The Actuarial Profession
making financial sense of the future

GIRO conference and exhibition 2010

Catherine BARTON, Partner – Ernst & Young Actuarial Services
Nicolas MALLISON, Director – Ernst & Young Fraud Investigations & Disputes

Proactive Data Driven Counter Fraud

Mining for digital gold

12-15 October 2010

A forensic data analytics perspective

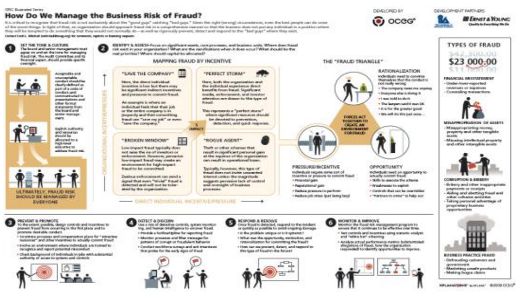
“The Governor of the Bank of France, Christian Noyer, said SocGen was not guilty of wrongdoing and that M. Kerviel's ability as a "computer genius" had allowed him to escape the bank's internal controls.”

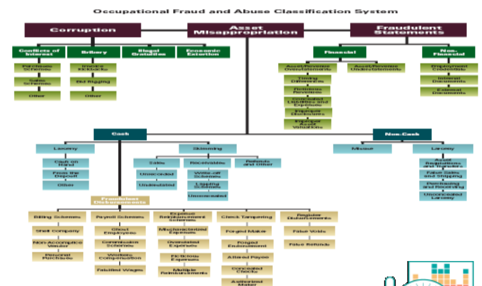
The Independent, Sean Farrell, Financial Editor

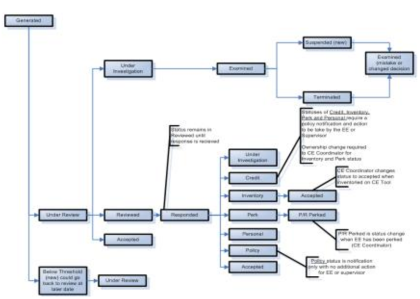
Introducing Proactive Data Driven Counter Fraud

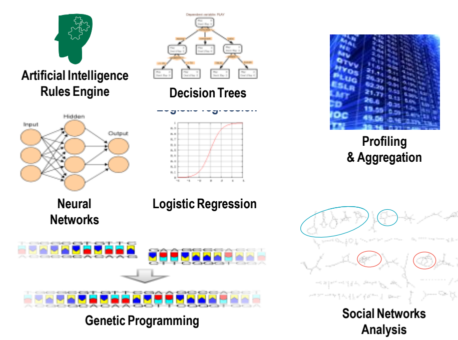
Manage fraud risks proactively

Quickly and cost effectively identify the most relevant information by leveraging the knowledge inside large scale datasets







