CyberSecurity MALAYSIA

People First, Performance Now

MOSTI Ministry of Science, Technology and Innovation

In conjunction with WIF-KL 2012

WIF KL WORLD INNOVATION FORUM KUALA LUMPUR

csm-ace 2012 CYBER SECURITY MALAYSIA AWARDS, CONFERENCE & EXHIBITION

CYBERSECURITY RISK & COMPLIANCE FOR ECONOMIC TRANSFORMATION

www.csm-ace.my

Case Study: Big Data Forensics

Neil Meikle, Associate Director,
Forensic Technology, PwC

6 November 2012

About me

- Transferred to Kuala Lumpur from PwC’s Forensic Technology practice in London, England
- Specialist in advanced data analytics, computer forensics and e-Discovery
- Background in IT consultancy and data analysis

Neil Meikle

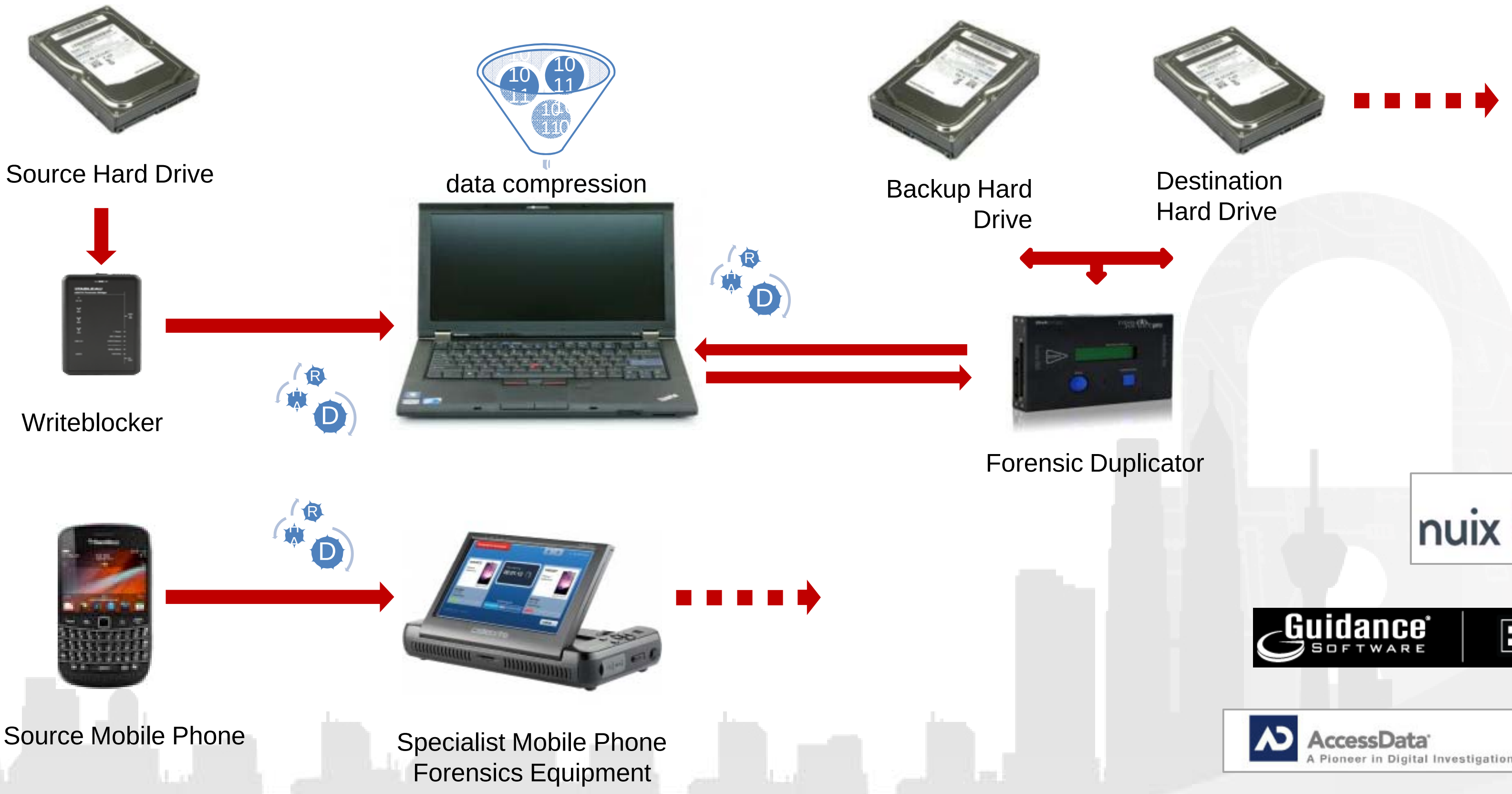
Forensic Technology, PwC

Tel: +60 3 2173 0488

Mobile: +60 17 243 7641

Email: neil.meikle@my.pwc.com

Some background: computer forensics enables the forensic capture and investigation of electronic devices



