

YEF Administrative Data Guidance

YEF online Evaluator Panel Conference

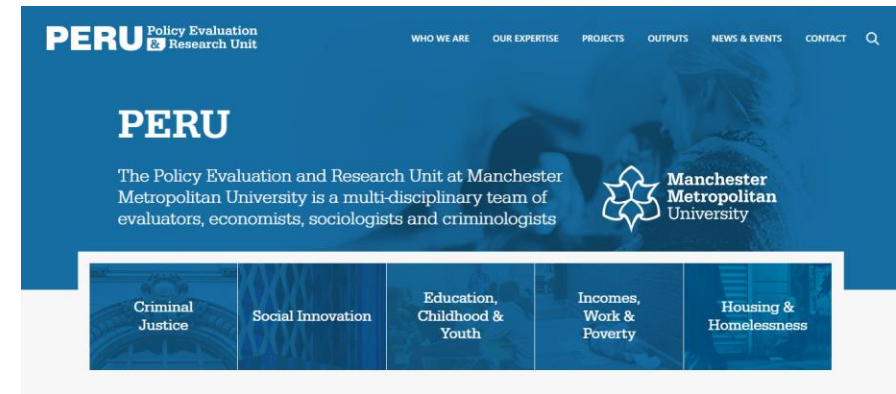
3rd December 2024

Mark Ellison
Will Cook

Structure of session

- Introduction
- Research (Desk work / Fieldwork)
- Report (Considerations, processes, access procedures, strengths & limitations, recommendations)
- Q& A

Introduction



Mark Ellison (Associate Head of Department, Sociology / Research Fellow, MMU)

Police forces / reconviction studies (PNC / JDL)

Dr Will Cook (Reader in Evaluation and Policy Analysis)

ADR UK Fellow / National Pupil Database-Police National Computer (NPD-PNC) datasets as part of the MoJ DataFirst initiative.

Policy Evaluation and Research Unit (PERU) www.peru.co.uk

Objectives of the project

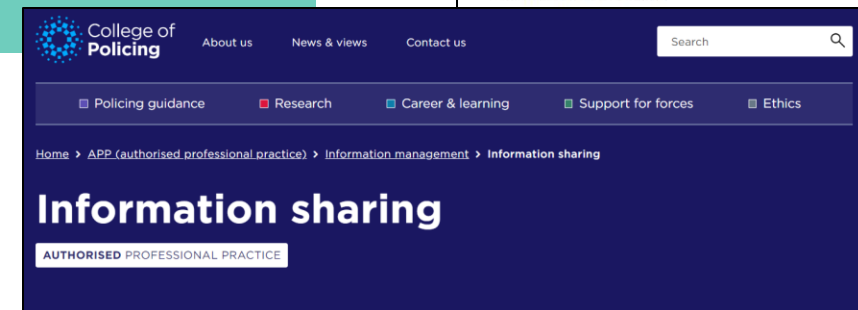
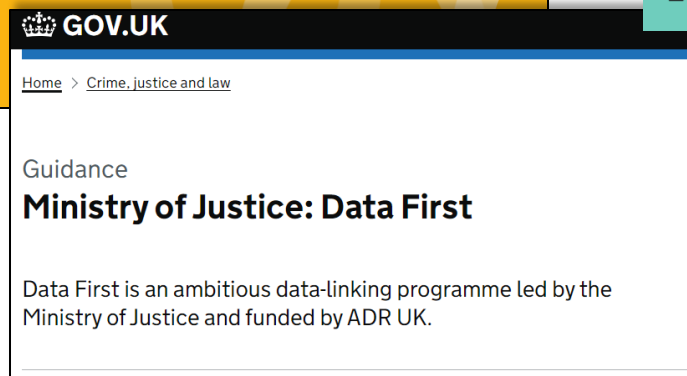
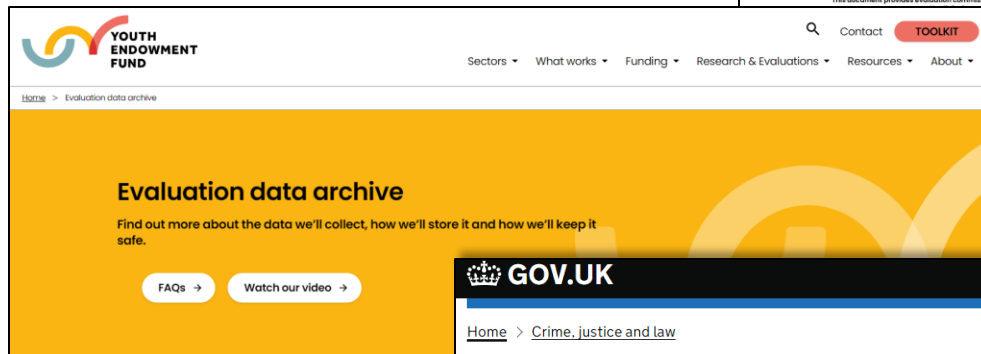
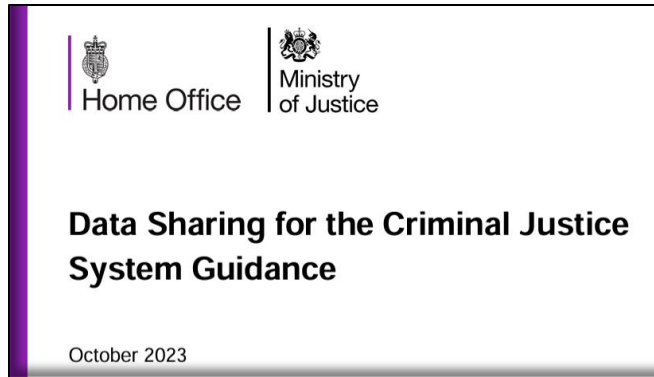
- To identify key individuals and organisation with knowledge of datasets that contain outcomes related to youth offending and violence.
- To define the contents of each data set and summarise the quality of the data.
- To document the procedures for evaluators to access the data.
- To assess the considerations that need to be taken into account when using each dataset for evaluation.
- To publish guidance that can be used in planning and executing YEF evaluations.

