Ethics Committee Briefing Note

Project Reference: National Data Analytics Solution – Modern Slavery (MS)

Purpose of data analysis:

The strategy document 'Policing Vision 2025' outlines the need for technology to be central to how law enforcement operates, calling on forces to embrace innovation so that policing can adapt to new threats and opportunities posed by the 21st century.

The National Data Analytics Solution (NDAS) aims to become a centralised advanced analytics capability for UK policing. UK police forces have access to a vast amount of digital data, but arguably lack the technological capability to use it effectively.[1] By proving that advanced analytical methods can be applied to existing law enforcement datasets, it is hoped that actionable insights grounded in data could be used to guide local intervention efforts and support the cross-cutting outcomes that evolved from the reform strands within the Policing Vision 2025. Putting information at the heart of decision-making in policing by connecting existing datasets for new insights should inform risk assessment and resource prioritisation.

NDAS demonstrated this capability during its Foundation Phase, which ran from September to April 2019. Three high-priority use cases were run as a proof of concept: Most Serious Violence, Workforce Wellbeing and Modern Slavery. Two of these use cases, Most Serious Violence and Modern Slavery, are nearly ready for operationalisation

The founding partners of the NDAS are: West Midlands Police; Warwickshire Police; West Mercia Police; West Yorkshire Police; Greater Manchester Police; Merseyside Police, the Metropolitan Police Service; the National Crime Agency; and Staffordshire Police.

This submission to the WMP Ethics Committee concerns the Modern Slavery (MS) use case, which looks to create a network-driven insights platform that aids the identification of events and people related to Modern Slavery, utilising advanced analytics. The MS use case relies on data from West Midlands Police and West Yorkshire Police—the participation of Warwickshire Police and West Mercia Police during this phase is to be confirmed.

Source of analytical question / hypotheses to be examined:

Through the Foundation Phase, NDAS worked with a range of Modern Slavery SMEs to identify a key operational challenge in managing the complexity of Modern Slavery, and effectively identifying cases. Due to the hidden nature of the crime and with a widely-held belief that the scale of the problem is greater than currently reported, the Home Office supported NDAS in developing an exploratory Proof of Concept to assess how advanced analytics could develop actionable insights into Modern Slavery.

The joint effort between the MS SMEs and NDAS team resulted in the development of the MS capability, which is being refined and intended to be operationalised during this phase of work.

Data to be used:

- Intelligence information – 2010 onwards
- Crimes information – 2010 onward
- People information across Crimes & Intelligence

[1] A Babuta, Big Data & Policing https://rusi.org/sites/default/files/201709_rusi_big_data_and_policing_babuta_web.pdf

Level of analysis

☐ **Individual**

Individuals aggregated?

☐ **Yes**

☐ **No**

☐ **Specific Area:**

☐ **Output Areas**

☐ **Super Output Areas - Lower**

☐ **Super Output Areas - Mid**

☐ **Wards**

☐ **Districts**

☐ **West Midlands, West Yorkshire**

Reliability of data:

The data is sourced from core systems used daily by the West Midlands Police and West Yorkshire Police systems. Once the data has been extracted it is reviewed and verified by the NDAS team and any obvious data quality issues are addressed with the guidance of data subject matter experts (SMEs).

Systems that support the analysis are used by the forces on a day-to-day basis. This use case has sourced data from Crimes and Intelligence systems, although trialled other source systems during the Foundation Phase. Trialled source systems included; Missing Persons, Custody, Command & Control and Stop & Search however are not in-scope for the MS use case during this phase.

Sample or entirety:

Entirety

If sample:

N/A

Method of sampling:

N/A

Method of choosing sample size:

N/A

Sample size:

N/A

Type of analysis:

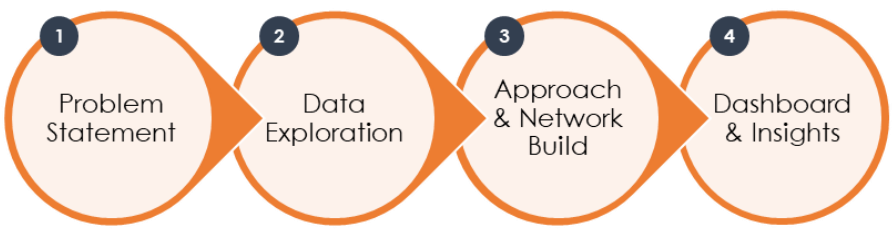
☐ **Exploratory**

☐ **Explanatory**

☐ **Predictive**

☐ **Optimisation**

Proposed methodology:



1. **Problem Statement**

The first stage of this process is defining the problem statement. The problem statement takes the initial guidance from the Partner Forces engaged with NDAS to assess how advanced analytics could develop actionable insights on Modern Slavery, aiding the identification process, and assist in refining what is to be investigated. It has been decided that an exploratory analytical approach would be the most suitable for this use case. The lack of historical data and limited understanding of the challenge as it stands ruled out a problem statement alluding to predictive modelling.