Source Hard Drive

data compression

Backup Hard Drive

Destination Hard Drive

Writeblocker

Forensic Duplicator

Source Mobile Phone

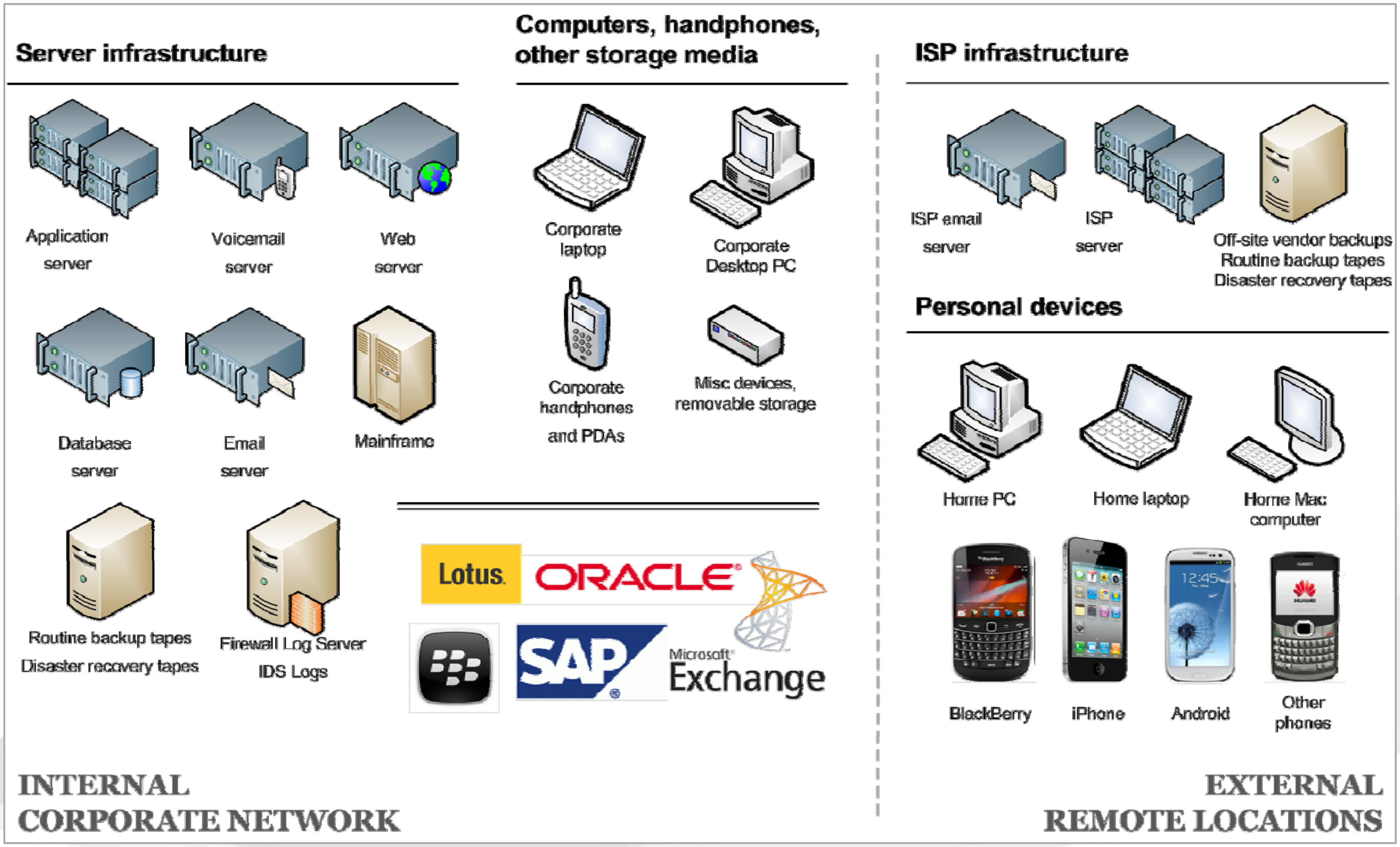
Specialist Mobile Phone Forensics Equipment

nuix

Guidance SOFTWARE | EnCase

AccessData A Pioneer in Digital Investigations Since 1987

A key challenge in fraud investigations: the typical sources of electronic information are expanding...



Server infrastructure

Application server
Voicemail server
Web server
Database server
Email server
Mainframe
Routine backup tapes
Disaster recovery tapes
Firewall Log Server
IDS Logs

Computers, handphones, other storage media

Corporate laptop
Corporate Desktop PC
Corporate handphones and PDAs
Misc devices, removable storage

Lotus. ORACLE SAP Microsoft Exchange

ISP infrastructure

ISP email server
ISP server
Off-site vendor backups
Routine backup tapes
Disaster recovery tapes

Personal devices

Home PC
Home laptop
Home Mac computer
BlackBerry
iPhone
Android
Other phones

INTERNAL CORPORATE NETWORK

EXTERNAL REMOTE LOCATIONS

How information forensic methods are changing

- Fraud investigations have made use of information forensics for many years to extract relevant information from electronic devices:
  - *A deleted document on an individual's laptop*
  - *A set of messages recovered from a Blackberry mobile phone*
- Relevant information will continue to be found in new places:
  - *A set of posting fragments from an individual browsing on Facebook on their laptop*
- But relevant information will also increasingly be found in larger data repositories and new data sources:
  - *An incriminating email on a corporate email server*
  - *Illicit transactions in a financial system*

We can use a new set of tools and techniques to process and analyse “big data”

- For unstructured data
  - We need to take large numbers of documents, emails, posts and other messages, automatically filter out the majority, then present the remainder for analysis (e.g. by a team of reviewers)
  - **This is E-DISCOVERY**
- For structured data
  - We need to transform large volumes of raw structured data into insight, e.g. identifying fraud, uncovering suspicious behaviour
  - **This is DATA ANALYTICS**

“Big data” isn’t just vast databases...
it can be huge numbers of emails and files too

Case study: Project codenamed “Apple”