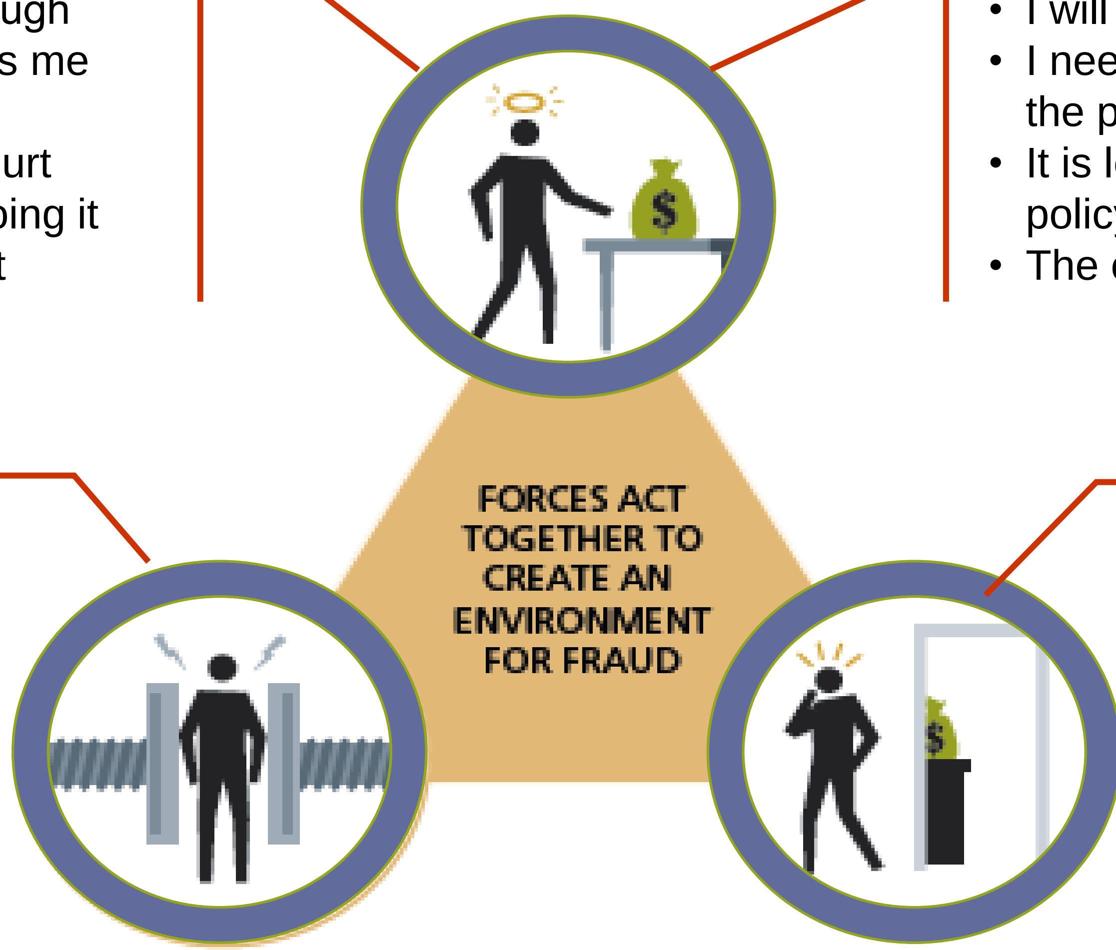


The Fraud Triangle

- I don't get paid enough
- The company owes me anyway
- Nobody is getting hurt
- Everyone else is doing it
- Nobody will find out

RATIONALISATION

- I will just do this once
- I need to recoup some of the premium I paid
- It is less than 1p on every policy
- The company can afford it



- Financial gain: large or small
- Reputational gain
- Reduce pressure to perform
- Reduce job stress
- Malignant narcissism: Psychological gratification

PRESSURE/INCENTIVE

OPPORTUNITY

- Skills to execute the fraud
- Weaknesses to exploit
- Controls that can be overridden
- Partners in crime to help out
- Access to confidential business information

Fraud Triangle in Insurance

Source	Rationalisation	Pressure/Incentive	Opportunity	Example Fraud
Employee / Internal – Theft from company	"The company owes me a raise"	Large debts	Working in claims department	Pays false claims in collusion with accomplice ; steals mass data
Soft – Opportunistic low level crimes	"It's a large company – it's not like we're robbing anyone"	Bit of easy extra cash	Belief genuine claims element can't be distinguished from fake	Camera stolen and additionally decides to add on ipod as want to get new model
Hard – Organised crime	"Their claims handling is weak – they will never find us out"	Large financial gains	Willing partners in crime with skills to carry out scheme	"Crash for cash" schemes.