Fieldwork: Consultation with evaluation and data experts

- Dr Daniel Acquah (Youth Endowment Fund)
- Dr Nick Axford (Plymouth University)
- Professor Iain Brennan (Hull University)
- Steve Boxford (Cordis Bright)
- Professor Simon Coulton (Kent University)
- John Flatley (Home Office) Programme Director Crime & Policing Statistics and Acting Home Office Chief Statistician
- Sukhjit Gill ((Home Office)
- Professor Peter Langdon (Warwick University)
- Mike Parker (South Yorkshire Police / Violence Reduction Unit VRU)
- Kirby Seward (Ministry of Justice)
- Kevin Wong (Manchester Metropolitan University)



Desk Research: Review of Guidance Documentation



Guidance Report

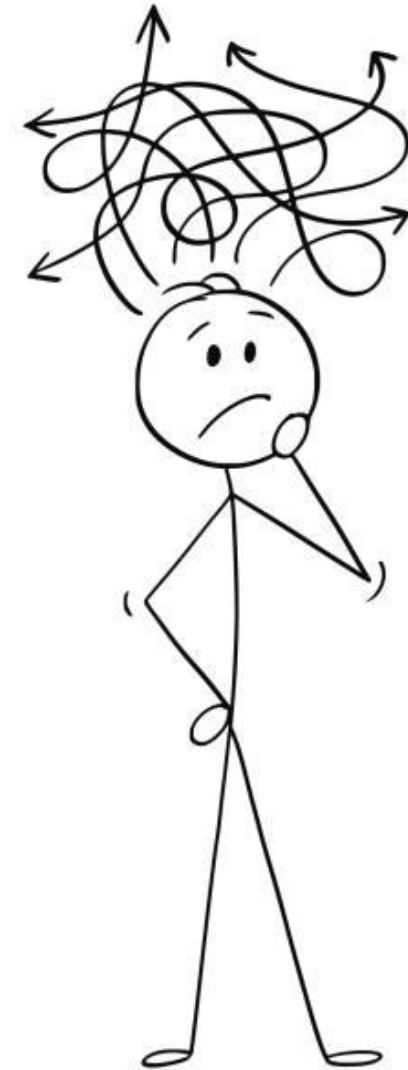
- Overview of the data access process
- Available datasets
 - Possibilities
 - Strengths and limitations
- Access procedures
- Recommendations



<https://youthendowmentfund.org.uk/wp-content/uploads/2024/09/YEF-Admin-Data-Guidance-Sept-2024.pdf>

Where to start?