To define the problem statement for this use case, time has been spent with several Modern Slavery Subject Matter Experts (SMEs), spanning both operational and strategic roles. Using a combination of the in-depth knowledge and experience from the SMEs, with the NDAS team's understanding of possible analytical approaches to assisting with current challenges, the most appropriate problem statement has been defined.

From the explorative work carried out during the Foundation Phase, the problem statement was further refined for the Acceleration Phase to ensure it would be fit for purpose for operational usage.

The problem statement:
Creating a network-driven insights platform that aids the identification of events and people related to Modern Slavery, utilising advanced analytics.

2. **Data Exploration**

Data exploration happens alongside defining the problem statement. This involves investigating hundreds of tables across the key data sources to find data fields that help to turn the problem statement from a business problem into a technical solution.

For this use case, this process is driven by initially understanding what key data points are important to MS operational and strategic investigators and then identifying the data sources for these data points and building the logic to create them. In some instances, these data points could be taken directly from source tables and in other cases some derivation is required to transform the raw data into logic-driven "Risk Factors", as identified by MS SMEs. The NDAS team also investigates all tables across the key data sources to find other information that could be useful for the use case. Once data exploration is complete, the raw data is transformed to create analytical tables in preparation for analytical modelling (e.g. NLP and network analytics). The process of data exploration assisted in identifying the initial source systems that would deliver the most value within the Acceleration Phase, and those that could be incorporated at a later stage.

3. Approach & Network Build

Approach

Natural Language Processing (NLP), specifically Document Classification, is performed on free text fields that capture key information for Intel and Crime events. Through data exploration these fields have been highlighted as worthy fields to assist in the identification of Modern Slavery. The hybrid approach of both NLP models and SME inter-lock allow the Document Classification process to identify Modern Slavery events based on terminology used within the free text fields.

The process formed is an iteration of the common 'scanning' approach deployed by many forces, where key words or phrases are used to identify events of interest dependant on the topic of interest. The difference is that the NDAS Document Classification process formed utilises known cases of Modern Slavery to highlight relevant phrases and words that should be utilised to scan for cases of Modern Slavery, as opposed to relying on common descriptive phrases, such as "Modern Slavery". Identify new words to scan for, and relationships between words of interest, then inform a revised list of scanning words that a Modern Slavery SME can review

and refine, prior to the scanning occurring in an automated fashion.

As well as scanning for cases of Modern Slavery, the approach also allows for the grouping of Modern Slavery events into relevant categories – largely supported by learnings from both the Home Office Modern Slavery Typology and the groupings applied by the National Referral Mechanism Reports. This structured information further empowers the ability to effectively manage Modern Slavery through the relevant streams, by quickly identifying the type of modern Slavery crime that could be occurring. The effectiveness and efficiencies experienced by this approach further support the potential use of the tool.

There are often many reasons why a Crime or Intelligence report, which relates to MS, would not be tagged accordingly. An example would be a case tagged as Public Protection, where the associated long text describes the cases of a young female in a vulnerable state who is believed to be groomed, potentially for sexual exploitation. In the first instance, an officer may highlight this as the need for Public Protection, which would mean that the relevant Modern Slavery teams would perhaps not be initially notified of this Modern Slavery event. Through the Document Classification process, the NDAS team would be able to identify that key phrases mentioned within the free text field are indicative of Modern Slavery (inferred from correctly tagged cases), such as vulnerable-groom-sexual-exploitation. Following SME review, these words would be added to the Document Classification process, and when collectively used as stated above, would allow for the process to tag the event as a potential case of Modern Slavery, and more specifically Sexual Exploitation. Ensuring the relevant team was informed within a timely manner of the event being logged. =

Network Build

One of the biggest factors in effectively understanding, identifying, and investigating criminal activity is understanding the connections between people. A key insight from engagement with MS operational case officers is that MS is viewed as a network-based problem, where MS offenders or victims are often part of a wider network of individuals (e.g. a human trafficking network).