The 2 key stages of fighting insurance fraud

Detection versus Investigation / Review

DETECTION

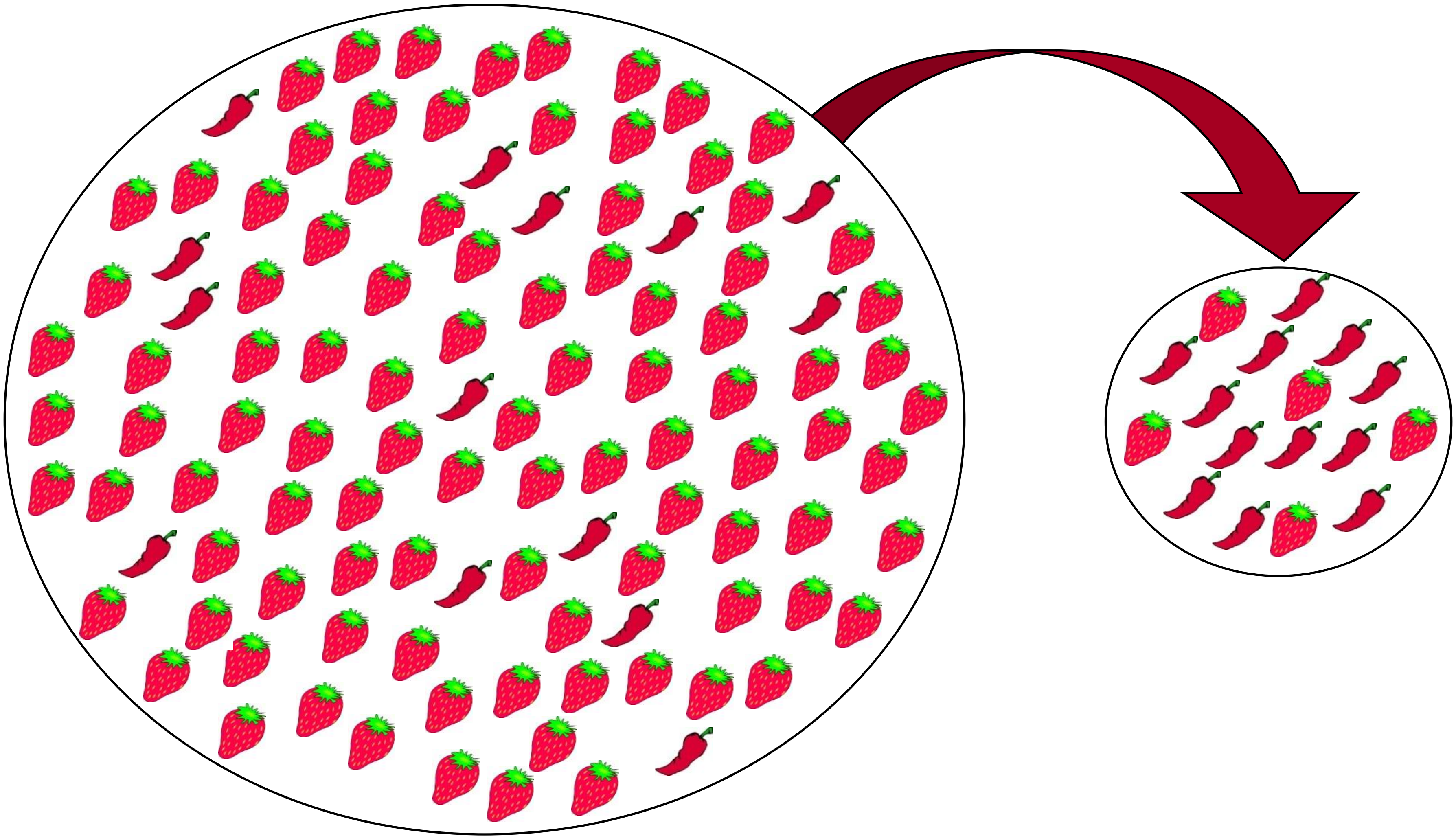
- “Needles in the big haystack” perspective: Detection is an analytical statistical game with goal improving the odds of finding the fraud (“needles”) and focus / prioritise further investigation efforts
- The end game is isolating small “needle rich” pool of hay rather than seeking to pick out individual “needles” (hard). It is not about finding the “needles”, it is about removing the hay
- Effectiveness of Detection can be measure with 2 metrics:
  - detection rate: % of confirmed fraud case in identified potential fraud pool
  - False positive rate: % of non fraud case in identified potential fraud pool

REVIEW

- Process of manually reviewing referred potential fraud cases with the aim to prove or disprove fraud
- Use of technology tool (data visualisation / link analysis, database lookup) to help a robust and cost effective identification and collation of evidence amongst databases and electronic documents
- Understanding of modus operandi of confirmed fraud cases can be fed back in data acquisition/preparation and business understanding
- Investigation outcomes can be used to train machine learning predictive technology that can continuously improve the logic and performance of detection