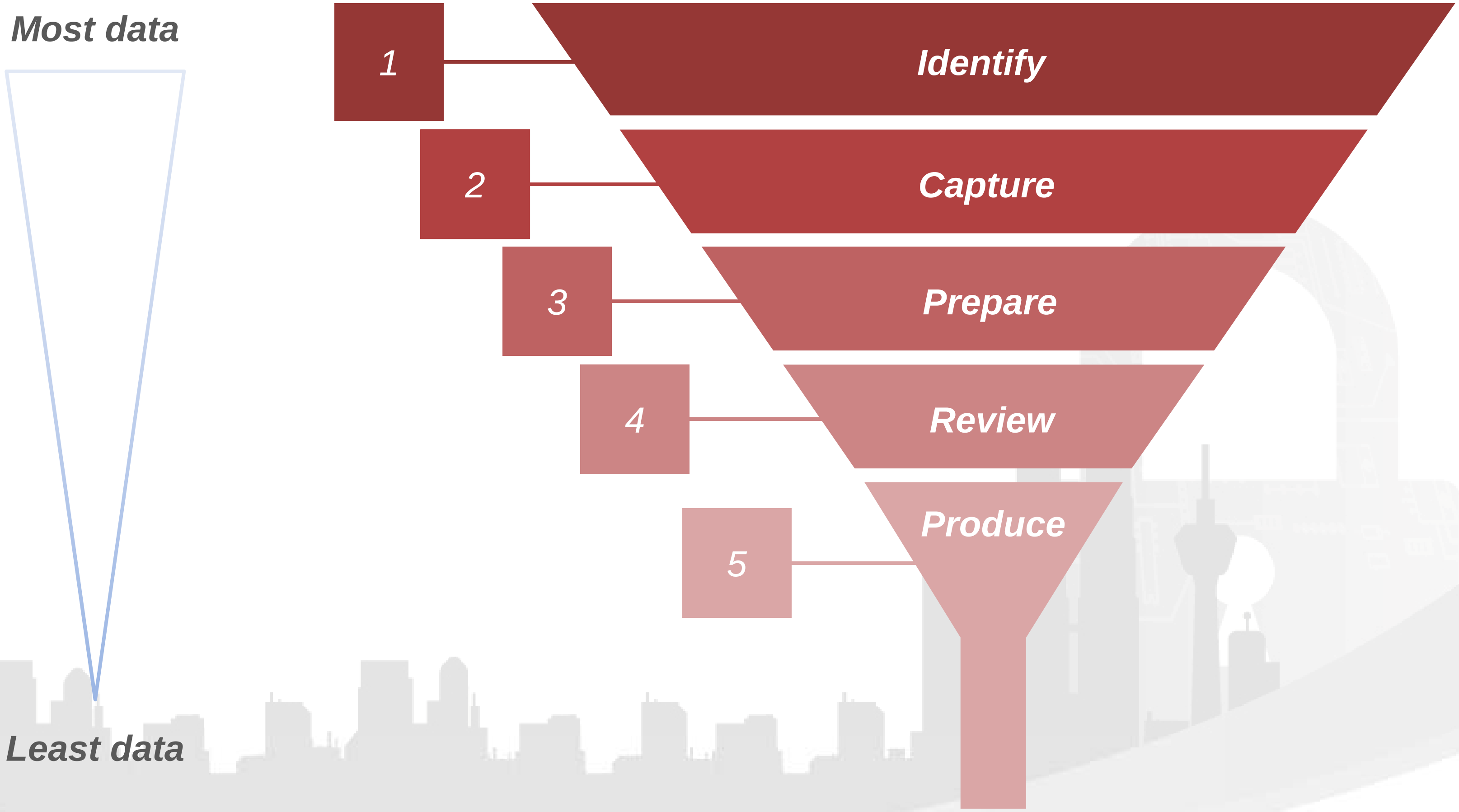
- An investigation and litigation e-disclosure exercise
- A financial organisation
- Billions of dollars of allegedly misappropriated funds
- Large volumes of structured and unstructured data
- Complex demands with non-standard (i.e. complicated) legal review

The unstructured data challenge

The e-Discovery challenges on Project Apple

- Capture of hundreds of thousands of documents from a foreign legal jurisdiction
- Review of hundreds of thousands of documents
- Translation of large numbers of documents into English
- Court deadlines
- Large number of reviewers
- Complex systems and processes
- Quality review
- Reconciliation

The e-Discovery filter: identify large amounts of data, but produce a much smaller set



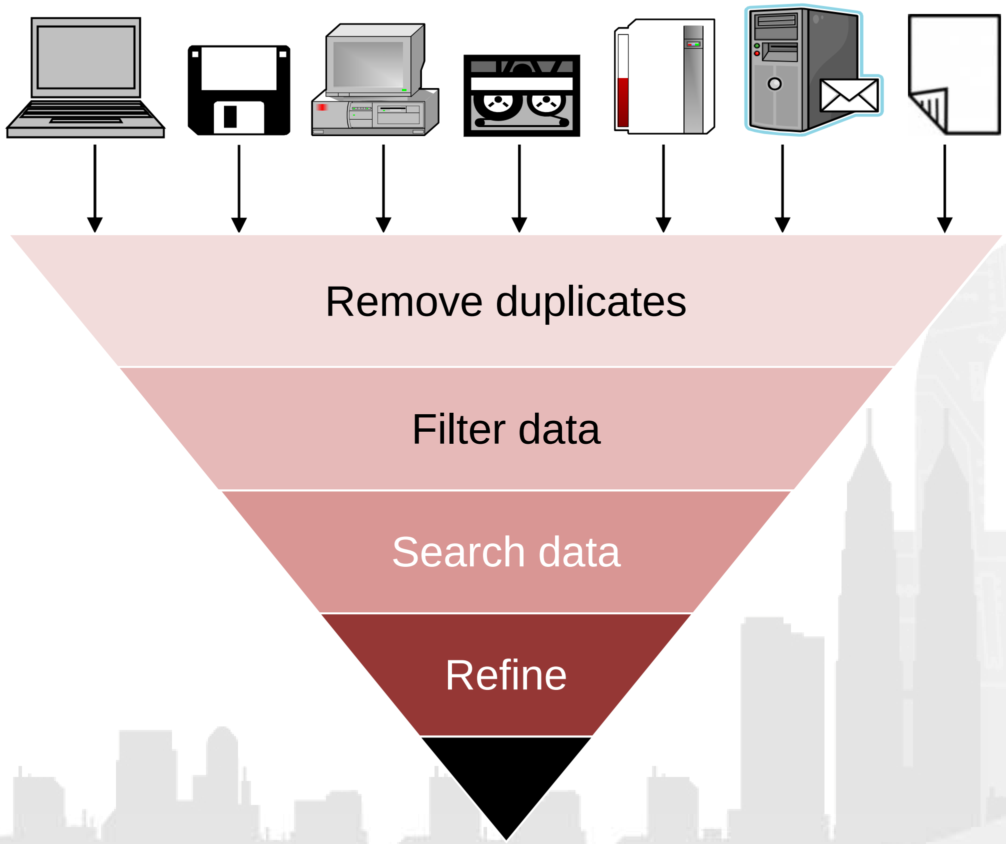
The e-Discovery filter

1 – Identify and 2 – Capture

- Sources of data?
- Relevant time periods and custodians
- Electronic vs hard copy
- Live vs static vs backup
- Early Case Assessment (ECA)

