West Midlands Police

Data Driven Insight &
Data Science Capability for UK Law Enforcement

‘The increasing availability of information and new technologies offers us huge potential to improve how we protect the public. It sets new expectations about the services we provide, how they are accessed and our levels of transparency’ (NPCC 2016, p.6).
Agenda

• Introductions and National Context
• Definitions
• ‘Influencers’ Data Science Project
• PTF supported National Data Science Capability project
• Data Science requirement gathering session
• Plenary
• Wrap-up
Framing of terms

— Big data:
  • Data sets which are beyond human cognitive capacity to make sense of and traditional data processing application software is inadequate to deal with them.

— Police Intelligence analysis:
  • Applying human domain expertise to, generally, pre-selected, cleansed and operationally relevant data sets for specific queries or testing hypothesis.

— Data analytics
  • Application of data processing and data engineering techniques to match, merge, de-duplicate, mine and then visualise synthesised data for broadly operational purposes.

— Data science
  • Data science – Application of mathematical and statistical methods to gain insight, understand patterns in data, describe the present and predict the future. A data led approach to complex policing challenges, which can be operational or organisational.
The Analytics Lab focused on the development of a risk model that assessed the likelihood and risk of a particular nominal becoming influential in co-offending.

Who are the influencers and who is at risk of becoming an influencer in co-offending?

Co-offending is one pathway into crime, vulnerability and harm
Experienced based hypothesis that when people first engage with crime they often do so with someone else
Reduce demand on conventional crime by identifying opportunities for early intervention

£109m estimated cost to West Midlands society over 5 years associated to active influencers & co-offending

Rich and unexplored data on nominals, networks and co-offending
Initial segmentation of 600,000 highlighted a large number of nominals who had co-offended multiple times and with multiple people
Structured and Unstructured data from key WMP and national systems has been analysed

Sourced and integrated data from 9 core WMP data sources and text mining was performed on IMS logs

Over 1TB of data volume
437 Source tables
10,908 Columns

Nominal table and interim data structure with:
• over 10m nominal entries
• ~30m event entries
• ~14m associations for over 4m nominals
Nominal Analytical Record with over 1300 KPIs

Over 267m rows
~900k Intel log entries
text mined
The DDI algorithm achieved 99% matching accuracy.

The matching process addressed 230m candidate pairs and produced 5.27 millions of resolved identities.

10.4m nominals entries

Simple matching:
- 7.1m unique identities
- 1.46 Entries (number of events) per identities

After matching algorithm:
- 5.2m unique identities
- 1.98 Entries (number of events) per identities

In one case the algorithm matched 108 nominal entries containing 27 different combinations of first name, surname, CRO Number and PNC ID.
A data driven process has been adopted to produce the risk model.

The journey to the risk model outputs has been based on the insight provided by the data validated with WMP SMEs.

- Core group of influencers identified based on behaviors and associations (RRR)
- Number of associations separating a nominal from the Core group (SNA)
- Target variable created (i.e. those who transition to become an “influencer” in co-offending) to drive KPIs and development of the model
The initial phases to understand and segment the data have provided some insight on the co-offending demographic.

- The younger you are when you commit a co-offending offence the more likely you are to go ahead to commit other crimes. Only 35% of people who first co-offend at age 23 or above go on to commit another crime.

- Insight from the data can be gained at each step of the analysis.

- Where a person is known to us (in any role) before they are 11 and their first solo offence is classified as wounding or above, 76% will have co-offended prior to the wounding offence.

- 52% of those who have committed a co-offending crime before the age 11 go on and commit a violent crime. The percentage drops to 29% if the age is 23 or more.
Influential Nominals are those nominals who are in the Core Group or associated with the Core group with co-offending crime association either directly or through another nominal.

#### Identifying the influencers starting from the Core Group

- The Core Group tend to be **3 years or older** and had committed **2 or more plus crimes** than their associates
- They co-offended with a wider network
- On average they committed **42 crimes and 10 co-offending crimes**
- Their associates committed 10% of all crimes ever
- 60% of the have been active within last 3 years
- Average years **36 years**
Based on the target group, 1349 KPIs have been identified.