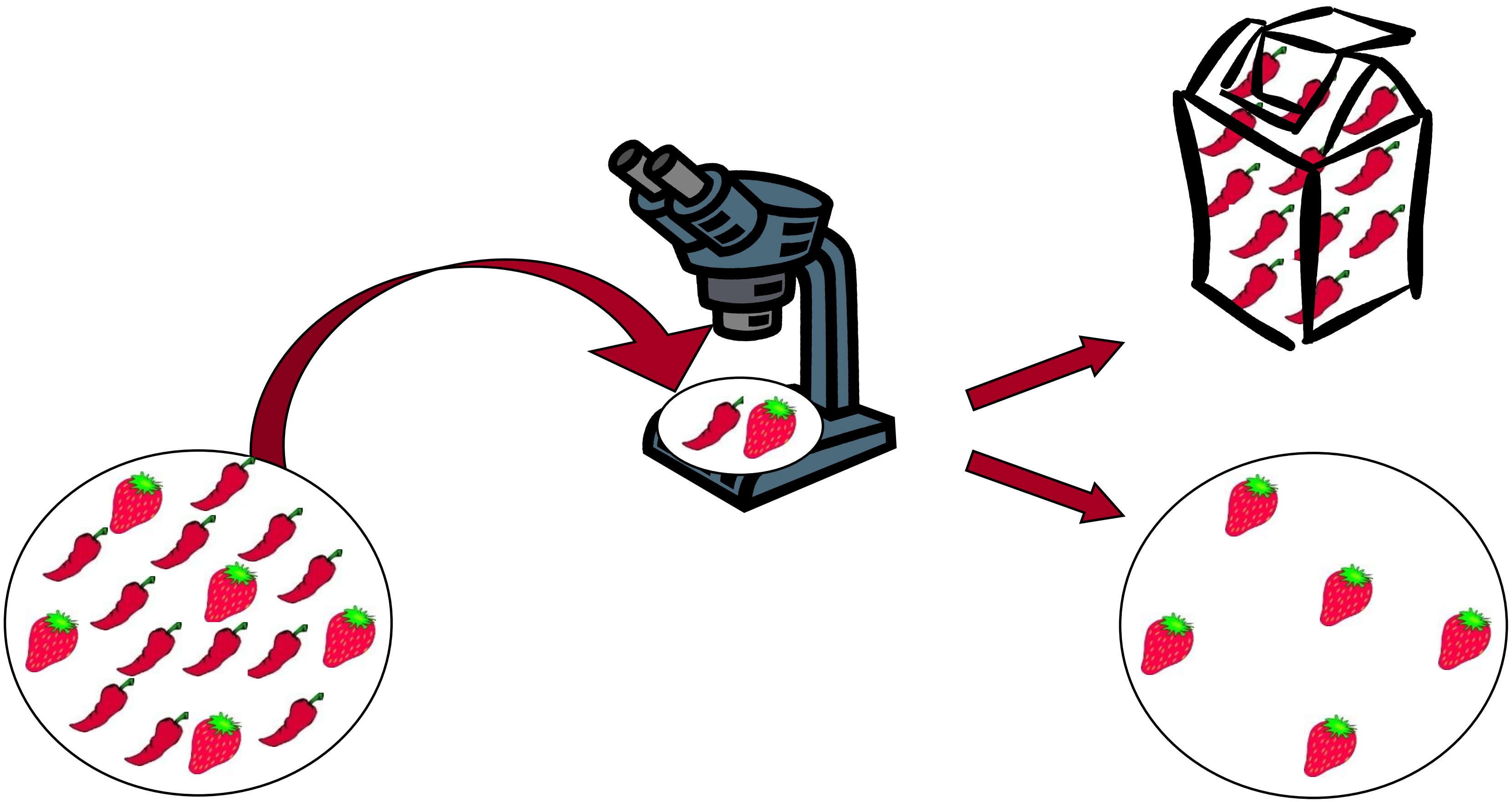
The 2 key stages of fighting insurance fraud

Detection



The 2 key stages of fighting insurance fraud

Investigation / Review



Where Proactive Data Driven Counter Fraud technology can help?

Determine the right claims to focus on

– Identify and prioritise the most suspect cases at the earliest possible opportunity based on consistent non judgemental data driven inspection of all claims

Decide what is the most appropriate action

– Integrate in the claim handling workflow risk based decisions points based on claim fraud propensity and differentiate actions based on suspicion level and mitigation potential
– Utilise insight into the key dimensions driving each type of fraud detected in order to “fraud proof” existing processes and systems

Data driven investigation case management

– Presentation of referral cases to fraud investigators in a user friendly web case management interface giving them interactive access to all the underlying data for link analysis and investigative search purposes

Measure and manage results

– Analysis of trends in fraud detection and mitigation performance through interactive web based management dashboard

What can Proactive Data Driven Counter Fraud achieve?