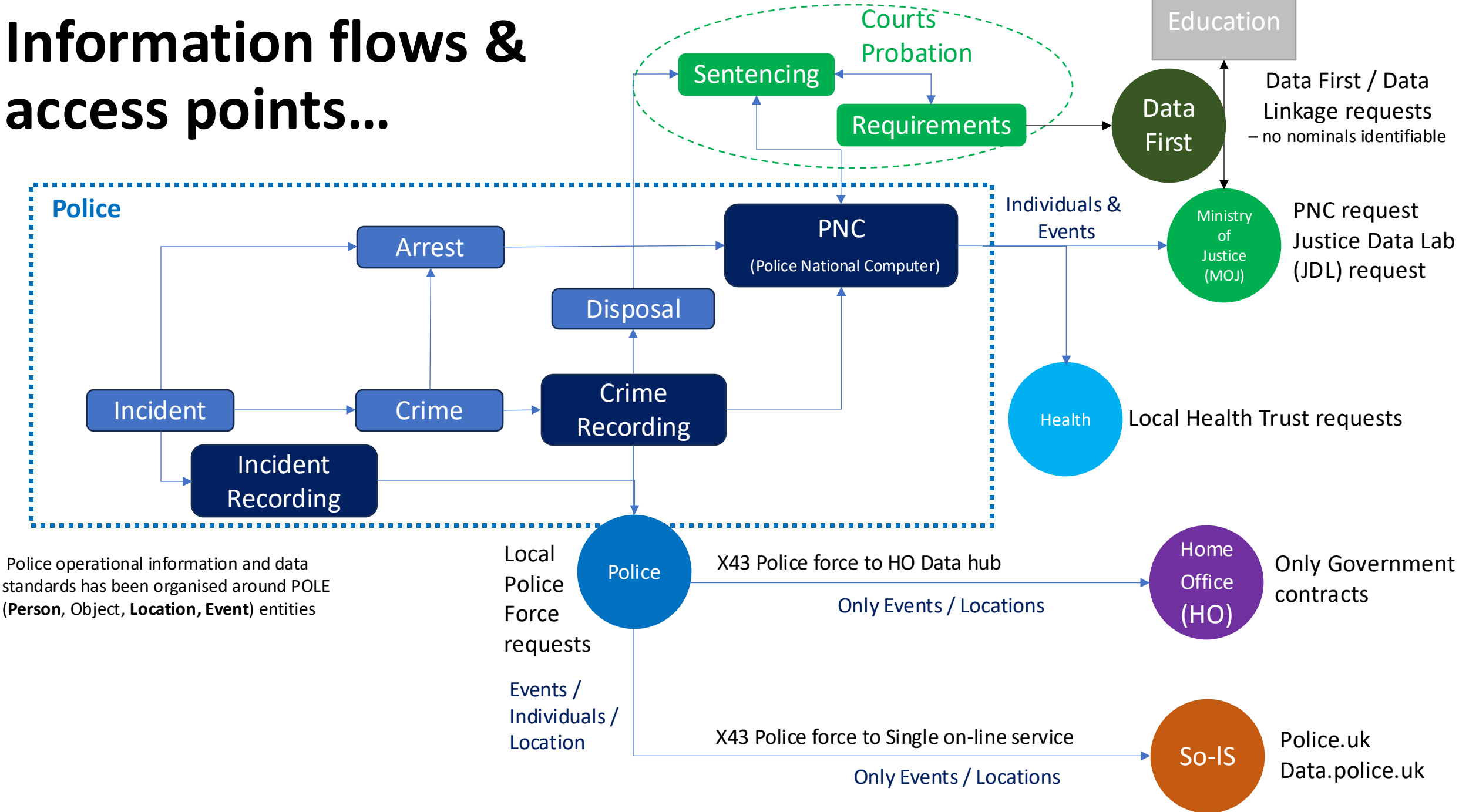
- What is available
- Considerations about trial design
- Data access: the process (modelled on local police data)
- Data access: where to go
- Strengths and Limitations of datasets
- Recommendations from study
- Sorry about the multiple process diagrams ☹️



