



Data Mining in the ATO

Warwick Graco
Director Operational Analytics
Office of the Chief Knowledge Officer
ATO



Outline

- ATO
- Change Program
- OCKO
- Roles and Responsibilities
- Some Challenges
- Career Prospects and Education



The logo consists of a vertical black line intersected by a horizontal black line. To the left of the intersection are three overlapping squares: a yellow one at the top, a red one on the left, and a blue one at the bottom. To the right of the vertical line, the letters 'ATO' are written in a blue, sans-serif font.

ATO

- The ATO is the major revenue collector for the Australian Federal Government raising over 90 percent of revenue
- It is responsible for raising revenue from a variety of sources including income tax, GST, superannuation, excise and duties, fringe benefits tax, company tax and agriculture levies



Change Program



Change Program

- Deliver new capabilities to move ATO into 21st Century with e-Commerce
- Cost is approximately \$0.5 Bn
- Core Capabilities include:
 - Case Selection
 - Case Management System
 - Customer Management System
 - Revenue Management System
 - Channel Management including Outward Bound



Additional Capabilities

- Evidence Management System
- Litigation Management System
- Intelligence Support System

These capabilities required to manage complex cases and issues



Office of the CKO



Office of the Chief Knowledge Officer

Information Management including

- Corporate Reporting
- Enterprise Data Warehouse

Content, Document and Records Management

Knowledge Management

Corporate Intelligence and Risk

Analytics and Operational Analytics



Analytics Staff

- Have approximately 30 miners and modellers employed. Most work in the Change Program
- There are other staff who are competent in statistics, econometrics etc
- A large number of employees can do data cube analysis and spreadsheet work
- Some competent in SQL