Significantly increase the detection rate and decrease the false positive rate of currently detected types of fraud

– This means the return on fraud investigation/review spend can be optimised as special investigation units, loss adjusters and claims handlers can prioritise their efforts according to the fraud propensity of each claim

Detection of a new unknown and emerging type of fraudulent claims through proactive identification of unusual activity and anomalous patterns

– This enable a much earlier identification of new and emerging type of fraud enabling over time the saving of a large quantity of payments that would have occurred until the new type of fraud is detected either by chance or because the repetition patterns in data become obvious

Proactive Data Driven Counter Fraud

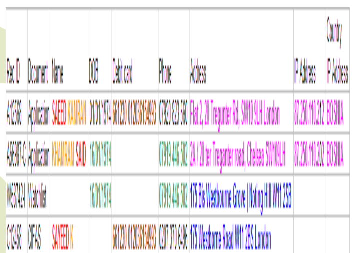
Data Driven Fraud Detection & Intelligence Framework

Acquire Data → **Transform Data** → **Analyse Data** → **Access Data** → **Decide**

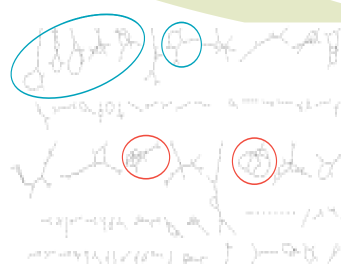
INTERNAL
- Claims history
- Policy history
- Investigation outcomes
- Watch lists
- Access Control

EXTERNAL
- 3rd party
- Fraud data
- Cross industry data sharing

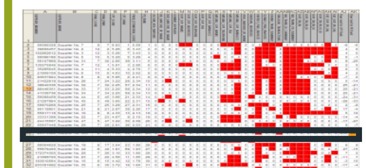
INVESTIGATIVE DATA LINKING



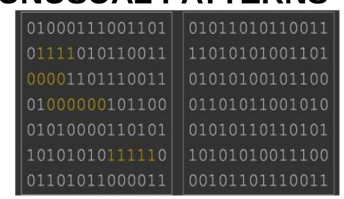
SOCIAL NETWORKS



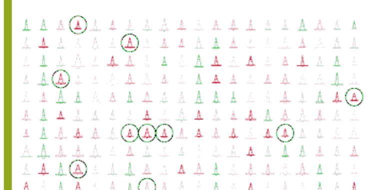
RED FLAGS RULES



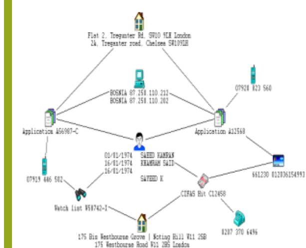
UNUSUAL PATTERNS



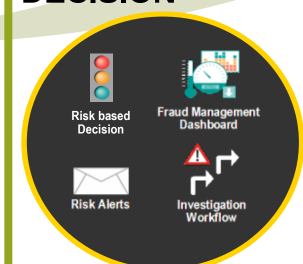
PREDICTIVE MODELS



SEARCH, EDIT & VISUALISE



RISK BASED DECISION



"It is not about finding the needle, it is about removing the hay"

OUTCOMES CAPTURE

Approaches to Fraud Detection

Whistle blowers

Investigative Data Linking / Multi-level Association

- Advanced data matching techniques and algorithms that seek to automate on large data sets the thought process of a human investigator
- Analysis of link patterns between policies, claims and entities in the resulting event networks

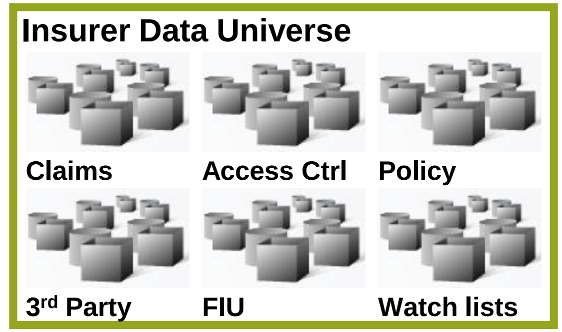
Text mining

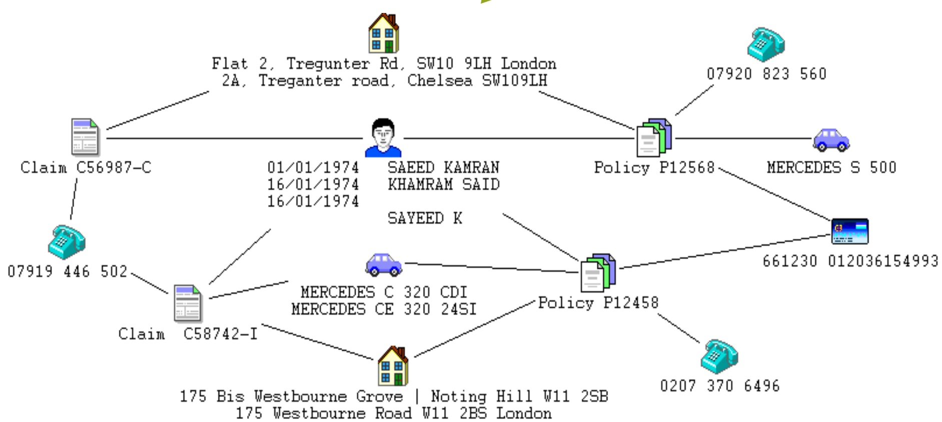
- Automatically extract information from free text notes, maps to concepts / entities and taxonomies to facilitate further down searches for unusual and predictive patterns in data across vast amount of records and variables