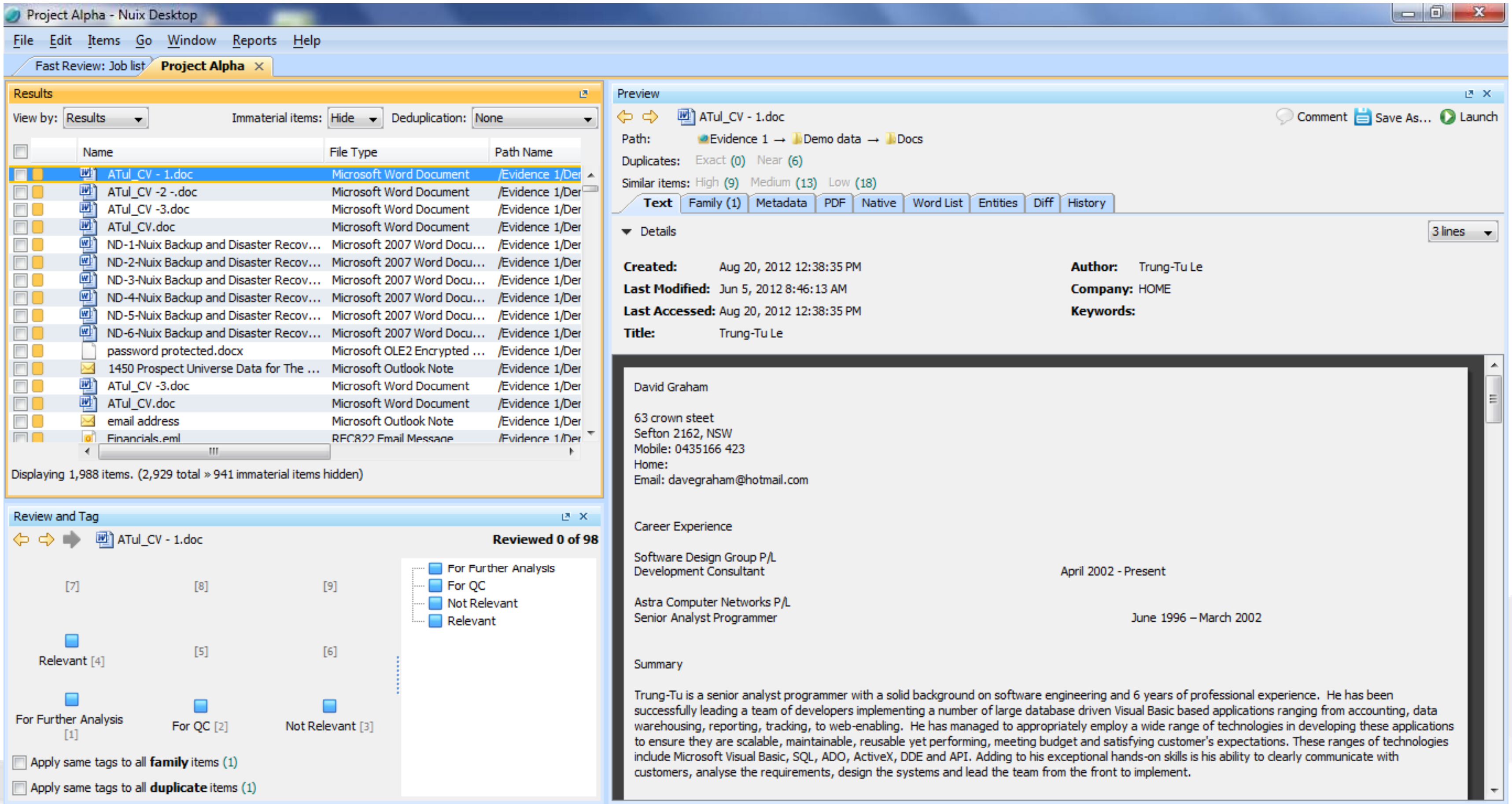
The e-Discovery filter

3 – Prepare



The e-Discovery filter

4 – Review



View by: Results
Immaterial items: Hide
Deduplication: None

Name	File Type	Path Name
ATul_CV - 1.doc	Microsoft Word Document	/Evidence 1/Der
ATul_CV -2 -.doc	Microsoft Word Document	/Evidence 1/Der
ATul_CV -3.doc	Microsoft Word Document	/Evidence 1/Der
ATul_CV.doc	Microsoft Word Document	/Evidence 1/Der
ND-1-Nuix Backup and Disaster Recov...	Microsoft 2007 Word Docu...	/Evidence 1/Der
ND-2-Nuix Backup and Disaster Recov...	Microsoft 2007 Word Docu...	/Evidence 1/Der
ND-3-Nuix Backup and Disaster Recov...	Microsoft 2007 Word Docu...	/Evidence 1/Der
ND-4-Nuix Backup and Disaster Recov...	Microsoft 2007 Word Docu...	/Evidence 1/Der
ND-5-Nuix Backup and Disaster Recov...	Microsoft 2007 Word Docu...	/Evidence 1/Der
ND-6-Nuix Backup and Disaster Recov...	Microsoft 2007 Word Docu...	/Evidence 1/Der
password protected.docx	Microsoft OLE2 Encrypted ...	/Evidence 1/Der
1450 Prospect Universe Data for The ...	Microsoft Outlook Note	/Evidence 1/Der
ATul_CV -3.doc	Microsoft Word Document	/Evidence 1/Der
ATul_CV.doc	Microsoft Word Document	/Evidence 1/Der
email address	Microsoft Outlook Note	/Evidence 1/Der
Financials.eml	RFC822 Email Message	/Evidence 1/Der