The KPIs investigate and describe direct and indirect behaviour, the nominal network and how nominal interacts in the network. The metrics and variable contributes to define the risk model and prediction.
Among the 1349 KPIs, 32 KPIs have been identified for their strong predictive power.

The most important KPIs are related to the social network of the nominal.

4 out of the top 10 KPIs are related to network.
- Number of co-offending crimes
- Crimes committed by the members of your network
- Co-offending patterns - followed or before solo crime pattern (specific sequence)

Social network KPIs are around 3 times more predictive compared to "age of nominal".

More crimes are committed in your network more risky you are.

Nominals are most likely to turn influencers when they are between age 14.67 and 20.5. The risk of becoming influencer is still high up to age of 29 years. Past this age nominals more and more less likely to become influencers.

<table>
<thead>
<tr>
<th>No</th>
<th>Predictor name</th>
<th>Predictor importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of solo crimes committed by nominal associates prior to transition</td>
<td>614</td>
</tr>
<tr>
<td>2</td>
<td>Average number of solo crime per nominal associate prior to transition</td>
<td>435</td>
</tr>
<tr>
<td>3</td>
<td>Months nominal known to WMP for any reason</td>
<td>342</td>
</tr>
<tr>
<td>4</td>
<td>Change of nominal Page Rank between 211th and 5th week prior transition</td>
<td>269</td>
</tr>
<tr>
<td>5</td>
<td>Age of nominal in years</td>
<td>241</td>
</tr>
<tr>
<td>6</td>
<td>Number of times nominal has been stopped and searched within 48 months prior transition</td>
<td>234</td>
</tr>
<tr>
<td>7</td>
<td>Nominal Page Rank 1 week prior transition</td>
<td>224</td>
</tr>
<tr>
<td>8</td>
<td>Number of IMS logs for nominal prior to transition</td>
<td>209</td>
</tr>
<tr>
<td>9</td>
<td>Number of times nominal has been mention on IMS within 48 months prior transition</td>
<td>207</td>
</tr>
<tr>
<td>10</td>
<td>Recency in months for any event one month prior transition</td>
<td>194</td>
</tr>
<tr>
<td>11</td>
<td>Change of nominal Page Rank between 13th and 5th week prior transition</td>
<td>190</td>
</tr>
<tr>
<td>12</td>
<td>Change of nominal Page Rank between 26th and 13th week prior transition</td>
<td>183</td>
</tr>
<tr>
<td>13</td>
<td>Change of nominal Page Rank between 105th and 5th week prior transition</td>
<td>172</td>
</tr>
<tr>
<td>14</td>
<td>Age nominal first appeared on any system, in years</td>
<td>167</td>
</tr>
<tr>
<td>15</td>
<td>Change of nominal Page Rank between 53rd and 5th week prior transition</td>
<td>155</td>
</tr>
<tr>
<td>16</td>
<td>Nominal Page Rank 211 weeks prior transition</td>
<td>141</td>
</tr>
<tr>
<td>17</td>
<td>Number of crimes nominal co-offended prior transition</td>
<td>140</td>
</tr>
<tr>
<td>18</td>
<td>Sources pattern length prior transition</td>
<td>125</td>
</tr>
<tr>
<td>19</td>
<td>Nominal network degree 1 week prior transition</td>
<td>117</td>
</tr>
<tr>
<td>20</td>
<td>Share of acquisitive crime in nominal crimes portfolio prior transition</td>
<td>93</td>
</tr>
<tr>
<td>21</td>
<td>Number of times nominal was stopped and searched prior transition</td>
<td>75</td>
</tr>
<tr>
<td>22</td>
<td>Nominal mention in location specific IMS reports - V14</td>
<td>67</td>
</tr>
<tr>
<td>23</td>
<td>Nominal mention in location specific IMS reports - V10</td>
<td>63</td>
</tr>
<tr>
<td>24</td>
<td>Co-offending pattern length prior transition</td>
<td>63</td>
</tr>
<tr>
<td>25</td>
<td>Number of solo crimes committed by nominal prior transition</td>
<td>62</td>
</tr>
<tr>
<td>26</td>
<td>Nominal mentioned in prison related IMS report - V8</td>
<td>61</td>
</tr>
<tr>
<td>27</td>
<td>Nominal mention in IMS report related to drug habit/addiction - V1</td>
<td>60</td>
</tr>
<tr>
<td>28</td>
<td>Nominal mentioned in IMS report related to drug supply - V5</td>
<td>47</td>
</tr>
<tr>
<td>29</td>
<td>Nominal mentioned in administrative IMS reports - V20</td>
<td>43</td>
</tr>
<tr>
<td>30</td>
<td>Nominal minial distance to influencer in non co-offending social network 5 weeks prior transition</td>
<td>37</td>
</tr>
<tr>
<td>31</td>
<td>Nominal minial distance to influencer in non co-offending social network 1 week prior transition</td>
<td>37</td>
</tr>
<tr>
<td>32</td>
<td>Nominal mention in location specific IMS reports - V17</td>
<td>37</td>
</tr>
</tbody>
</table>
The network of a nominal is critical in his journey to influencer.