A Network-based approach is therefore an important part of the problem statement definition. The NDAS team for the MS use case is currently developing:

- An interactive network of connected individuals to aid the identification of Modern Slavery networks
- Calculate relevant Network Metrics i.e. how an individual is connected to other known victims and offenders, to assist in the formation of actionable insight

Through both the police tagging of Modern Slavery events, as well as the tagging via the Document Classification process, the capability will be able to present a coherent network of people involved with Modern Slavery. With the visualisation and relevant network metrics, the user will be empowered to review and act upon relevant information within the Modern Slavery domain and use this information to target networks and safeguard victims.

4. **Dashboard & Insights**

The final stage of the analytics process is to present the data in a way that informs and assists the end user in identifying Modern Slavery and taking the suitable action. This encompasses not only the network visualisation and metrics as discussed, but the relevant contextual information related to the events and associated people. This information is presented at both low and high levels of granularity, allowing the tool to serve two audiences (these were identified in consultation with MS SMEs). The low-level granularity enables an operational team to follow-up on identified cases through existing processes. The high-level aggregated level of granularity allows for management teams to utilise the tool to inform strategic decision making, with improved understanding of Modern Slavery through trends & statistics.

Will the project eventually be automated?

☒ **Yes**
☐ **No**

The MS analytical models and approach to form the word set will require input from Modern Slavery SMEs to confirm the phrases & words to scan for via the Document Classification process. The NDAS will supply insights to force intelligence so that networks of modern slavery, victims and offenders may be understood more clearly; for organised crime groups (OCGs) to be scored against OCGs involved in other serious crime; and for police and partner resources to be tasked with targeting MS offenders and safeguarding victims.

Means of evaluation:

Internal Technical Evaluation

The problem statement, methodology and Proof of Concept model results were reviewed as part of an internal evaluation to provide recommendations on how to improve the validity and accuracy of the analytical model. The reviewer was not involved with the NDAS project, and internal evaluations such as this are conducted as part of general quality assurance activities by the NDAS' delivery partner.

These were the recommendations from the internal evaluation to improve the data quality of the tool:

- Creating a labelled data set to validate MS versus non-MS events that is currently being determined through key-word matching
- Potentially using machine learning instead of key-word matching to classify modern slavery events

The NDAS team are working through the above recommendations and another technical evaluation will be planned in the next few weeks. Our own methodology/approach will consist of regular model monitoring and capturing/reviewing the results that are generated as part of the output.

Independent Evaluation

In addition, it is critical that the models and output developed by the NDAS are independently evaluated before moving into large-scale deployment as planned after this phase of work. The NDAS will partner with Anjali Mazumder of the Turing Institute, who specialises in data science decision-making for evidence-based modern slavery and human rights policy, to provide an independent evaluation of both the MS analytical model and its outputs.

ALGO-CARE considerations:

Advisory

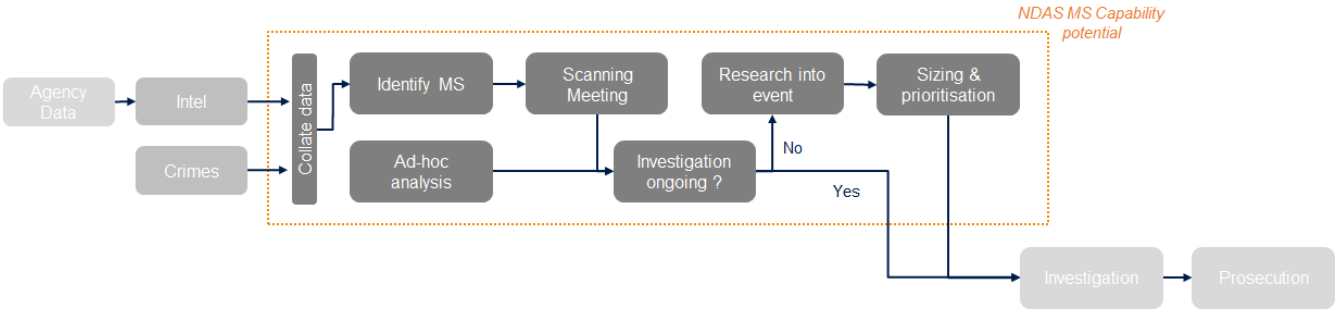
If applicable, are the outputs from the algorithm to be used in an advisory capacity?

The output of the MS analytical models will be used as a source of information to assist police intelligence in the understanding and identification of Modern Slavery and associated insight. The NDAS capability is not a tool that substitutes the professional judgment and discretion of law enforcement practitioners for automated decision-making. It is designed to assist human decision making, with added efficiencies through automated processing.

Does a human officer retain decision-making discretion?

Yes. The MS output dashboards will only be used to supplement existing processes designed to target organised crime and safeguard victims of Modern Slavery.