Roles and Responsibilities



Compliance Model

Attitude

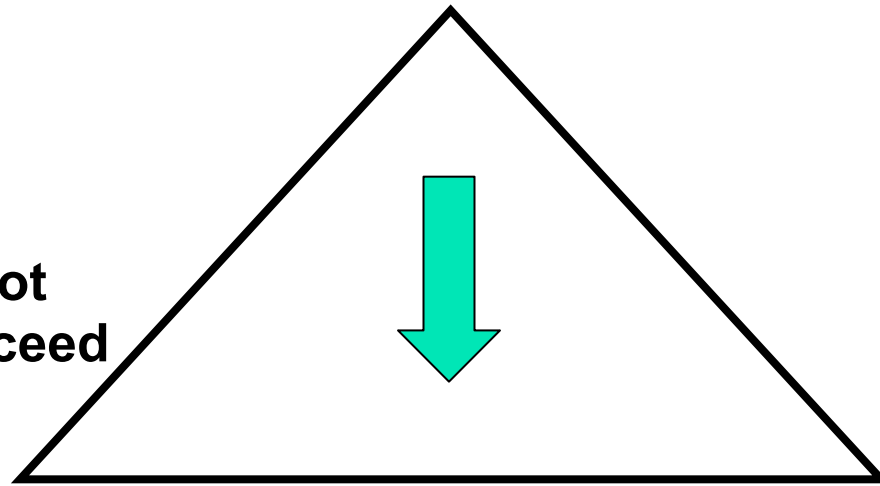
Not Comply

Don't Want
To Comply

Try but do not
Always Succeed

Comply

Push Down



Compliance Measures

Use full Force of Law

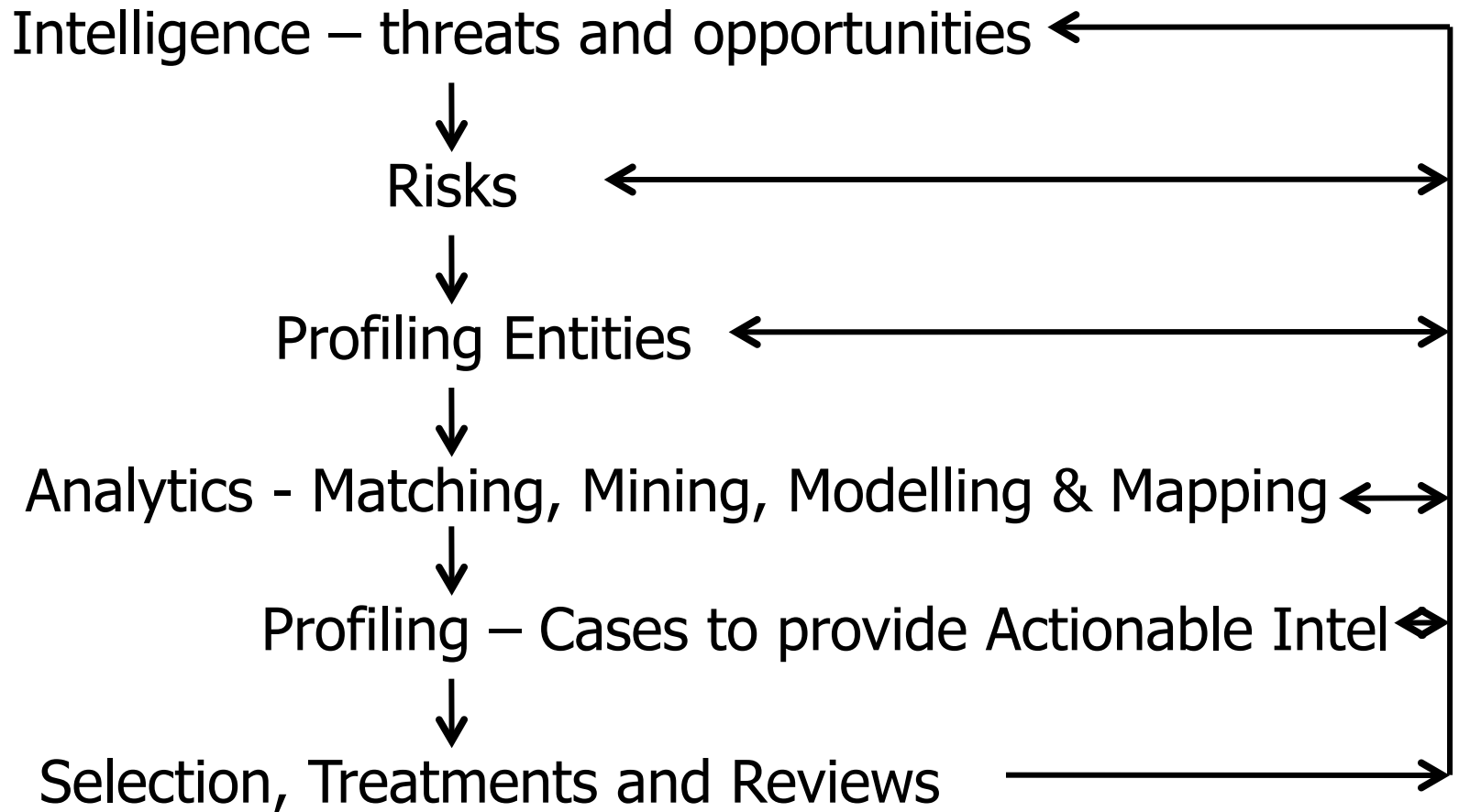
Deter

Assist to Comply – eg
Educate

Make it Easy



Analytical Cycle





Qualitative Disciplines

- Intelligence Analysis
 - Identify threats and opportunities
 - Determine their capabilities and intentions
- Risk Assessments – identify risks associated with each threat and opportunity and work out mitigation strategies
- Profiling – identify defining attributes and behaviours of entities of interest eg tax payers using tax-avoidance schemes



Analytics

- Those trained in Analytics perform the following functions:
 - **Matching** – ie link datasets and match data items
 - **Mining** – ie discover relationships, patterns and trends in datasets
 - **Modelling** – ie develop classification and prediction models
 - **Mapping** – ie identify the links and associations between entities such as people who live at the same address and make high-risk claims



Operational Analytics

- Those who perform this function have the following responsibilities including
 - Assist business owners to identify the risks they want models developed
 - Work out the business impacts of the models and the treatments they will apply to cases identified by the models
 - deploy models that meet required standards into production

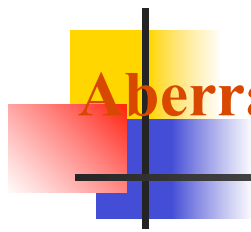


Analytical Models

- These produce a pool of high-risk cases based on the compliance risks identified by the business owner
- They have a long cycle and are changed periodically to keep current with the latest frauds, abuses and other patterns of non compliance
- They provide an actuarial basis for case selection based on parameters such as
 - Strike Rate
 - Probability of Adjustment
 - Estimated Dollar Adjustment