Administrative Datasets

- *“Administrative data are a by-product of administrative systems developed primarily for operational purposes. Administrative data are used extensively in the compilation of many sets of official statistics about a wide range of topics”* (Office for Statistical Regulation, 2024).
- **Local police data (LPD)** includes police-recorded crime data collected by one of the 43 local police forces across England and Wales. LPD includes details of crime events (i.e. offence type, location and date/time) or suspects/offenders' demographic information (age, gender, ethnicity)).
- **Police National Computer (PNC)** data is a national dataset which includes information about police cautions and court convictions for individual offenders in England and Wales. The Ministry of Justice (MOJ) receives a data extract to examine offenders' convictions over time and conduct re-offending analyses by offender characteristics.
- **Hospital episode data** includes accident and emergency (A&E) attendance or hospital admission for injuries associated with violence. Data on individual (patient) episodes include demographics.
- **Linked datasets** across the criminal justice system and other government datasets (e.g. MOJ Data First) enable accredited researchers across government and academia to access anonymised, research-ready datasets ethically and responsibly.

Information flows & access points...



Police operational information and data standards has been organised around POLE (**P**erson, **O**bject, **L**ocation, **E**vent) entities

Local Police Force requests

Events / Individuals / Location

X43 Police force to HO Data hub

Only Events / Locations

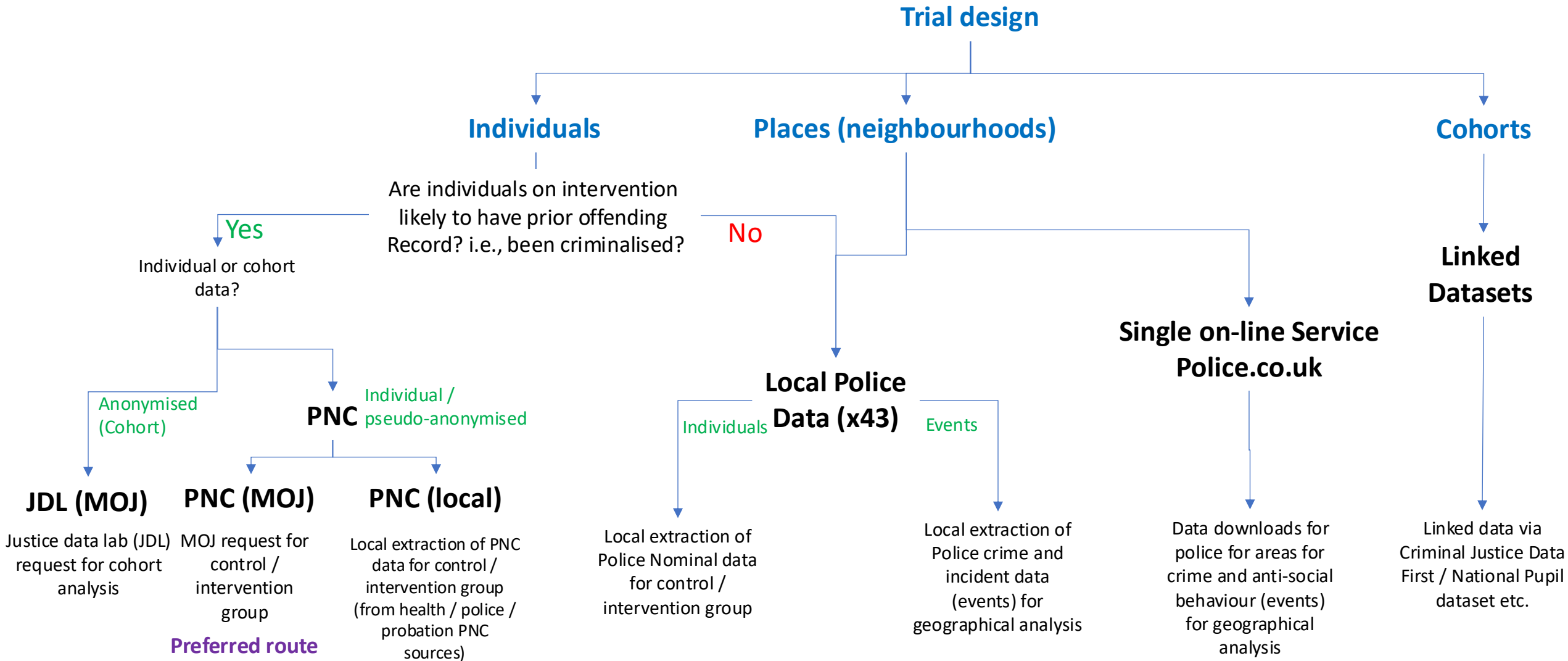
X43 Police force to Single on-line service