The following is a high-level overview of how Modern Slavery will assist in the identification of cases, prior to an investigation and prosecution taking place.



It is anticipated that the interactive dashboard which encompasses outputs from this use case will support the overall process by identifying connections between information already held within police systems. The NDAS team is exploring 2 possible audiences for this output: force intelligence and MS SMEs (operational staff within SOCEX, focused on understanding details surrounding an MS case and all known connections across data systems); and strategic analysts (focused on understanding high level statistics and trends relating to MS).

Lawful

What is the policing purpose justifying the use of the algorithm (means and ends)?

The purpose is to better identify and understand how networks of people are connected through modern slavery events, enabling better understanding of the scale of such networks to improve intelligence, guide investigations, and inform the delivery of possible interventions for potential victims and perpetrators. The MS use case demonstrates the ability to see networks of modern slavery in an innovative, interactive manner at an unprecedented scale, which would provide a simplified view of an otherwise very complex crime, where it is difficult and time-consuming to identify links between multiple nominals and demonstrate this on a network level.

Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes?

There is minimal impact to the privacy of individuals as the MS use case relies on data that is already held on police systems and identifies connections between said data to visualise network information; details of nominals within the networks; nominal trends and statistics; and event trends and statistics.

In what way will the tool improve the current system and is this demonstrable?

During the Foundation Phase for this use case it quickly became evident, from engagements with SMEs, that all the data required to understand, identify, and tackle modern slavery already existed within police source systems. However, the time required to find and understand the information and form the necessary connections to enable actionable insight can be extremely time-consuming, and in certain cases not possible.

During our engagement sessions with a number of groups, the National Referral Mechanism (NRM) was discounted as a data source for the MS use case during the Foundation Phase given that once an NRM is submitted it is often too late for impactful preventative measures. Therefore, the focus of the use case adapted to understanding how data could be joined from across multiple source systems, to enable improved insight into understanding, identifying, and investigating modern slavery. WMP had several tools available for this such as Corvus and WMP's Search platform—both tools enable the searching of information across source systems. However, results were siloed and did not enable insight specifically around Modern Slavery and did not present a holistic view of the information that exists within police data.

It was also evident that the MS team at WMP had difficulty in easily highlighting and monitoring cases of MS within the police systems. This is particularly the case when MS events had not been correctly labelled with a Home Office Crime Reporting Code or given the correct Intel Subject header—a common thread across engaged Partner Forces.

As was seen both at WMP and other partner forces that we worked with, the use of keyword scanning to highlight potential events related to MS within the police source systems is a common process. This would involve an analyst manually searching for key words stored within systems, such as 'Human Trafficking', 'Modern Slavery', 'Exploit' and so on. This is a lengthy process, both to undertake the search, but also to review and prioritise the results. Depending on the timeframe, such keywords could return a list in the hundreds, and possibly thousands if events were picked up out of context – e.g. 'traffic' being a common word associated with traffic offences, and the string 'traffic' existing within the search term 'Human Trafficking'.

The process of scanning for key words would likely be reviewed during a Scanning meeting – where prior to the meeting, the results of a key word search would be captured to then be reviewed as a team. While this provided the team with a well-rounded understanding of current MS activity over the prior scanning period (a week for WMP), it was often a meeting that focused on filtering and sorting, rather than actioning and preventing. A fair portion of the content returned by the scan was already known to the MS team and in some cases already actioned. Once these events have been identified, it is then a process of understanding who will action the Intel (as Intel is normally the system being scanned) and prioritising by risk. The latter two steps often used the experience of the SMEs to assist in judgement, as opposed to a systematic process. Scanning via this manual approach is time-consuming and inefficient. The NDAS MS capability will improve the approach through systemisation and process consistency.

Once the SMEs had manually located the MS information of interest, they needed to move onto forming an improved understanding of an event or nominal. Again, this included manual searches across source systems, forming a network to connect learnings as investigations developed. This could be done in systems like IBM i2 iBase/Analyst Notebook (used in both WMP and WYP), however these tools relied heavily on manual input.

The MS use case has the potential to mitigate the following challenges throughout the MS identification and investigation process:

- Improving data accessibility to end users
- Capturing incorrectly tagged Modern Slavery events
- Allowing insight on large sets of unstructured data
- Reduce the time required to process information
- Improve the ability to present actionable insight

To date, building the Proof of Concept solution for the Modern Slavery use case has led to several early-stage successes:

- Utilising the network metrics, the NDAS team were able to identify that, for the MS nominals that WMP are aware of, on average, an MS nominal is connected to 54 other nominals. This provided initial insight into the type of networks formed related to this crime
- Through scanning a subset of the Intelligence system, the use case proved the ability to highlight and identify individuals and events which would not have led to significant, coordinated operational activity if not for the application of the model
- For the Foundation Phase, 6 source systems were transformed into just three base tables to enable the efficiency gains and analysis to take place. This consolidation of different data sources reduced insight generation for an MS analyst from 200 hours to only a few minutes (200 hours has been quoted by an MS analyst as an example of how much time is ordinarily required to collate and form an MS report). In this phase of work we are considering 2 source systems.

Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim?

All data used by the NDAS is derived from existing police systems, meaning the data was obtained and processed for criminal law enforcement purposes under Section 3 of the Data Protection Act 2018. In addition, the acquisition, processing, and retention of data by the NDAS on behalf of West Midlands Police is governed by an information sharing agreement (ISA) between partner agencies. 4 partner agencies signed the ISA during the Foundation Phase. These include West Yorkshire Police, Warwickshire Police and West Mercia Police.

The ISA stipulates:

'Each national analytics assignment commissioned through NDAS governance will look to answer a specific problem (or "use case") on behalf of the Partner Forces, in line with one or more of the following policing purposes:

- Protecting life and property
- Preserving order
- Preventing the commission of offences
- Bringing offenders to justice, or
- Any duty or responsibility of the police arising from common or statute law.'

In this way, all data sources will be shared for a common, lawful and specified purpose.

In accordance with the Information Commissioner's guidelines, a full Data Protection Impact assessment was conducted for both the Foundation and Mobilisation phase. It is attached in the appendix to this submission. A legal review from the WMP legal team will follow in due course.

Is the operation of the tool compliant with national guidance?

The Modern Slavery Police Transformation Unit currently have a database of ongoing MS investigations. This is designed to allow a national view of the offending and law enforcement response. The use of data analytics is seen as an improvement in the work that they are doing as it will enable forces to understand the landscape of MS more comprehensively.

The NDAS is part of the Home Office's Digital Policing Portfolio, which has articulated a national aim for a data analytics capability for UK law enforcement. In line with this aim, recommendations on national guidance establishing minimum standards on how data analytics platforms should be developed and used by law enforcement need to be produced. Considering its important role in shaping what these recommendations should look like, the NDAS welcomes this endeavour and will participate fully. In the interim, the NDAS will consider the independent research—commissioned by the CDEI and conducted by the RUSI—into the potential for bias to occur in predictive analytics technologies being developed by police forces.[2] The NDAS has been asked to participate in the Centre for Data Ethics and Innovation's effort to develop a code of practice for the trialling of predictive analytics technology in policing. [3] This builds on previous engagement undertaken by the NDAS with a range of stakeholders in navigating this landscape, including the Home Office Biometrics Commission; the National Police Chiefs' Council; the Royal United Services Institute; the National Law Enforcement Data Programme; and the National Policing Information Risk Management Team.

[2] Centre for Data Ethics and Innovation, https://www.gov.uk/government/publications/report-commissioned-by-cdei-calls-for-measures-to-address-bias-in-police-use-of-data-analytics

[3] Centre for Data Ethics and Innovation Work Programme 2019 – 2020, https://www.gov.uk/government/publications/the-centre-for-data-ethics-and-innovation-cdei-2019-20-work-programme

In the absence of a framework regulating analytics in law enforcement, the NDAS has looked to ensure that its general operation remains aligned to the relevant existing national guidance that applies to law enforcement, particularly with regard to relevant data protection and administrative laws. Building on our completion of a Data Privacy Impact Assessment, NDAS is engaged in the Office for the Information Commissioner's Project DALE (Data Analytics in Law Enforcement) and is committed to continuing to ensure that all operations adhere fully with general data protection requirements for law enforcement.

Granularity

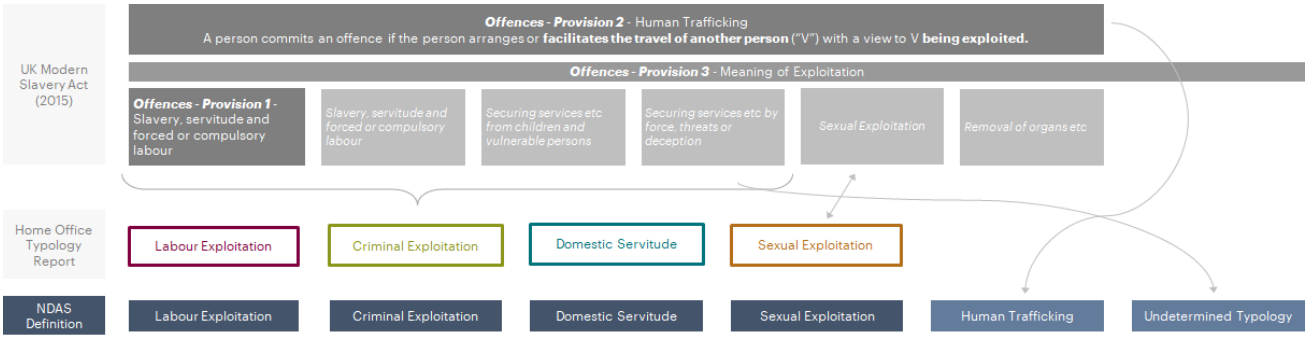
Does the algorithm make suggestions at a sufficient level of detail given its purpose and the nature of the data processed?

