National Data Analytics Solution Submission to WMP Ethics Committee, March 2020

Proposed Approach to develop an Ethical Framework

This document provides an overview of our approach to developing an ethical framework for the National Data Analytics Solution (NDAS) project.

The need for an ethical framework

Our previous submission to the WMP Ethics Committee in November 2019 highlighted the need for national guidance and standards to govern how data analytics platforms are developed by law enforcement and used in practice. This was reflected in our response to the following question:

> *"Is the operation of the tool compliant with national guidance?*
>
> *The NDAS' s role is to fulfil the National Police Chief Council's strategic ambition for an advanced 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.[1] 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. [2] 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.*
>
> *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.*

---

[1] 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
[2] 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

Work is underway across the data ethics policy landscape to define national guidelines and minimum standards for how analytics should be used in the public sector, including law enforcement. Guidance for the UK public sector, while piecemeal and lacking a coordinated, regulatory-driven approach, is extensive and useful: for example, the Department for Digital, Culture, Media and Sport (DCMS) is refreshing its data ethics framework in 2020; the Information Commissioner's Office have released multiple blogs and guidance covering the explainability of AI-based decisions and launched a consultation on their draft AI auditing framework; and the Alan Turing Institute have developed a guide for the responsible implementation of AI in the public sector. This list is by no means exhaustive, as international governments and regulatory bodies have also begun to develop new ethical standards for AI. Law enforcement—as an area where the ethical and legal risk of AI systems is high and well-documented—is catching up with the academic and policy landscape: the NPCC and APCC's National Policing Digital Strategy has proposed a new National Data Ethics Governance Model; and the Centre for Data Ethics and Innovation is working towards a draft framework for the ethical development and use of data analytics tools in policing.

In November 2019, we reiterated our desire to engage with a range of stakeholders across the area of data ethics; to participate in efforts to develop national standards on ethical applications of AI in policing; and to ensure that our operations remain aligned to general data protection regulation and administrative laws that currently apply to policing in England and Wales. While data ethics policy and academic discourse progresses, and fresh resources are developed to shape the use of advanced analytics in the public sector, a newly-published report by RUSI and the CDEI reveals 'widespread concern across the UK law enforcement community regarding the lack of official national guidance for the use of algorithms in policing, with respondents suggesting that this gap should be addressed as a matter of urgency.'[3]

We echo this sentiment—as a project seeking to evolve into a national capability for advanced analytics in policing,[4] we believe it is crucial for us to develop a project-based framework that grounds principles of data ethics within the organisational and governance standards established in policing culture. Alongside our desire to build transparent systems and processes, and to maintain accountability in how these systems are used in practice by police practitioners, we consider this effort to be the beginning of a coordinated approach towards the long-term sustainability of the use od advanced data analytics in policing, which necessitates the responsible development, deployment, and use of data-driven technology within the project and with our partner forces. Our stated desire to work with stakeholders in data ethics and participate in the development of national standards is unchanged but has evolved into a more proactive effort. We have begun this work and invite the Ethics Committee and other stakeholders to have input into the process.
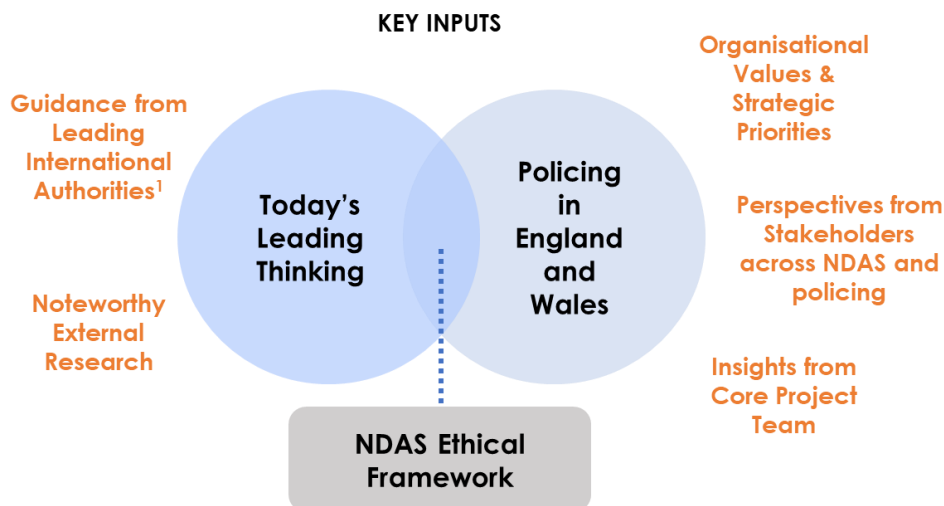
Our approach and methodology

Law enforcement as a sector is unique in that although its use of advanced analytics to inform decision-making and interventions can come with high legal and ethical risks, we believe it has an established culture of governance and accountability that lends itself well to data ethics, subject to some augmentation.

