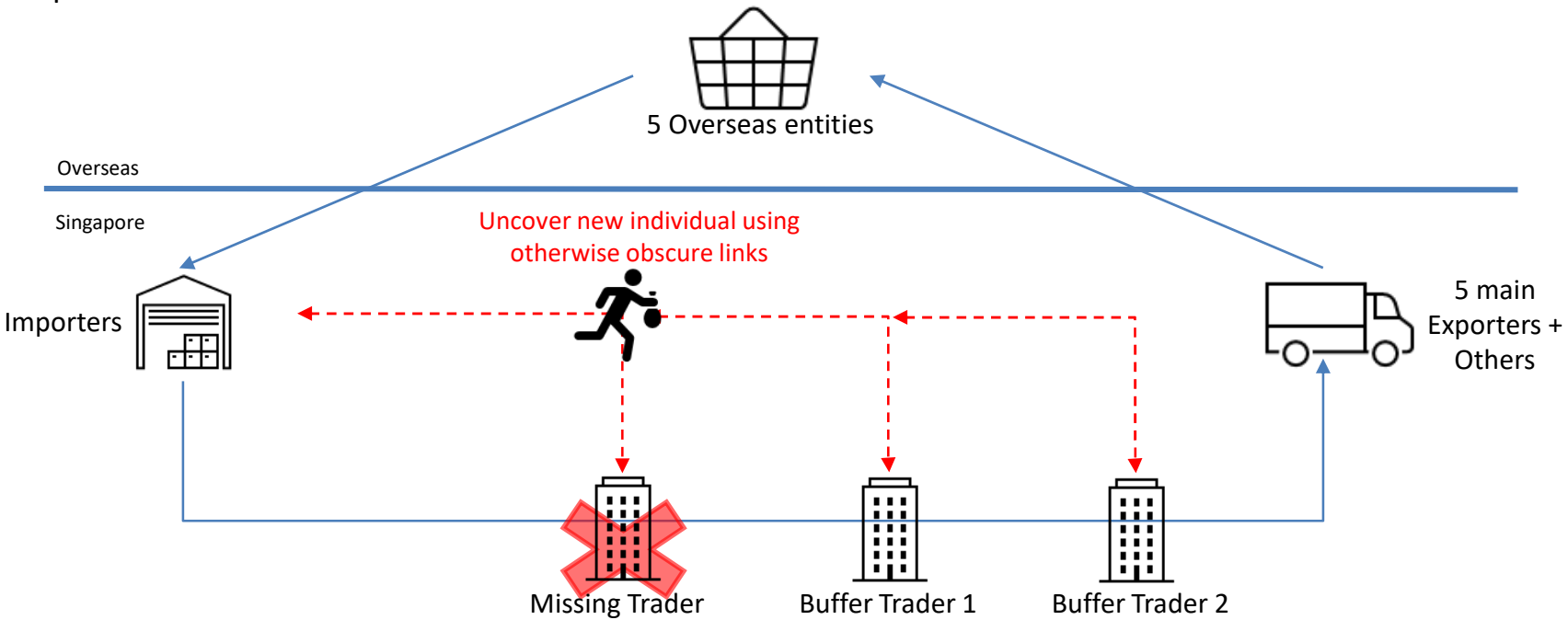
Case Studies



Case Sharing – Using Network Tool to Link Entities

Operation Wand

- Suspected GST Refund Fraud case
- More than 20 entities / individuals in cluster
- Complex network and transactions



Large Language Models (LLMs)

Advisory from the Prime Minister Office (“PMO”) on use of LLMs for Government Agencies

Accuracy & Accountability
Risks due to the tendency of LLMs to produce convincing yet inaccurate output (“hallucinations”) or provide inappropriate responses.

Security
Risks related to the potential compromise of Government data and networks. LLM providers may log data, often in overseas servers, which increases the risk of sensitive data being leaked if included in the user’s prompt.

IRAS' Experiment of Large Language Models (LLMs)

IRAS are currently limiting the exploration of LLMs in the domain of productivity aid and information retrieval while managing the risks involved. One potential use case for investigation is to use LLM to extract key names mentioned in statement / document.

