WMP Briefing Paper

Missing Persons (Entity Extraction and Linkage)

Ethics Committee (6th March 2026)

This project is at the proposal stage and is presented to the committee 'in principle' so that any immediate concerns can be raised.

The finer details of the methodology, exact data to be used and mode of communicating the results will not be determined until after the exploratory data analysis (EDA) phase has been completed.

Once the analyses have been completed the project will be presented to the Committee again so that the data used, methodology, findings, intention for deployment and communication plans can be examined in more detail.

Legal opinion has been sought and the Data Protection Impact Assessment (DPIA) is being reviewed by the Force Data Protection Officer (DPO).

Tasking

This project was requested by the Intelligence Department.

The aim of the project is to analyse the fields section within COMPACT (the missing persons tracking system) and other WMP systems to identify intelligence relating to vulnerable individuals which may not be recorded within CONNECT with a focus on child sexual exploitation and abuse (CSEA) offending.

Purpose

The purpose of the analysis is to capture information (essentially entity extraction) from free text fields in COMPACT and CONNECT relating to phone numbers, locations, addresses and individuals linked to missing persons, the degree of frequency, etc. This information will then be used to support the safeguarding of vulnerable missing persons (including, but not limited to) those who are subject to grooming type offences.

Context

One of the core principles of policing is to prevent crime. The efficient utilisation of WMP systems in gathering intelligence around vulnerable missing persons will support the gathering of actionable insights to enable a proactive and preventative approach to policing. Such an approach will support the intervention in missing episodes, the safeguarding of individuals and reduce calls for service.

Intended activity resulting from the project

The intention is to develop an analysis of the COMPACT and other WMP systems with a focus on CSEA offending (but not limited to this) to identify key information presented within records of missing persons. The focus will be on information including names, phone numbers, addresses and locations. It is then intended to enable a network delineation of these items of information to enable their quick identification in relation to missing persons.

Ethical considerations

In the context of the current pressures on policing, this project aims to ensure that WMP utilises the systems and data available relating to (frequent) missing episodes. The aim is to highlight actionable intelligence to police proactively to reduce calls for service, ensure efficient safeguarding and to prevent crime. This in turn will provide support in building greater trust and confidence in policing.

The project will make use of COMPACT and CONNECT data relating to missing persons. Special category data will not be used, and the tool will be used as an information gathering and identification tool.

The Intelligence Department will work closely with the Data Analytics Lab (DAL) in the development of the tool and will use the output as an enhancement to current practices.

Data

Data to be used:

Any other Force systems which give information about either demand or WMP resources – likely to include:

- COMPACT
- CONNECT

Level of analysis:

☑ Individual
Individuals aggregated?
☐ Yes
☑ No
☐ Specific Area:
☐ Output Areas
☐ Super Output Areas - Lower
☐ Super Output Areas - Mid
☐ Wards
☐Districts
☑ West Midlands
☐ Other (type of crime)

Reliability of data:

An extensive exploratory data analysis (EDA) phase will be undertaken to examine the extent of any data quality issues. The data comes from standard WMP data sources which are routinely used in DAL projects and known data quality issues are accounted for.

Sample or entirety: Entirety

If sample:

Method of sampling: N/A

Method of choosing sample size: N/A

Sample size: N/A

Type of analysis:

☑ Exploratory
☐ Explanatory
☐ Predictive
☐ Optimisation
☑ Dashboard

Proposed methodology:

The following questions will be addressed:

1. Identify useful pieces of information in relation to (frequent) missing persons (names, phone numbers, locations, addresses, etc.); essentially entity extraction from free text.
2. Explore network charting / analysis of the useful information to enable better safeguarding of missing persons

Will the project eventually be automated: ☑Yes ☐No
Means of evaluation: Testing during and following the development of the process to test both the identification of relevant information, it's extraction and, where applicable, to minimise the potential for incorrect inference as to the kind of link between a missing person and a piece of information.

ALGO-CARE considerations

As this project is at the proposal stage and is presented to the committee 'in principle' in order that any immediate concerns can be raised, the finer details of the methodology will not be determined until after the EDA. Once the analyses have been completed the projects will be presented to the Committee again so that findings and methodology can be examined in more detail.