The MS analytical model provides detail at an individual, network level and aggregated view. This is considered sufficient to understand, identify, and support investigations of modern slavery based on the use case's problem statement, as validated through feedback with MS SMEs.

Are data categorised to avoid broad-brush grouping and results and therefore issues of potential bias?

The following diagram describes how MS categories have been defined to support our analysis. Aligning with both the Home Office Typology of Modern Slavery and existing groupings within the National Referral Mechanism reports (reflecting the Modern Slavery Act 2015). The NDAS Modern Slavery Categories formed will allowing the grouping of events and infer the type of MS a related person (to an event/s) may be involved with.

Note: an event or person could have multiple flags for the following categories, depending on the tagging applied in the police data and output of the document classification. The undetermined category is used to capture cases of Modern Slavery, where the form of exploitation is unknown or not clear, e.g. 'Modern Slavery'.



Do the potential benefits outweigh any data quality uncertainties or gaps?

Yes. Firstly, no significant data quality issues or gaps were identified. Moreover, there is the social benefit of being able to identify networks of modern slavery within the existing data, which could lead to more efficient safeguarding of victims.

Is the provenance and quality of the data sufficiently sound?

The data required comes from core source systems used on a day to day basis by forces and do not show evidence of significant data quality issues.

If applicable, how often are the data to be refreshed?

Data will be refreshed daily, but the frequency of dashboard refresh will be confirmed in due course, in partnership with the forces' needs (in-line with existing scanning timelines).

If the tool takes a precautionary approach in setting trade-offs, what are the justifications for the approach taken?

Not applicable. The model does not set accuracy trade-offs. However, it is arguable that there is a net positive result of identifying and disbanding an OCG network, with the aim of long-term public protection and restricting its ability to exploit victims. In addition, where there are unidentified offenders, the model could improve the policing ability to target the highest-risk OCGs in this field. Police resources are not infinite, and it is not expected that every single insight identified can be managed and actioned.

Ownership

Who owns the algorithm and the data analysed?
WMP owns all models developed as part of the National Data Analytics Solution, on behalf of the Home Office. Each partner force owns their own data being analysed and the insights derived.

Does WMP need rights to access, use and amend the source code and data?
No.

Are there any contractual or other restrictions which might limit accountability or evaluation?
No.

How is the operation of the algorithm kept secure?
Partner Forces will be given a choice of Secure File Transfer Protocol (SFTP) or a Police Encrypted Device as a data transfer mechanism. Once received data will be uploaded to a WMP on-premise staging environment and then transferred to the secure NDAS AWS cloud, where all data will be held and processed throughout the delivery of the use cases. More detail on the security specifications can be found in the attached Data Protection Impact Assessment. This cloud solution will also go through a penetration test (estimated date November 2019).

Challengeable

What are the post-implementation oversight and audit mechanisms, e.g. to identify any bias?
A range of governance mechanisms are in place to monitor the progress of the NDAS project as it moves toward, and beyond, its first operationalisation phase:

- Strategic oversight is provided by the Home Office, as the NDAS is part of its Digital Policing Portfolio
- Tactical oversight is currently provided by the NDAS Stakeholder Governance Group, which includes the Chair of the West Midlands Ethics Committee
- Funding for the project is overseen by the West Midlands Office of the Police and Crime Commissioner
- The NDAS is engaged with the development of an ethical framework for the use of algorithmic systems in law enforcement, overseen by the CDEI
- The NDAS is engaged with the Office for the Information Commissioner on developing guidance over the use of data analytics in law enforcement

In addition, algorithmic systems deployed in this context need close supervision using ongoing human monitoring and auditing of performance against metrics such as accountability, bias, and security. As part of ongoing model maintenance and monitoring, the NDAS team will conduct regular reviews of the model to monitor against bias, including providing a statistical analysis of the impact of errors in the output (e.g. false positives and false negatives). The NDAS will work with the end users — MS operational investigators and strategic analysts—to support the development of a decision-making oversight and audit mechanism.