Displaying 1,988 items. (2,929 total » 941 immaterial items hidden)

ATul_CV - 1.doc
Reviewed 0 of 98

□ Apply same tags to all **family** items (1)
□ Apply same tags to all **duplicate** items (1)

ATul_CV - 1.doc
Path: Evidence 1 → Demo data → Docs
Duplicates: Exact (0) Near (6)
Similar items: High (9) Medium (13) Low (18)

Created: Aug 20, 2012 12:38:35 PM
Last Modified: Jun 5, 2012 8:46:13 AM
Last Accessed: Aug 20, 2012 12:38:35 PM
Title: Trung-Tu Le
Author: Trung-Tu Le
Company: HOME
Keywords:

David Graham

63 crown steet
Sefton 2162, NSW
Mobile: 0435166 423
Home:
Email: davegraham@hotmail.com

Career Experience

Software Design Group P/L
Development Consultant April 2002 - Present

Astra Computer Networks P/L
Senior Analyst Programmer June 1996 – March 2002

Summary

Trung-Tu is a senior analyst programmer with a solid background on software engineering and 6 years of professional experience. He has been successfully leading a team of developers implementing a number of large database driven Visual Basic based applications ranging from accounting, data warehousing, reporting, tracking, to web-enabling. He has managed to appropriately employ a wide range of technologies in developing these applications to ensure they are scalable, maintainable, reusable yet performing, meeting budget and satisfying customer's expectations. These ranges of technologies include Microsoft Visual Basic, SQL, ADO, ActiveX, DDE and API. Adding to his exceptional hands-on skills is his ability to clearly communicate with customers, analyse the requirements, design the systems and lead the team from the front to implement.

The e-Discovery filter

5 – Produce (disclosure rules)

- **UK:**
  - Civil Procedure Rules Practice Direction 31B – Disclosure of Electronic Documents
- **Malaysia*:**
  - The Rules of High Court 1980 (RHC) and the Subordinate Court Rules 1980 (SCR) govern discovery process
  - Unlike the UK CPR, the rules on discovery under both court rules remains unchanged, even with developments in IT
  - There is no specific provision in the RHC 1980 or any Practice Direction that contains guideline on e-discovery of electronically stored information (ESI)

** From: Discovery of electronically stored information (ES1) or e-discovery: the law and practice in Malaysia and other jurisdictions*

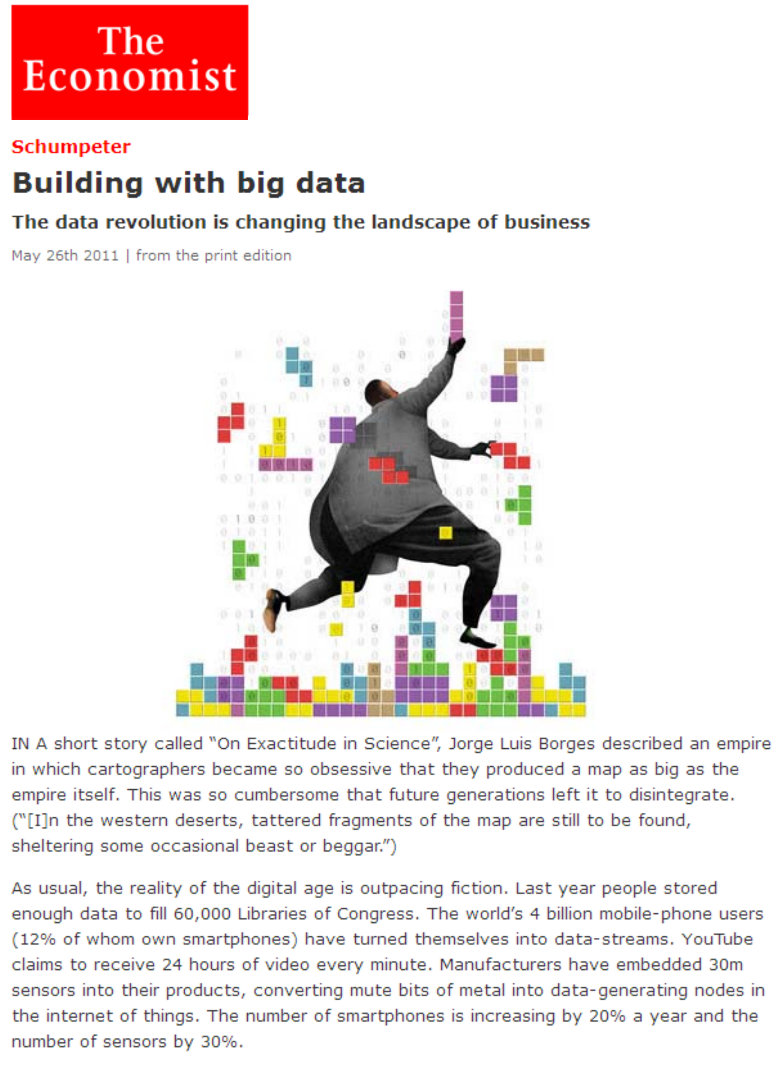
The e-Discovery filter
5 – Produce (case study example)

- Electronic vs printed
- Appropriate, agreed format
- Provided in a format that can be loaded into the opposing party's e-review platform

The structured data challenge

Big data = more potential insight, more evidence in fraud investigations

- ***Finance and retail*** (e.g. pricing and risk analytics)
- ***Utilities*** (e.g. smart usage analysis)
- ***Pharmaceuticals and health*** (e.g. smart patient monitoring and diagnosis)
- ***Supply chain and inventory*** (e.g. efficiency improvement through simulation modelling)
- ***Marketing and CRM*** (e.g. customer profiling and segmentation, customer acquisition and retention , customer value and profitability)
- ***Fraud investigation and prevention*** (e.g. suspicious transaction identification, bribery and corruption)



The Economist

Schumpeter

Building with big data

The data revolution is changing the landscape of business

May 26th 2011 | from the print edition

IN A short story called "On Exactitude in Science", Jorge Luis Borges described an empire in which cartographers became so obsessive that they produced a map as big as the empire itself. This was so cumbersome that future generations left it to disintegrate. ("[I]n the western deserts, tattered fragments of the map are still to be found, sheltering some occasional beast or beggar.")