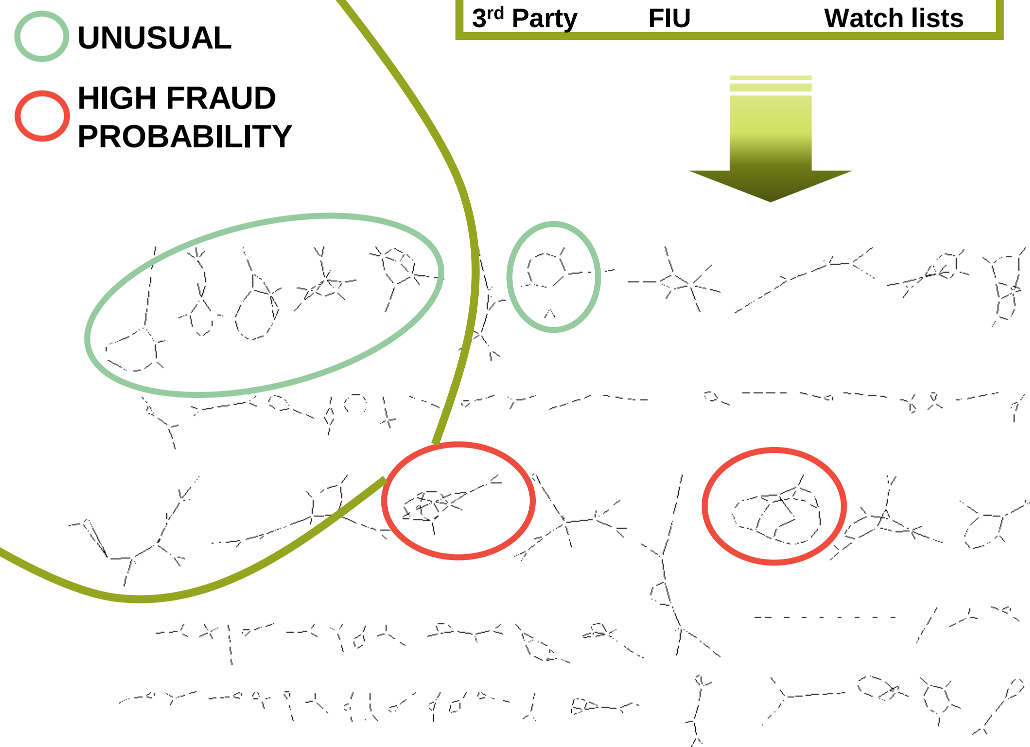
Investigative Data Linking / Multi-level Association

Illustrative example

Rec ID	Document	Type	Name	DOB	Debit card	Phone	Address	Car
P12568	Policy	holder	SAEED KAMRAN	01/01/1974	661230 012036154993	07920 823 560	Flat 2, 20 Treguner Rd, SW10 9LH London	MERCEDES C 320 CDI
C56987-C	Claim	claimant	KHAMRAM SAID	16/01/1974		07919 446 502	2A / 20 ter Treganter road, Chelsea SW109LH	MERCEDES CE 320 24SI
C58742-I	Claim	injured		16/01/1974		07919 446 502	175 Bis Westbourne Grove | Noting Hill W11 2SB	
P12458	policy	holder	SAYEED K		661230 012036154993	0207 370 6496	175 Westborne Road W11 2BS London	MERCEDES S 500







Rules and models to identify suspects

- Links with previous suspected fraud activity
- Unusual "ring" – circular link patterns
- Unusual combination of red flags such as:
  - Indirect links between claimants and 3rd parties
  - Unusual pattern of information access
  - Evidence of potential data obfuscation
  - Hit on fraud hot list / third party risk indicator