Advisory	
If applicable, are the outputs from the algorithm to be used in an advisory capacity?	The output would be advisory. The intention is to highlight potential information held on COMPACT and other WMP systems which may not be on CONNECT and to link information to support in the assessment of actionable intelligence for the Intelligence Department.
Does a human officer retain decision-making discretion?	Yes. The information would be considered in conjunction with other information and analysis conducted by the Intelligence Department. (In other words, it would be treated as intelligence.)
What is the policing purpose justifying the use of the algorithm (means and ends)?	This project supports the core principles in policing to prevent crime and supports the force strategy through delivering a service that works for local people and engaged communities through increasing efficiency, investing in prevention and through the efficient use of data.
Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes?	While data regarding individuals will be processed, this will not include special category data. This is a proportionate use of personal data which has been collected for a legitimate policing purpose.
In what way will the tool improve the current system and is this demonstrable?	At current, it is not possible to link information across frequent missing episodes and to highlight missing intel from COMPACT into CONNECT requires a manual process of reading each missing and found episode for each person and comparing against intelligence held on CONNECT.
Are the data processed by the algorithm lawfully obtained, processed and retained,	The data is from WMP systems and is collected as part of normal operational and organisational activity. As such data is collected in the appropriate manner and for the appropriate purposes.

according to a genuine necessity with a rational connection to a policing aim?	
Is the operation of the tool compliant with national guidance?	The analyses proposed would accord with the Government Digital Service Data Ethics Framework 2020[1]
Does the algorithm make suggestions at a sufficient level of detail given its purpose and the nature of the data processed?	The output will provide the Intelligence Department with information to support better intelligence gathering to support better data led decisions.
Are data categorised to avoid broad-brush grouping and results and therefore issues of potential bias?	NA
Do the potential benefits outweigh any data quality uncertainties or gaps?	The project will include an extensive EDA element and this should highlight areas of heightened uncertainty in the data or where potential gaps exist. There is currently no tool available to help the force in linking COMPACT intel in relation to CSEA and so this tool will reduce the level of uncertainty.
Is the provenance and quality of the data sufficiently sound?	The data has been gathered during day-to-day work of WMP and will enable analyses of the type envisioned for this project
If applicable, how often are the data to be refreshed?	To be agreed with end users – likely to be daily
If the tool takes a precautionary approach in setting trade-offs, what are the justifications for the approach taken?	NA
Ownership	

[1] https://www.gov.uk/government/publications/data-ethics-framework

Who owns the algorithm and the data analysed?	WMP would own the analysis and the data.
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?	The data and the analyses are contained wholly within WMP systems and the security measures employed therein.
What are the post-implementation oversight and audit mechanisms, e.g. to identify any bias?	Any process which is developed and productionised will have checks included to monitor its accuracy on an on-going basis as well as any consistent patterns that may represent biases or missed information. In this instance however it is unlikely that bias (in the sense generally discussed in the context of predictive modelling) will be present as this is aimed at entity extraction.
If the algorithm is to inform criminal justice disposals, how are individuals notified of its use?	NA
Accuracy	
Does the specification of the algorithm match the policing aim and decision policy?	Yes, this project supports the core principles in policing to prevent crime and supports the force strategy through delivering a service that works for local people and engaged communities through increasing efficiency, investing in prevention and through the efficient use of data.
Can the accuracy of the algorithm be validated periodically?	The productionisation of any process resulting from the project would include checking its accuracy on an ongoing basis.
Can the percentage of false positives / negatives be justified?	NA

How was the method chosen as opposed to other available methods?	Currently the broad approach has been identified due to the nature of the business question and the data available.
What are the (potential) consequences of inaccurate forecasts?	NA
Does this represent an acceptable risk?	NA
How are the results checked for accuracy and how is historic accuracy fed back into the algorithm for the future?	NA
How would inaccurate or out-of-date data affect the result?	Inaccurate data could lead to inefficient intelligence gathering and safeguarding interventions.
Responsible	
Would the operation of the algorithm be considered fair?	Given the am of the project to help in the finding and protection of missing persons, it is likely that the proposed analyses be considered fair. The potential for biases would be examined as part of the process development.
Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review?	The development of the process will involve detailed consultation with the Intelligence Department as the end users. A technical paper will be provided with the details of any process arising.
Would it be considered to be used in the public interest and to be ethical?	It would be considered to be in the public interest to ensure that limited resources are used effectively and that intelligence is gathered and used from all policing systems to safeguard vulnerable missing persons.
Is information available about the algorithm / decision-making rules and the impact of each feature?	A technical report will be produced which will include information about the methods used and assumptions made. The process will be developed with the Intelligence Department to ensure they make decisions based on sound understanding of the methodologies.

Appendix 1: Glossary of Terms

WMP / Law Enforcement Terminology	
DAL	Data Analytics Lab
HR	Human Resources
WMP	West Midlands Police
Missing persons	Missing persons

Data Science Terminology	
ALGO-CARE	All projects have used the ALGO-CARE to consider ethical implications: Advisory, Lawful, Granularity, Ownership, Challenge, Accuracy, Responsible, Explainable
EDA	Exploratory Data Analysis