---

[3] RUSI report Feb 2020

[4] A Centre for Data Analytics has been proposed by the NPCC, with a view that the NDAS will develop the capability that evolves into the CDA for policing.

Our approach to this exercise is twofold, building on both principles and processes.

To build our principles, we are conducting a literature review of relevant academic discourse and leading UK and international guidance on data ethics. This review will inform a list of considerations for advanced analytics more broadly and will be used to augment a framework that relies primarily on organisational directives in policing.

KEY INPUTS

Guidance from Leading International Authorities[1]

Noteworthy External Research

Today's Leading Thinking

Policing in England and Wales

Organisational Values & Strategic Priorities

Perspectives from Stakeholders across NDAS and policing

Insights from Core Project Team

NDAS Ethical Framework

With this in mind, we are relying on the following organisational directives in policing as a starting point for the NDAS ethical framework:

- The College of Policing's Code of Ethics (built on the Seven Principles of Public Life)[5]
- The National Decision Model
- Participating forces' vision and core values

The core principles enshrined in the above directives will be mapped onto the data ethics considerations explored in the literature review, and any gaps will be augmented. The framework itself will be principles-based with accompanying guidance for our project teams, partners, and police practitioners. We do, however, recognise the need to take practical steps towards embedding guiding practices, policies, and processes to uphold these principles. This is generally accepted in the academic discourse and leading guidance on data ethics, and is best summarised by the Committee for Standards in Public Life:

> "Clear and authoritative ethical principles…need to be further elaborated and specified in codes of conduct that are explicit about what is expected of public office-holders in different contexts. It is likely that sector-specific AI ethics codes will be necessary, particularly in high-risk policy areas such as policing, criminal justice, health and social care. Sector-specific codes can help make abstract ethical principles clearer and more tailored to particular

---

[5] The Committee for Standards in Public Life argue that the Nolan Principles, which are the foundation of the College of Policing's Code of Ethics, are 'strong, relevant, and do not need reformulating for AI systems', particularly because these principles are already relevant in terms of the ethical challenges that public bodies will have to meet when developing and using AI systems. The report is available here: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF

professional settings, while retaining the link to the standards expected of public office-holders across the whole of the public sector. " [6]

Taken together, the ethical framework will guide how the NDAS operates internally and externally, from how use cases are commissioned; how models are built with data ethics in mind; how results are communicated and explained to police practitioners; and how these results might support police decision-making in terms of delivering interventions.

To this end, we are considering the following key resources as we build the processes required to underpin a principles-based framework:

- 'A guide to using artificial intelligence in the public sector',[7] published by the Office for AI, the Government Digital Service (GDS), and the Alan Turing Institute (The Turing)
- 'Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of AI systems in the public sector'[8] by The Turing Public Policy programme (as part of a partnership with the Office for AI and the GDS)
- The Data Ethics Framework[9] launched by the Department for Digital, Culture, Media, and Sport[10]
- Draft guidance on 'Explaining decisions made with AI',[11] co-developed by the Information Commissioner's Office (ICO) and The Turing
- Draft guidance on the AI auditing framework[12], developed by the ICO
- 'Artificial Intelligence: Real Public Engagement'[13] by the Royal Society of the Arts

This list has been kept intentionally narrow as this is a crowded policy landscape—the above resources are not only sufficiently robust but are intended to support the ethical use of advanced analytics in the public sector.

An invitation

We recognise that the exercise of building an ethical framework for an advanced analytics project in policing that is fully grounded in governance and accountability cannot be undertaken alone. This submission to the WMP Ethics Committee is not simply an articulation of our purpose and methodology, but an invitation to contribute expert insight to the development of the NDAS ethical framework (to the extent that the necessary oversight function of the Committee is not conflicted by its participation).