EXAMPLE ONLY USING ALTERED DATA

Text Mining for fraud detection

- Intelligent spell checking
- Dictionaries, External or user-built thesauri
- Categorization
- Clustering
- Entity extraction
- Natural language search
- Interactive multi-dimensional analysis
- Visual link analysis

- Pattern definition language:
  - Proximity (N terms, sentences, paragraphs)
  - Sequence
  - Negations
  - Sentiment
  - Synonyms
  - Hierarchical thesaurus relations
  - Phonetics
  - Regular expressions
  - Morphology

NAME: [redacted]
CHART NUMBER: 16635
DATE: 07/14/03
NEW PATIENT CONSULTATION

REFERRING PHYSICIAN: Dr [redacted]

OTHER PHYSICIANS: Dr. [redacted]

REASON FOR REFERRAL: Evaluation of breast cancer.

PATIENT'S CHIEF COMPLAINT: Breast cancer.

DATA SOURCE: The patient who is a good historian and she brought extensive old records.

HISTORY OF PRESENT ILLNESS: This is a 67 year-old Caucasian female with a prior history of rectal cancer and a more recent history of carcinoma of the left breast status post lumpectomy, adjuvant radiation therapy and now adjuvant Tamoxifen. She

Technology Demonstration Text Mining on claims notes data with Megaputer's PolyAnalyst

PAST MEDICAL, SURGICAL AND PSYCHIATRIC HISTORY:

1) Carcinoma of the left breast.
2) Status post lumpectomy and lymph node dissection and adjuvant radiation therapy.
3) She has a history of asthma.
4) She has a history of restrictive pulmonary disease status post thoracotomy.
5) She has had a thoracotomy because of prolonged pneumonia versus radiation pneumonitis.
6) She has diabetes.
7) She has had rectal cancer.
8) She has had resection of here colon.
9) She has a history of multiple polyps.

Approaches to Fraud Detection

Rule-based approach

- look for matches with known fraud schemes – low hanging fruit
- Relatively rigid, slowly or non adaptative and easy to circumvent

Model-based / Forensic Data Mining approach

- Look for specific patterns in data across vast amount of records and variables
- Unusual patterns may point towards new type of emerging fraud
- Automatically infer from data patterns closely resembling previous case of fraud without having to specify a rule (predictive fraud propensity model).

Interactive aggregation, visualization and reporting

The difficulty of fabricating data

Unusual Patterns

Which one of the 2 sequences of numbers below is NOT a random string of 0/1 with ½ probability ?

010001110011011	010110101100110
011110101100110	110101010011011
000011011100111	010101001011001
010000001011000	011010110010101
010100001101010	010101101101010
101010101111101	101010100111001