If the algorithm is to inform criminal justice disposals, how are individuals notified of its use?
The insights generated by the MS analytical model will not be used to inform criminal justice disposals,

including decisions on charge or bail; or decisions as to whether to continue an investigation into allegations concerning a subject. There will, however, be a process for notification, challenge and complaint within existing police systems.

In addition to notification, external observers and data subjects shall be able to challenge the process by which an outcome was reached, to 'ensure that such tools are being used in accordance with the requirements of the relevant data protection legislation and principles of accessibility and natural justice under the Human Rights Act 1998'.[4] In line with this aim, NDAS governance forums should work with partner forces to embed standards for fairness, accountability, and transparency not just in the analytical models developed but also in the overall decision-making process that uses the outputs generated. For example, if a data subject wishes to challenge a decision that has been made with NDAS output (providing supplementary information to the decision-maker), a process will be established to allow the subject to scrutinise the model outputs.

Accuracy

Does the specification of the algorithm match the policing aim and decision policy?
Yes. The MS use case was developed directly as a response to the problem statement which was defined in collaboration with partner forces.

Can the accuracy of the algorithm be validated periodically?
As the product in question is a discovery tool to support analysis of officers, accuracy scores are not calculated.

Can the percentage of false positives/negatives be justified?
Not applicable (see previous response).

How was the method chosen as opposed to other available methods?
The text search method was chosen based on discussions with MS SMEs who have identified which words they generally use to identify the category of 'modern slavery' in their reports. Natural language processing methods have been used to build on the list of words that are used and finds similar words. This method has higher explainability value over a multi-class classification model. Moreover, certain modern slavery categories had a very low sample size, which would have made it difficult to train the model with high enough accuracy in those categories.

What are the (potential) consequences of inaccurate forecasts?
As stated above, the output generated during the MS use case will be provided as a supplementary source of information for MS operational officers and strategic analysts with their work, as opposed to stating an automated recommendation for them. In practice this means that an inaccurate forecast would have minimal impact, as the officer looking into the insights would be advised to use the solution and its insights as just one piece of information in the wider context of additional information, in combination with professional judgement. The output of the model alone should not be used to drive direct action.

Does this represent an acceptable risk?
Yes. A plain-language explanation of how the MS output was generated and the factors that influenced the output will be produced alongside the output itself. In addition, it is intended that police end users review and interpret the model's results—in combination with the associated key predictive indicators—to complement other sources of information in order to develop a targeted, well-informed interventions approach.

[4] Alexander Babuta, Marion Oswald and Christine Rinik, 'Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges' (2018) https://rusi.org/sites/default/files/201809_whr_3-18_machine_learning_algorithms.pdf.pdf

How are the results checked for accuracy and how is historic accuracy fed back into the algorithm for the future?
Not applicable as accuracy scores are not calculated.

How would inaccurate or out-of-date data affect the result?
Not applicable as accuracy scores are not calculated.

Responsible

Would the operation of the algorithm be considered fair?
One of the risks of deploying predictive analytics technology is the amplification and reinforcement of existing human biases. Some of these biases are unintended and arise from a lack of diverse perspectives when developing and training an analytical model. In other cases, decision-making can be skewed by reliance on incomplete data where other relevant factors are omitted. Finally, the historical data on which analytical models are trained may be biased. An analytical model merely manipulates data in order to produce an outcome. If the source data itself is biased, the model will inadvertently reflect biases inherent in the dataset.

In a law enforcement context, the perpetuation of these biases can be especially pernicious: for example, there has been much discussion over the potential for indirect racial bias to creep into analytical models used in law enforcement. It is well-documented that postcode information can function as a proxy variable for ethnicity or social deprivation, resulting in algorithmic outcomes that perpetuate bias.

Recognising these risks, ethnicity and gender variables were not used as inputs to the MS analytical model. Similarly, location variables were also omitted to avoid this information becoming a proxy for other factors, helping to mitigate any undue bias that may arise in the analytical model. The NDAS team is monitoring any bias in the data that might affect the model output.

Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review?
As part of our continuous engagement with MS SMEs, including operational and strategic decision-makers, we will support accountability in the decision-making process by helping decision-makers fully understand the output generated by the MS use case. We will continue to work with SMEs to define what this process and associated standards will look like. It is also important that the use of the MS dashboards is periodically reviewed if it is deployed into an operational environment—as part of regular data and dashboard refresh processes, a reporting system that flags up any potential issues will be embedded.