Extract Key Names mentioned in the Statement

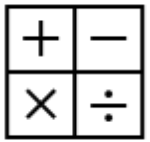
Entity Name	Email	Contact Number	Address
POI 1	d@ione2u.com	N/A	N/A
POI 2	@ione2u.com	N/A	N/A
COI A Pte Ltd	@ione2u.com	N/A	221 Address S159557
POI 3	N/A	N/A	N/A
COI B	N/A	N/A	N/A
POI 4	N/A	N/A	N/A
COI C	N/A	N/A	N/A
COI D	N/A	N/A	N/A
COI E	N/A	N/A	N/A
POI 5	.com.sg	N/A	N/A

- **Pair Chat** is a free, fast and secure version of ChatGPT, and is currently available and free to use for all public officers. It is currently powered by the same Large Language Model underlying ChatGPT.
- Pair Chat has been cleared by the Smart Nation and Digital Government Office (SNDGO) for public officer use (up to certain data classification).
- Depending on the prompt, the no. of extracted entities extracted differ.

New Development in LLM

- New development in LLM*

Snap a picture of a landmark while traveling and have a live conversation about what's interesting about it.



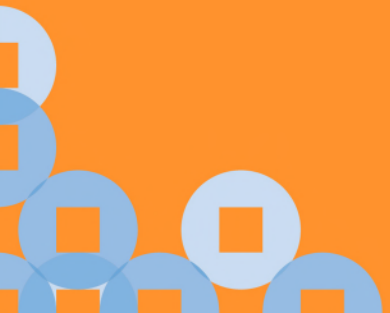
Help your child with a math problem by taking a photo, circling the problem set, and having it share hints with both of you.

*Source : ChatGPT can now see, hear, and speak (openai.com)

Hands-on Exercise for LLM

- https://openai.com/gpt-4

What’s next

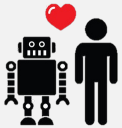


What is next in IRAS AI Strategy

Rich Entity-centric Data Portfolio

Enrich understanding of taxpayers using data/big data

AI for Everyone



Leverage AI/data science for intelligent tax administration/ enterprise disbursement

Future-Ready Data Technology with Optimal Connectivity

Strengthen capabilities to receive, store and use big and real-time data and scale AI solutions

Democratise Data Science and AI Capabilities

Drive pervasive data usage across all staff

Trusted Data Ecosystem

Use and share AI and data responsibly

New Data Sources

- To continue to expand our sources of data (e.g. InvoiceNow)
- To tap on new data to refine existing solutions (e.g. fraud detection using bank transaction data) or explore new solutions (e.g. sentiment analysis on audio data)

Expand AI Applications

- By identifying strategic use-cases with high impact and readiness, to increase the effectiveness and efficiency of our service delivery
- By enhancing and scaling existing AI applications

Modernise Data Architecture

- By migrating the UDP to the Cloud to exploit Cloud capabilities and continuously improving it to strengthen AI capabilities
- By exploring and identifying technologies required to enhance governance in the ML/AI development lifecycle

Develop Expertise

- By continuously investing in our people and relevant tools
- By exploring partnership opportunities with external stakeholders (e.g. short-term attachment/exchange)

AI Ethics & Governance

- By establishing and formalizing a set of principles and measures to inculcate the responsible adoption of AI in IRAS without stifling digital innovation

AI Safety Summit in UK (Nov 2023)

AI Safety Summit (early Nov 2023) at Bletchley Park, in Buckinghamshire

[Extract of BBC article : "Rishi Sunak: AI firms cannot 'mark their own homework'" dated 2 Nov 2023: "The Bletchley Declaration calls for global cooperation on tackling the risks, which include potential breaches to privacy and the displacement of jobs.

*Signed by 28 countries and the EU, it also says AI should be kept "**safe, in such a way as to be human-centric, trustworthy and responsible**".]*

Governance for AI

https://www.bbc.com/news/technology-67285315

Thank You

The information presented in the slides aims to provide a better general understanding of taxpayers' tax obligations and is not intended to comprehensively address all possible tax issues that may arise. This information is correct as at the date of presentation. While every effort has been made to ensure that this information is consistent with existing law and practice, should there be any changes, IRAS reserves the right to vary its position accordingly.