[6] Ibid 33
[7] https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector
[8] https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf
[9] NDAS was part of a working group session in consultation for the relaunch of this framework in February 2020
[10] https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework
[11] https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-and-the-turing-consultation-on-explaining-ai-decisions-guidance/
[12] https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf
[13] https://www.thersa.org/discover/publications-and-articles/reports/artificial-intelligence-real-public-engagement

We are in the process of developing the principles-based framework and accompanying guidance based on the methodology outlined above and are preparing to launch a consultation to all relevant stakeholders in the coming months. The consultation is an opportunity for those with interest and expertise in data ethics and policing to provide their input, and will include partner forces, legal and information assurance representatives, public affairs representatives, and data science practitioners. We welcome a diversity of perspectives in shaping this framework and believe that the Ethics Committee could play a vital role in supporting this work.

Interim position and going forward

In the meantime, we will continue to submit our plans and updates to the Ethics Committee as required. Finally, we are aware that it is not necessarily the role of the NDAS project to dictate an ethical framework for analytics in policing; we will tailor our approach to any new guidance that emerges.

Most Serious Violence Update to WMP Ethics Committee, March 2020

This is an update to the WMP Ethics Committee on the Most Serious Violence use case.

During the operationalisation activities—the impact analysis of refining MSV features based on a new logical data model—for the MSV Proof of Concept (PoC) use case, a coding error was found in the definition of the training dataset which has rendered the current problem statement of MSV unviable. It has proven unfeasible with data currently available, to identify a point if intervention before a person commits their first MSV offence with a gun or knife, with any degree of precision. However, it should be noted that the investment in developing the ability to ingest data from different systems and forces, to develop a data model, and tools to analyse behaviour, networks and natural language processes, as well as visualising the results, have been well worthwhile. The team have begun to explore alternative problem statements (by changing the components of the MSV problem statement) to find an alternative point of intervention in MSV, and the results from this are looking positive. These will be refined over the coming weeks, with input from stakeholders such as the West Midlands Violence Reduction Unit, offender management leads, and Project Guardian.

A plan is being developed to engage various stakeholders to establish an alternative point of intervention which is suitable to help prevent harm to reframe the use case, and continue to refine the data science techniques so that the precision rate of the model is as good as it can be. We invite the Ethics Committee to provide input into this effort.

Modern Slavery Response to WMP Ethics Committee, March 2020

This submission provides a response to general comments and requests for further information by the WMP Ethics Committee on the Modern Slavery use case.

Response to comments

The following comments and requests for further information were made by the committee. Responses to each comment/request is shown below.

"The Committee has not seen any legal advice regarding the model and there is no DPIA specific to this model, meaning the committee cannot advise to progress."

There is a DPIA covering all NDAS Acceleration phase activity, which was provided as part of our submission to the Ethics Committee in November 2019..

NDAS has now received a legal review of the MS use case which is attached in this submission.

"The Committee would like to see the independent evaluation referenced on page 5 of the project proposal."

We are currently exploring options and partners for the commissioning of an independent academic evaluation of the MS model before it is operationalised. The proposal for the evaluation, once confirmed, will be shared with the Committee.

"The Committee asked about the process for engaging the natural language processing method."

There are 3 elements to the natural language processing (NLP) methodology developed for the MS use case, referred to as Document Classification. These are:

- Document Classification: assigning a document (event) into categories, utilising Rule Generation & Rule Application
- Rule Generation: this is the process of forming the Rule Set of primary, secondary, and tertiary words linked to modern slavery, based on known cases of modern slavery. The words defined in the Rule Set were developed through input from modern slavery SMEs and Home Office offence typology[14].
- Rule Application: NDAS then apply the Rule Set to identify cases of modern slavery. This automates the scanning approach deployed by many forces, where keywords or phrases are used to identify events of interest depending on the topic of interest. The difference is that

---

[14] Home Office Typology of modern slavery offences, available at <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652652/typology-modern-slavery-offences-horr93.pdf>

the NDAS Document Classification process, utilises known cases of modern slavery to highlight relevant phrases and words that should be utilised to scan for other cases of modern slavery, as opposed to relying on common descriptive phrases, such as "Modern Slavery". Identifying new words to scan for, and relationships between words of interest, then informs a revised list of scanning words that a modern slavery SME can review and refine.

- Test Set: a proportion of MS and non-MS cases are manually reviewed to assess the effectiveness of the approach.