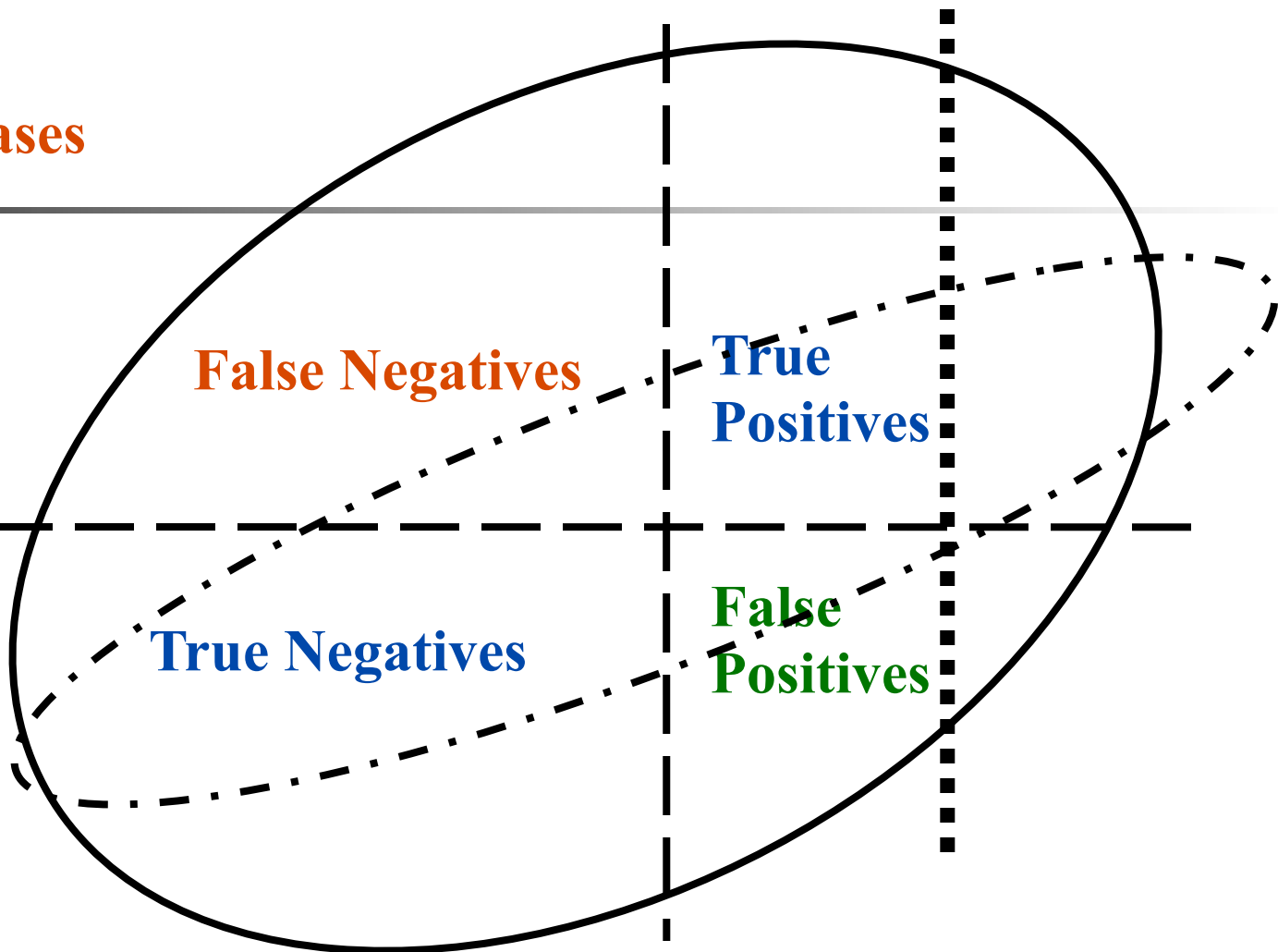
Lift Chart



Aberrant Cases

**Baseline
Separating
Aberrant
from
Acceptable
Cases**

Acceptable Cases



False Negatives

**True
Positives**

True Negatives

**False
Positives**

Cutoff used by Classifier



Business Models

- These include the expert rules used by compliance staff to select cases and to assign treatments
- These are based on expert judgment & experience
- They capture the nuances that apply to particular cases and issues
- They have short cycles



Case Selection

- It is truism, backed up by extensive scientific research, that the best case selection decisions are based on a combination of the following:

Expert + Actuarial
Judgment Prediction



Case Selection

- Actuarial prediction gives case selection staff the probability of a case being a 'true positive' rather than a 'false alarm' while
- Expert judgment includes factors and issues that are not included in the Analytical model thus improving the overall precision of the selection decision



Some Challenges



Staffing

- Underestimated numbers and types of skills required
- Critical Skills needed include:
 - Linking and Matching
 - Model Evaluation especially to do cost-effectiveness studies
 - Data Analytics
 - Business Engagement
 - Model Integration and Tuning



Software

- Staff idiosyncratic preferences for different packages
 - SAS
 - Rattle and R
 - Weka
 - Teradata Teraminer
 - SQL
- Staff are allowed to use the packages they prefer. Some write their own routines



IT Support

- Originally forecasted ten servers required each having ten miners and modellers as users
- This part of the equation was correct
- Installed two servers and a some minor ones



IT Support

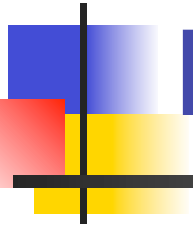
- The two major servers each have 8 nodes with 20 Gb RAM each
- Both have a half a Terabyte of storage
- They have proven totally inadequate to do mining and modelling on large datasets



IT Support

- Often run out of storage space
- Recognised need for 64 bit architecture and have set up a network of Linux servers
- ICT staff are ignorant of Analytics and do not know how to support this function. This has created delays with deliveries

Career Prospects and Education





Career Prospects

- Demand for those with Analytics and Intelligence skills is very high
 - Supply does not currently meet demand
- Career Progression can include:
 - Data Miner
 - Senior Data Miner
 - Chief Data Miner
 - Director
 - Chief Analyst



Education

- Time is ripe for a Masters Degree in Analytics just as there are now Masters Programs in Intelligence Analysis
- Feeder Courses at Undergraduate Level include:
 - Computer Science/Information Technology
 - Business Studies
 - Other Science and Engineering Disciplines



Education

- Need to include data mining as well as statistics in science, engineering and business studies programs at undergraduate level
- These are to train users in the application of these techniques to solve problems and reach decisions