Via an analysis of the events and the journey of a nominal we can identify the steps in the journey and critical transition events.
Influential Nominals are those nominals who are in the Core Group and/or associated with the Core group via co-offending crime association either directly or through another nominal (associate of associate).

5% of nominals who committed crimes (ca. 30,000)
~40% of co-offending crimes
~23% solo crimes
Interventions with existing influencers (reactive)

- Comparison with those nominals on IOM radar

Potential Benefits

- **New nominals** identified through Foundation Phase Analytics Lab risk model – work with IOM to understand why not, and identify potential costs/benefits of applying interventions
- **Nominals** currently subject to IOM interventions not in Analytics Lab risk model – work with IOM to understand why not, and identify potential costs/benefits of ceasing interventions
- Nominals in different positions on both lists – work with IOM to understand why and identify difference in cost of differing levels of interventions
- Where nominals appear in both lists, could assess impact of interventions through checking impact of historic IOM interventions on nominals’ co-offending trajectory. Compare with results of OM study on impact of interventions

Potential business application/ putting model into production

- **TBC**

Interventions with soon-to-be influencers (proactive)

- Comparison with Neighbourhood Policing or FCID at-risk nominals

Potential Benefits

- **New nominals** identified through Foundation Phase Analytics Lab risk model – work with NPT/FCID to understand why not, and identify potential costs/benefits of applying interventions
- **Nominals** currently subject to NPT/FCID interventions not in Analytics Lab risk model – work with NPT/FCID to understand why not, and identify potential costs/benefits of ceasing interventions
- Nominals in different positions on both lists – work with NPT/FCID to understand why and identify difference in cost of differing levels of interventions
- Where nominals appear in both lists, could assess impact of interventions through checking impact of historic Neighbourhood policing/FCID interventions on nominals’ co-offending trajectory

Potential business application/ putting model into production

- Add markers in source systems to those individuals who reach crime association distance 3 or 4
- **TBC**

Some immediate operational next steps have been identified
Foundation Phase Analytics Lab

Quantitative Benefits

A range of approaches to calculating benefits identify different degrees of potential savings that could be made through addressing the riskiest individuals.

Approach 1: Calculate the cost associated with any solo crimes committed by influencer A after their transition to be an influencer.
- The total number of solo crimes committed by nominal A post transition is 12,582.
- The cost associated with 68% of these crimes is £58.43m over 5 years.
- The cost associated with 100% of these crimes = £85.92 over 5 years.

Approach 2: Calculate the cost associated with any co-offending crimes (#1 and #2) between influencer A and nominals B and C.
- The total number of #1 and #2 co-offending crimes, committed by A with either nominal B or C, is 3346 over a period of 5 years.
- The cost associated with 77% of these crimes is £13.82m over 5 years.
- The cost associated with 100% of these crimes = £19.94 over 5 years.

Approach 3: Look for any link #1 and #2 between influencer A and nominals B and C, AND then subsequent co-offending crime (#3) committed between B and C. Calculate the cost associated with crime #3.
- The total number of #3 crimes committed between nominals B and C is 533 over a period of 5 years.
- The cost associated with 72% of these crimes is £2.44m over 5 years.
- The cost associated with 100% of these crimes = £3.33m over 5 years.