The diagram below provides an example of the NLP approach, how text is pre-processed to prepare for rule application, and how the rule set is then applied, and findings are captured within the output table.
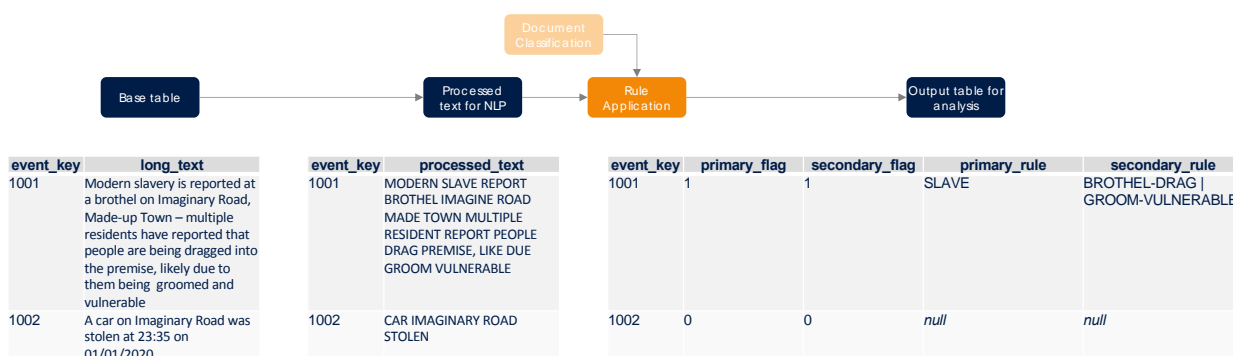
1. The base table holds all relevant information related to an event, including the long text field. In this example we have selected 2 event keys in the base table, 1001 and 1002.
2. The long text field for each event key passes through a number of steps for pre-processing, such as removal of numbers, punctuation, and lemmatization (lemmatization is the process of grouping together the different inflected forms of a word so they can be analysed as a single item, by linking words with similar meaning to a single word).
3. The words and phrases that form the Rule Application are maintained within a Reference File, the Rule Set.
    - This file can be easily updated to accommodate updates required by SMEs, as well as evolve over time in-line with the changing MO of the crime type. The Rule Set is then applied to the processed text by scanning for words and highlighting where there is a match.
    - The below Rule Set example documents the structure of the file, which when operationalised will contain around 50~ words and phrases. As the MS use case continues to be refined ahead of operationalising, the associated Rule Set is also updated.

Primary Word	Secondary Word	Tertiary Word
SLAVE	*null*	*null*
null	BROTHEL	DRAG
null	GROOM	VULNERABLE

4. Should a match occur, the respective words flagged by the rule are recorded in the output table so that they can be traced for review and analysis by an SME. Word flags are marked

from 0 (no hits flagged by the rule) to 1 (the rule flags a hit). In this example, the rule has identified that event key 1001 has hits on a primary flag and secondary flag. The primary rule associated with the primary flag is displayed (this is 'SLAVE') and the secondary rules associated with the secondary flag is also shown (2 rules: 'BROTHEL-DRAG' and 'GROOM-VULNERABLE'). As demonstrated in the table above, event key 1002 does not raise any flags as identified by the rules.

5. The output of this process then allows for these events to be flagged and visualised in the dashboard tool.

The Committee requests additional information about the application of the tool and the benefits of the tool in relation to modern slavery investigations.

Further to the benefits as documented in the last submission, NDAS have gone on to showcase the capability to a wider set of stakeholders and intended end users. This has led to noteworthy feedback with regards to the current capability and potential of the tool. West Yorkshire Police, whom will be one of the first partner forces to gain access, have supported the capability following a recent discussion on integrating the tool into existing processes.

A leading senior barrister, during a visit on the use case, commented on how innovative the MS capability is, calling out the potential savings for a modern slavery operation being in the millions as well as improved insight becoming available through the network visualisation.

Further engagements are planned over the coming weeks to ensure that the MS capability is fit for purpose ahead of being operationalised, so that it can supplement existing investigative processes within each respective force.

Further detail on the use of ethnicity data is needed, as this was not clear in the report.

Ethnicity data are not included within the MS capability, in terms of structured data fields. The model identifies individuals known or believed to be involved in modern slavery through being linked to modern slavery "tagged" events (crime reports or intelligence logs) or through NLP as described above.

Reference to ethnicity does however exist within the free text fields within police source systems, such as within intelligence reports and crime investigation logs. As the Rule Set of primary, secondary, and tertiary words is formed based on known modern slavery cases, it is possible that the document classification process could identify protected characteristics such as ethnicity as a result. However, a review of the document classification output ahead of forming the Rule Set allows for these characteristics to be removed (should it be required).