As usual, the reality of the digital age is outpacing fiction. Last year people stored enough data to fill 60,000 Libraries of Congress. The world's 4 billion mobile-phone users (12% of whom own smartphones) have turned themselves into data-streams. YouTube claims to receive 24 hours of video every minute. Manufacturers have embedded 30m sensors into their products, converting mute bits of metal into data-generating nodes in the internet of things. The number of smartphones is increasing by 20% a year and the number of sensors by 30%.

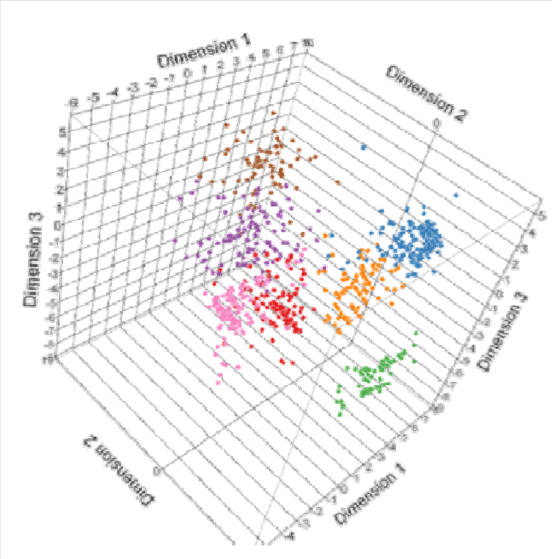
How we supported our investigation by transforming raw transactional data into insight

- Raw data = transactions
- Data recovered from financial systems
- Many transaction types
- Large volumes of data
- We needed to:

  (A) Transform

  (B) Visualise
- It can also be a requirement to:

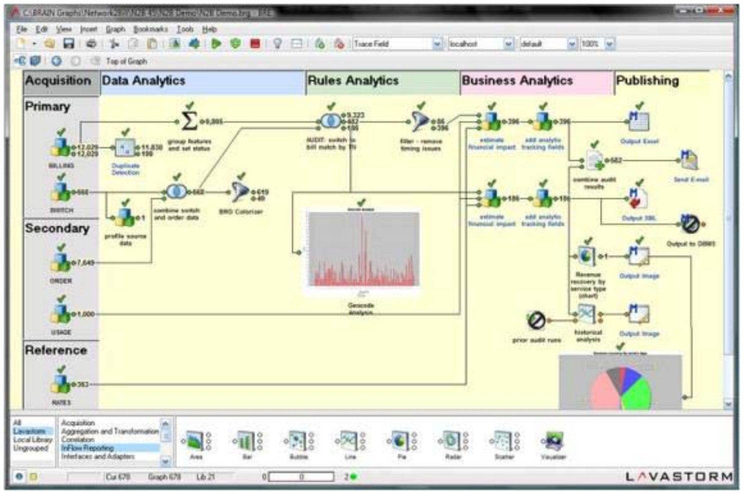
  (C) Statistically analyse



(A) Transforming data

Processing raw data to answer important questions

- Correcting data quality issues and parsing
- Profiling and analysing patterns
- Standardising and de-duplicating
- Matching, correlating and reconciling
- Aggregating and transforming
- Analysing complex data flows
- Producing dashboards

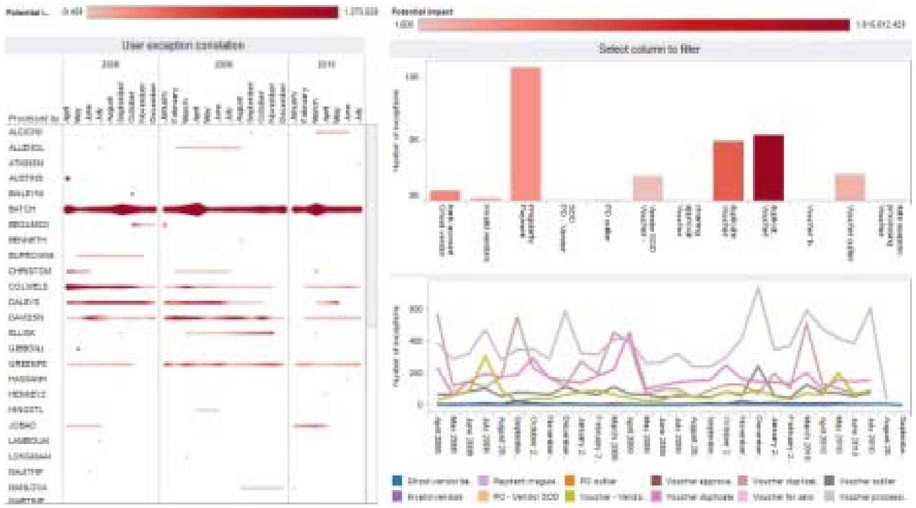


python™ Microsoft® SQL Server® LAVASTORM analytics

(B) Visualising data

Presenting data in an interactive, intuitive way

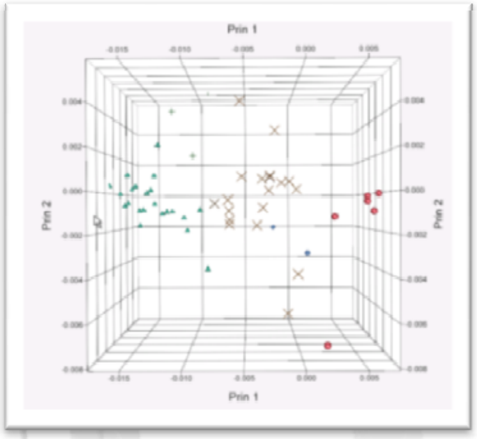
- Visualisation tools are used to explore, interpret and present data
- Visualisation dashboards enable interactive search and filtering
- A different perspective on large volumes of data