<u>Public engagement for transparency and accountability</u>
Ensuring there is public trust in policing is paramount. The risks of damaging public confidence and trust in law enforcement are manifold, and the consequences well-documented. The public may be less likely to support the police in the pursuit of crime prevention, their willingness to participate as witnesses or to come forward as victims weakened. The application of advanced analytics adds complexity to this: although such technologies have been broadly applied in the private sector, it is still relatively new within policing—however, in both cases the level of public discourse is nascent. Compounding this, the opacity of how personal data is collected and used has understandably raised surveillance and privacy concerns—in addition, the pursuit of new technological initiatives without public consultation has arguably led to diminishing public trust in technology.

Despite these challenges, there is an opportunity for the NDAS to engage citizens to influence its operation and build local accountability in developing ethical approaches to the use of analytics in law enforcement. A plan for meaningful public engagement should play a role in this. The Royal Society of the Arts (RSA)[5] suggests the application of 'a process of citizen deliberation' in the deployment of analytics across three phases:

1. Public scrutiny through consultation when such systems are being introduced
2. Technical oversight through testing predictions for accuracy or expert-led auditing; and
3. Monitoring how the system is used by humans and evaluating it for accuracy

We suggest that point 1, public scrutiny through consultation, be delivered through the Office of the Police and Crime Commissioner (at this stage, by the West Midlands OPCC). It is recognised that it takes real resource and commitment to deliver a plan for public engagement on this topic, and that consideration must be made towards being as representative of local citizens as far as possible to build local-level democratic accountability.

Would it be considered to be used in the public interest and to be ethical?

Yes. If it is demonstrable that the MS use case has the potential to rapidly identify information (that already exists, but would take longer to analyse manually) to support an investigation of modern slavery, it is arguably within the public interest for it to be trialled, in a way that is proportionate to any potential impact on individual rights.

It is appreciated that in identifying victims of modern slavery, the deployment of this use case may lead law enforcement to locate victims who do not wish to be found, including victims who will be liable for deportation and victims whose opportunities in their home country are even less than those offered to them by traffickers. We appreciate that the ability of the police to safeguard victims once found is reflective of existing national structures and policies and will vary depending on the particular context and experience of individual victims. The following factors should be taken into consideration:

1. A significant proportion of victims of trafficking identified in the UK are UK nationals – children from the UK are most likely to be trafficked, and adults the third most likely (2018 Annual Report on Modern Slavery, Home Office).
2. The second highest contributor to modern slavery figures is Albania (2018 Annual Report on Modern Slavery, Home Office), and the home Office is investing considerable resources to working with the Albanian government to make the community of Albania more resilient to modern slavery and human trafficking.
3. Recent changes to the National Referral Mechanism (NRM) mean that the support available to victims is bespoke to that victim, rather than a standard 45 days. This support includes the allocation of a support worker and access to legal support to fight any deportation proceedings. Where there is an investigation into a trafficking Organised Crime Group (OCG), UK Visas and Immigration are approachable to requests by investigators to keeping victims in the country to support the investigation.
4. There are tens of thousands of victims of modern slavery in the UK. Many are victims of serious sexual offences. Many exist under fear of serious violence or death, and many are the victims of serious violent offending. The risks of police activity adversely affecting a minority of these victims is more than offset by the opportunity to promote long-term public protection by dismantling the OCGs involved.
5. The avenue for operationalising this use case will be via force intelligence departments, to give them the ability to prioritise these OCGs amongst OCGs involved in other serious criminality. This prioritisation of OCGs means that activity will be directed towards the those that present the most physical risk to victims and communities, thereby ensuring that this is not just a mechanism to identify victims but rather a mechanism to understand and systematically dismantle the networks involved.

[5] RSA, 'Artificial Intelligence: Real Public Engagement' https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf

Explainable

Is information available about the algorithm / decision-making rules and the impact of each feature?
As part of our continuous engagement with MS SMEs, including operational and strategic decision-makers, we will work towards embedding overarching standards of algorithmic transparency and intelligibility throughout the analytical process so that we can support informed decision-making.

For the Foundation Phase, and for this Acceleration Phase, a description of all words and phrases is captured at an event level to detail why a piece of free text was deemed to be related to Modern Slavery.
For example, 'modern slavery' existed as a phrase within a long text description, causing the undetermined category flagged to be marked as '1', when reading the long text field, the user would be able to read how the mention of 'modern slavery' comes into play and then review all supporting contextual information. For an individual, it would be possible to observe all of the events related to the related individual, and why those events have been tagged as Modern Slavery, and therefore how the model's inference that the person is related to Modern Slavery has been reached (via the person-event relationship within the policing systems).