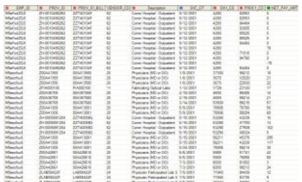
Sometimes less is more

Rules-based with model-based aggregation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
1	SUPPLIER_NUMBER	SUPPLIER_NAME	RAW_SCORE	RAW_RANK	CPT_SCORE	CPT_RANK	PROB_ESTIMATION_SCORE	PE_RANK	ACCOM_ADD_MATCHES	BIS_MATCHES_BY_NAME	BOE_MATCHES_BY_NAME	DUMMY_SUPPLIERS	DUP_SUP_NAMES	DUP_SUP_ADDRESSES	DUP_SUP_TELEPHONE	DUP_SUP_FAX	DUP_SUP_BANKACCS	MISSING_SUP_NAMES	MISSING_SUP_TELEPHONE	MISSING_SUP_FAX	MISSING_SUP_BANKACCS	OFAC_MATCHES_BY_NAME	UNUSUAL_ADDRESSES	RAPIDLY_PAID_INVOICES	SLOWLY_PAID_INVOICES	ROUND_TRANS_SUMMARY	SINGULAR_TRANSACTIONS	WE_TRANS_ID_SUMMARY	WE_TRANS_PD_SUMMARY	DUP_SN_IN	DUP_SN_IN_ID	DUP_SN_IN_IA	DUP_SN_IN_IA_ID	KEY_SUP_BY_VAL	KEY_SUP_BY_VOL	Raw rank Vs CPT Rank	Raw rank Vs PE Rank
2	56098329	Supplier No. 7	9	7	6.83	1	5.59	1	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	1	0	-6	-6
3	39608457	Supplier No. 2	12	2	5.26	6	3.43	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	4	0
4	152862612	Supplier No. 3	12	3	5.26	7	3.43	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	4	0
5	56096166	Supplier No. 5	10	5	5.28	5	3.29	4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	1	1	0	-1
6	56107968	Supplier No. 31	7	30	2.85	39	3.11	5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	1	0	0	0	1	0	9	-25
7	135575846	Supplier No. 7'	12	1	5.81	3	2.88	6	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	1	0	0	0	1	1	1	1	1	1	0	0	2	5
8	35295645	Supplier No. 4	11	4	5.33	4	2.70	7	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	1	0	3
9	21906108	Supplier No. 6	10	6	4.63	10	2.52	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1	4	2
10	93637824	Supplier No. 8	9	9	5.95	2	2.41	9	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	1	1	0	0	0	0	1	1	-7	0
11	21422919	Supplier No. 26	7	24	3.22	29	2.34	10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1	1	0	0	0	0	5	-14
12	38640336	Supplier No. 32	7	32	2.25	57	2.34	11	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	1	1	0	0	0	0	25	-21
13	88446351	Supplier No. 33	7	33	2.25	58	2.34	12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	1	1	0	0	0	0	25	-21
14	41336736	Supplier No. 34	7	34	2.25	59	2.34	13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	1	1	0	0	0	0	25	-21
15	56088459	Supplier No. 44	5	45	3.66	21	2.33	14	0	0	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	-24	-31
16	21297864	Supplier No. 45	5	48	3.65	22	2.31	15	0	1	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-26	-33
17	58970268	Supplier No. 25	7	25	3.29	27	2.31	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	0	0	0	2	-9
18	56118951	Supplier No. 21	8	21	2.87	38	2.26	17	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	1	1	0	0	1	0	17	-4
19	55817118	Supplier No. 10	9	8	4.33	13	2.13	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	5	10
20	33331389	Supplier No. 22	7	23	4.67	9	2.10	19	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	1	0	1	0	0	0	0	0	0	-14	-4
21	24415587	Supplier No. 47	5	47	3.22	30	2.09	20	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-17	-27
22	55937448	Supplier No. 29	7	28	2.91	36	2.03	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	8	-7
23	36153012	Supplier No. 30	7	29	2.91	37	2.03	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	8	-7
24	26570922	Supplier No. 305	2	304	1.94	68	2.00	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	-236	-281
25	55653801	Supplier No. 45'	2	305	1.97	66	1.97	24	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-239	-281
26	26644632	Supplier No. 46	5	49	3.42	25	1.97	25	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	-24	-24
27	56054628	Supplier No. 19	8	17	3.54	23	1.95	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	1	0	6	9
28	55679400	Supplier No. 42	6	42	1.91	69	1.93	27	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0	27	-15
29	172374300	Supplier No. 28	7	26	3.15	33	1.90	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	0	0	1	7	2
30	21888720	Supplier No. 23	7	22	4.56	11	1.85	29	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	-11	7
31	183810564	Supplier No. 15	8	15	4.42	12	1.78	30	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1	1	1	0	0	1	0	-3	15
32	203381682	Supplier No. 35	6	35	3.72	18	1.60	31	0	0	1	0	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	-17	-4
33	36153390	Supplier No. 39	6	36	2.41	54	1.57	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	0	0	0	18	-4

Technology Demonstration

Healthcare Behavioural Deviation Fraud Detection Solution



Raw data

(CMS-15000, UB-92,etc.)

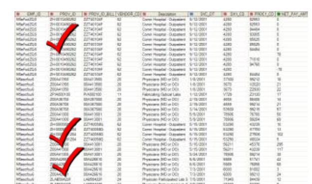
MediCop™

Determine the “norm”
for different situations
and identify deviations
from the norm

Identify anomalous
providers and their profiles

Combine obtained results
with relevant business
considerations

Suspicious
providers



YOUR CHOICE...

"IT IS BETTER TO DRAIN THE SWAMPS THAN TO FIGHT THE ALLIGATORS"

PROF IRVIN WALLER, DIRECTOR GENERAL,
INTERNATIONAL CENTRE FOR THE
PREVENTION OF CRIME - 8 MARCH 1996

Contact details

Catherine Barton

Partner
European Actuarial Services

Direct: +44 207 951 5503
Mobile: +44 771 104 8102
cbarton@uk.ey.com

Nicolas Mallison

Director – Fraud Data Analytics
Fraud Investigations & Disputes

Direct: +44 207 951 2750
Mobile: +44 792 082 3560
nmallison@uk.ey.com