(C) Advanced techniques (statistical analysis)

Sophisticated analysis to detect unusual activity

- What was the next step if visualising the data hadn't answered our questions?
- Use of aggregated metrics created during the transformation phase
- Automatic classification of loans into groups – data driven
- Creating groups with similar behaviour can separate the normal users from the suspicious users



A case study involving advanced analytics: Project Digital - detecting procurement fraud

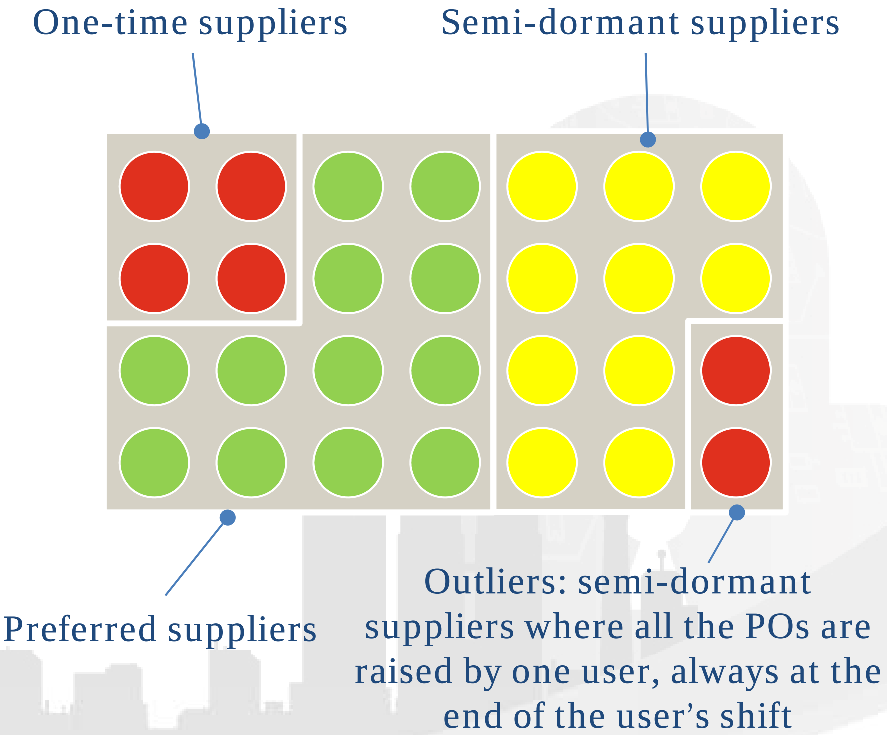
- A TV production and broadcast company uncovered a false invoicing fraud (by chance)
- The client suspected other instances of false invoicing fraud over a period of two years
- For the time period in question, procurements totalled approx. 200,000 transactions and 9,500 vendors
- These transactions exhibited a huge range of PO values: from a few pounds to hundreds of millions
- We were not informed of which transactions the client had identified as fraudulent

Can this type of problem be solved with data matching and red flag analysis?

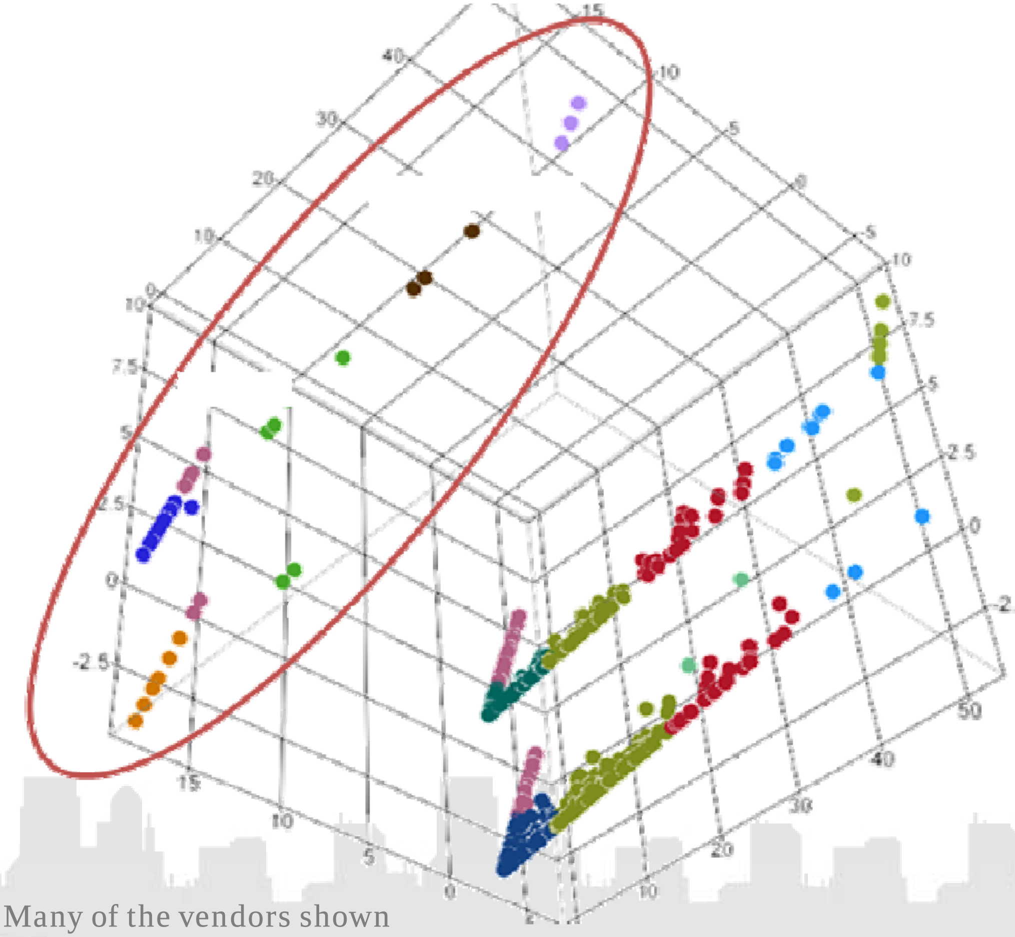
- Typically we would solve this type of problem with a traditional red-flag approach, i.e. decide whether any transactions broke pre-agreed rules
- But traditional data-driven fraud techniques have limitations
  - They tend to be **rule based**
  - Exceptions are only treated in **isolation**
  - They assume that the fraud pattern is **known**
- In this scenario there are multiple indicators but no clear rules that definitely show that fraud has occurred