These references were not removed during the Proof of Concept stage, in part to understand the extent to which ethnicity is referred in police source systems. The Rule Set formed during the PoC stage includes 2 tertiary words with links to ethnicity: 'eastern european' and 'traveller'. The former, when used as a secondary rule alongside 'brothel' as the secondary word, flagged 41 of 22,783 possible modern slavery events in the Intel dataset. The latter, when used as a secondary rule

alongside 'vulnerable' as the secondary word, flagged 22 out of 22,783 possible modern slavery events.

ms_category	rule_type	primary_word	secondary_word	tertiary_word
sexual_exploitation	secondary	*N/A*	brothel	eastern_european
undetermined_slavery	secondary	*N/A*	vulnerable	traveller

The Committee would like to see a method developed for analysing how effective the tool is including in relation to false positives from the use of key words.

In order to assess the effectiveness of the Document Classification process, a balanced test set was formed (these are positives equating to 20% of tagged cases across Intelligence and Crimes). The evaluation involved manually verifying modern slavery tags as applied by the police (via HOCR Code/Intel Subject) as well as manually reviewing the flags from the Document Classification process. Once the manual review had taken place, it was possible to analyse how effective the Document Classification process had been in correctly identifying cases of modern slavery.

The process accounted for Intelligence tags applied by the Police that were contrary to intended use. A small volume of events did not align with expectations, such that an event was tagged as modern slavery by the police, however a review of the long text description of the case would state otherwise e.g. 'having inspected the premise, it is not believed that modern slavery is occurring'. At present, the volume of cases that did not align with expectations was minimal (1.6%) and was only accounted for to improve the credibility of the analysis. In an operational setting, events tagged as modern slavery will not have the Document Classification process applied to them.

Taking into account the updated Intelligence reports: of the cases identified by the Document Classification process only 1.82% were deemed to be a false positive (i.e. incorrectly flagged as modern slavery), with an associated precision rate of 93.5% (which is the ability of the classification model to return only relevant instances of modern slavery). When assessing how effective the process is in identifying *all* cases of Modern Slavery, 17.02% were deemed to be a false negative (not captured as Modern Slavery), with an associated recall of 60.6% (this is the ability of the classification model to identify all relevant instances), highlighting that the process is effectively capturing a large proportion of cases.

To ensure the best use of time by operatives when engaging with the tool, it is important that the capability correctly identifies Modern Slavery, hence the project has prioritised the achievement of a low false positive rate and high precision. To further explore the cases that the Police did not tag as modern slavery, a stratified sample was taken and manually reviewed. Of the 100 cases reviewed thus far, no events have been incorrectly identified as Modern Slavery (0% false positive rate).

As NDAS receives new data the Rule Generation process can be re-run to accommodate for any changes in terminology used to describe cases of Modern Slavery, ensuring the Rule Set accounts for current trends or modus operandi. It is important to note that during the Document Classification process no updates or changes to source system information e.g. existing tags applied by the Police, were made. It is for the professional judgement of Police Staff/Officer to make the relevant updates post-review.

Further information is requested about the possibility of using the tool for investigative purposes or for other purposes, specifically in relation to immigration control. If vulnerable victims are identified how will this information be shared with other agencies?

As stated during the November 2019 Ethics Committee session, there are no plans to share the output of this model with immigration authorities. Where vulnerable victims are identified, support will be available through the national referral mechanism. We feel that the ability to identify and support victims, as well as target the organised criminals exploiting them, far outweighs any risk that vulnerable victims may be subject to immigration enforcement as a result of the use of this model.

The Committee would like to see a breakdown of how many victims are identified and also how many offenders.

As a result of the capability being focused at the identification of possible Modern Slavery events, a breakdown of victim/offender volumes as presented within the dashboard will be reflective of the tagged data contained within the current Crimes system only (via respective Modern Slavery HOCR codes). The capability does not currently identify potential victims, suspects, or offenders of modern slavery—it instead identifies a possible Modern Slavery event, and allows the end user to understand the context (using the flags described above), and then use professional judgment to make an inference of which nominals might be viewed as a potential victim or suspect. This exercise mirrors and supports the existing investigative process, enabling practitioners to assess each possible modern slavery event faster and with a visual representation of all the Crimes data available to them.

The Committee recommended that general consideration is given to the overall outcome that is sought and whether/if data science methods can contribute positively to achieving those outcomes.

The outcome sought by the MS use case is the identification of modern slavery events and associated individuals involved with the aim to improve intelligence for the safeguarding of victims and targeting organised crime groups. The proof of concept showed that the model can identify thousands of individuals and numerous networks, saving hundreds of hours of analysts' time and identifying previously unknown networks of both suspects and victims.