£74.69m (resolved) over 5 years
£109.19m over 5 years
Police Transformation Project

Data Science Solution for UK Law Enforcement

October 2017
Four core ambitions
These ambitions translate into twelve statements summarising the detailed ambition for NAS

CONNECT TO COLLABORATE

1. Be a hub for evidence-based policing and policy
2. Resolve legal and ethical barriers to the use of data in law enforcement
3. Connect all partners across a collaborative services network

LEARN TO INFORM

10. Provide a feedback loop to assess old and new interventions
11. Build an evidence base across law enforcement
12. Recommend the next best action once the data is sufficient

IDENTIFY TO DESCRIBE

4. Classify demand in terms of threat, risk and harm posed to communities
5. Identify factors leading vulnerable people towards becoming an offender or victim
6. Identify cross-cutting organisational issues impacting law enforcement workforces, assets or operations

PREDICT TO PREVENT

7. Predict harm to victims and communities
8. Predict the escalation of criminality
9. Predict internal operational issues that will impact law enforcement effectiveness
What the solution doesn’t intend to do

Boundary Conditions for the National Analytics Solution

• NAS will not perform local data integration
  – The NAS will ingest data from local law enforcement agency systems as it is currently formatted and stored. While there is benefit to agencies individually if they can integrate their local datasets (as some are already doing), NAS can utilise data transformation techniques to integrate the data it ingests from agencies. As a result, NAS may share enhanced datasets back with individual agencies in some cases, if it is considered important to achieving law enforcement outcomes.

• NAS will not create a centralised law enforcement database
  – The NAS will safely and securely create a repository of the data it ingests from across law enforcement agencies in accordance with data security procedures and ethical standards. Combined data sets will be created for the purpose of developing insights in response to specific issues expressed as use cases. NAS will, therefore, not offer agencies a single database that can be queried, such as Police National Database (PND) that information can be drawn from.

• NAS will not perform enterprise search
  – The NAS will utilise a subset of law enforcement data drawn from across disparate local systems, rather than fully integrating local systems to provide a central or local enterprise search capability. When undertaking local data integration, agencies may choose to complete this functionality locally.

• NAS will not prescribe the law enforcement interventions to be must be put into action.
  – The NAS will be generating insights and sharing them with agencies, providing appropriate support to each agency to interpret the evidence and design new interventions where it is desired; ensuring that agencies retain their operational independence.
The capabilities of the solution will expand as NAS scales over the next three phases.

Decision gates between phases will allow the partners to plan and implement changes in a structured way.

**TS0**  
**Define the Strategy**
- Use case list established and prioritised
- To-be architecture defined
- Analytics ‘pod’ team located at WMP develop 2 use cases
- 'Hub and Spoke' operating model operational
- 3 months  
  April 2017

**TS1**  
**Build the Foundations**
- 2 priority use cases developed
- Value of 2 use cases proved by Foundation Phase platform
- Project delivery team supports phase
- Gate 1
  - 12 months
  - July 2017

**TS2**  
**Industrialise and Scale**
- 4 use cases developed into analytical services every 12 months
- Full NAS Platform procured
- ‘Pod’ teams established to scale analytical services
- ‘Hub’ Leadership team established
- Gate 2
  - 24 months
  - July 2018

**TS3**  
**Drive Forwards**
- 34 use cases tested and productised by the end of TS3
- Solution scaled to 44 law enforcement agencies
- Other local service providers are integrated
- Gate 3
  - 24 months
  - July 2020

- Partner Agencies joined as ‘Spokes’ to operating model
NAS will provide the capabilities for agencies to shape and implement interventions locally

As the number of partners increase, data quality will improve and the range of interventions enabled by NAS will grow
An assessment of the ‘as-is’ across the agencies

Guided Survey: Full results in business case

Each founding partner was asked questions to evaluate the analytical capabilities of 8 subjects:

- Vision and Strategy
- Information Management
- Sponsorship and Governance
- Organisation, Structure and Talent
- Data to Analytics Insights
- Capability Development
- Insight Driven Decisions
- Outcome Management

![Information Management Chart]

Can you access the data that you need for analytics?

All founding partners have found that data access has evolved based on business needs. Access to data is often limited and the 2 most common processes for accessing data are wildcard searches within a system before copying the information to Microsoft Excel or requesting a data extract from someone with a higher level of data access.