Clustering suppliers to identify outliers

- Grouping together suppliers based on their characteristics (and generated events)
- Suppliers that are different in some way are identified and investigated further
- We looked for behaviours that differed from the "typical" vendor


One-time suppliers
Semi-dormant suppliers
Preferred suppliers
Outliers: semi-dormant suppliers where all the POs are raised by one user, always at the end of the user's shift

Project Digital: Key findings



Note: Many of the vendors shown on this diagram overlap with others

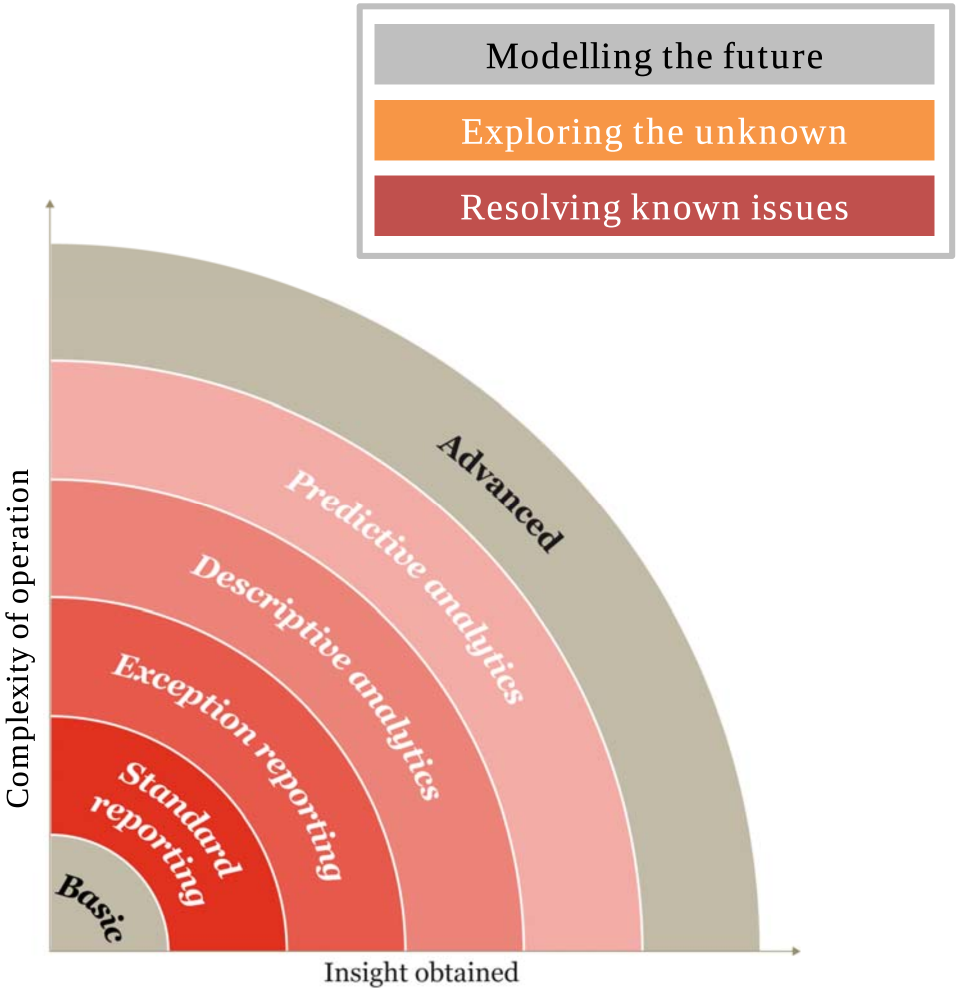
- Uncovered 42 “outlier” vendors for further investigation

- Two of these vendors were confirmed as the anonymised frauds

Structured data analytics is not just about reporting on known issues or frauds

Data analytics has an increasing role to play in supporting investigations and internal audit functions

- Proactively detecting fraud
- Helping make the investigations process more efficient
- Continuous transaction monitoring
- Predicting future events



Big data forensics - summary

- Fraud investigations have made use of information forensics for many years
- We also need a new set of tools and techniques to process and search “big data”

- *E-Discovery tools* take large numbers of documents, emails, posts and other messages, automatically filter out the majority, then present the remainder for review
- *Data analytics tools* transform raw structured data into insight through processing, transformation, visualisation, and statistical analysis

Thank you

Neil Meikle

Forensic Technology, PwC

Tel: +60 3 2173 0488

Mobile: +60 17 243 7641

Email: neil.meikle@my.pwc.com