Is unstructured data captured and stored?

Unstructured data includes both free-text fields within a system and types of files that cannot be structured into a standard database format, such as images or PDF documents. Many different sources of unstructured data are stored – e.g. officer notes, crime reports saved within word documents and images. However, this data is often not utilised because analysts do not have the time to read large volumes of text or process multiple images because relevant information is not easily identifiable.

Are there guidelines for storage of data?

Most founding partners indicated that their data storage guidelines were based upon legal regulations. These founding partners could not guarantee that their data management processes were robust to ensure data was always disposed of to meet these guidelines. Some of the workshops identified instances where founding partners did not appear to have any data storage guidelines in place.
95 Use Cases into 57 Unique Questions into 7 Themes

The themes

- Threat Risk & Harm Identification
- Vulnerability Identification
- Demand Management
- Crime Identification
- Organisational Insight
- Advanced Visualisation
- Intervention
The two use cases which have been chosen address issues which are high priority amongst partners

Both are complex problems which require indicative and predictive insight to target interventions effectively

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Target Population</th>
<th>Prioritised Because</th>
<th>Likely Insight</th>
<th>Likely Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify and predict the common behaviours and factors which indicate low workforce wellbeing</td>
<td>Partner agencies’ own officers and staff (and potentially those of all 43 forces in England &amp; Wales)</td>
<td>Comparatively high abstraction rate in policing is creating a high cost of sick leave (£150m p/year across 8 forces)</td>
<td>Insight into key factors influencing workforce wellbeing and identification of who is at risk of suffering negative outcomes in the workplace</td>
<td>Enable partner agencies to act on low workforce wellbeing in a meaningful way to deliver tangible improvements to workforce abstraction rates, productivity and wellbeing</td>
</tr>
<tr>
<td>Assess the likelihood of people and places falling victim to modern day slavery</td>
<td>The vulnerable - individuals or places which do not have the full capacity to take action to protect themselves from harm</td>
<td>Modern day slavery is a complex new threat. Those vulnerable to it are often hidden from the police, and are likely to be at risk of threat or harm</td>
<td>Identification of who and where is at risk or vulnerable to modern day slavery, and increasing levels of risk around known persons</td>
<td>Enable partner agencies across policing, law enforcement and border control to recognise vulnerability and take early preventative action to support and protect the most vulnerable</td>
</tr>
</tbody>
</table>
Police Transformation Project

Workshop

October 2017
## Requirements gathering

<table>
<thead>
<tr>
<th>Content</th>
<th>Structure</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role/Actor</td>
<td><em>As a...</em></td>
<td>Explain who is interacting with the solution, why they are interacting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with it, and what they are trying to achieve</td>
</tr>
<tr>
<td>Task</td>
<td><em>I want...</em></td>
<td>Explain what information, capability or outcome the actor requires in</td>
</tr>
<tr>
<td></td>
<td></td>
<td>this scenario</td>
</tr>
<tr>
<td>Goal</td>
<td><em>So that...</em></td>
<td>Explain the desired result or outcome achieved as a result of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>actor completing that task</td>
</tr>
</tbody>
</table>
‘The increasing availability of information and new technologies offers us huge potential to improve how we protect the public. It sets new expectations about the services we provide, how they are accessed and our levels of transparency’ (NPCC 2016, p.6).
57 insight requirements

7 categories

<table>
<thead>
<tr>
<th>Threat, Risk &amp; Harm Identification</th>
<th>Demand Management</th>
<th>Interventions</th>
<th>Vulnerability Identification</th>
<th>Organisation</th>
<th>Crime Identification</th>
<th>Advanced Visualisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>9</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Threat, Risk &amp; Harm Identification</td>
<td>Demand Management and Effective Resourcing</td>
<td>Early Intervention</td>
<td>Victim &amp; Pre-Victim Identification</td>
<td>Organisational Health</td>
<td>Crime &amp; Pre-Crime Identification</td>
<td>Insight Generation</td>
</tr>
<tr>
<td>Staffing</td>
<td>Effective Policing Interventions</td>
<td>Offender Identification</td>
<td>Staff Welfare</td>
<td>Emerging Crime</td>
<td>Geo-mapping</td>
<td></td>
</tr>
</tbody>
</table>