Only Events / Locations

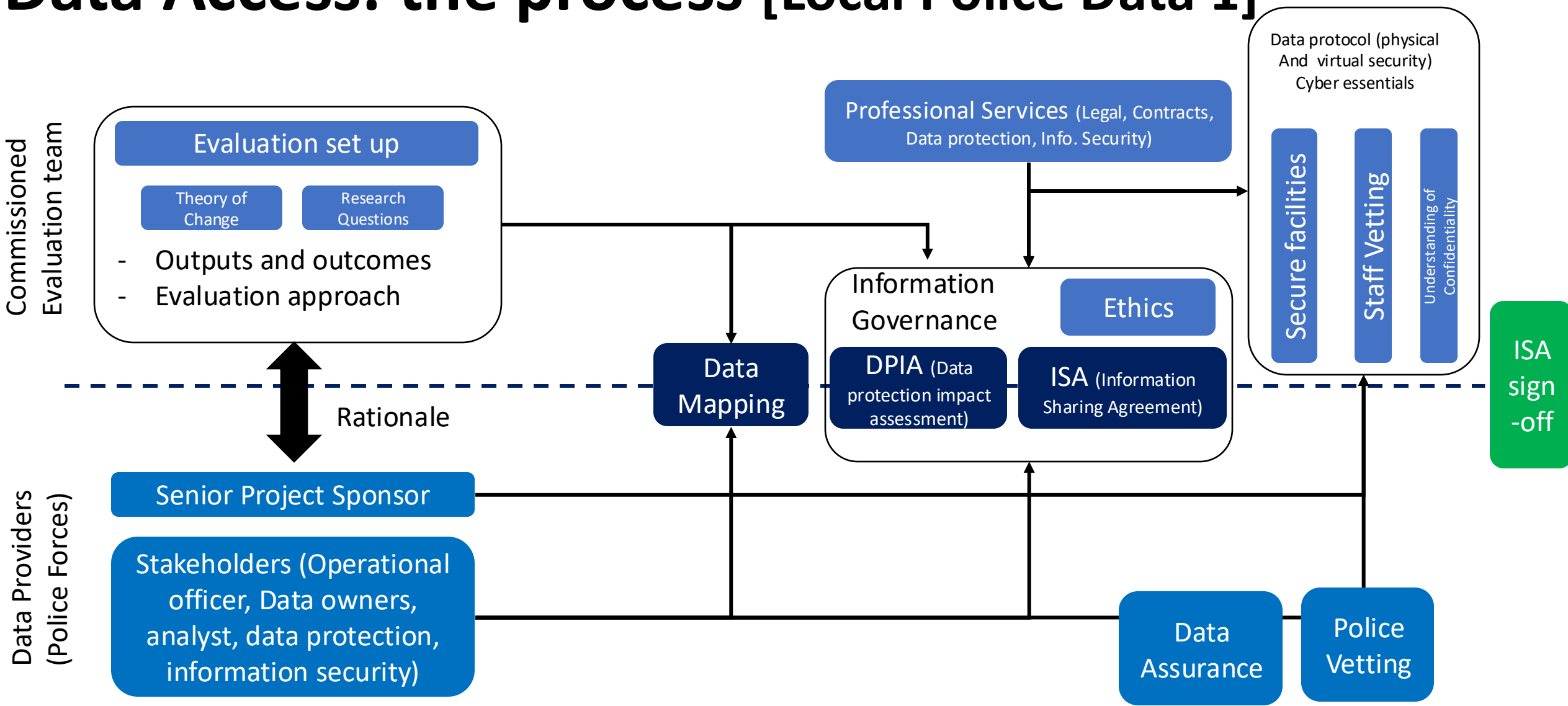
Only Government contracts

Police.uk
Data.police.uk

Trial Design decisions...

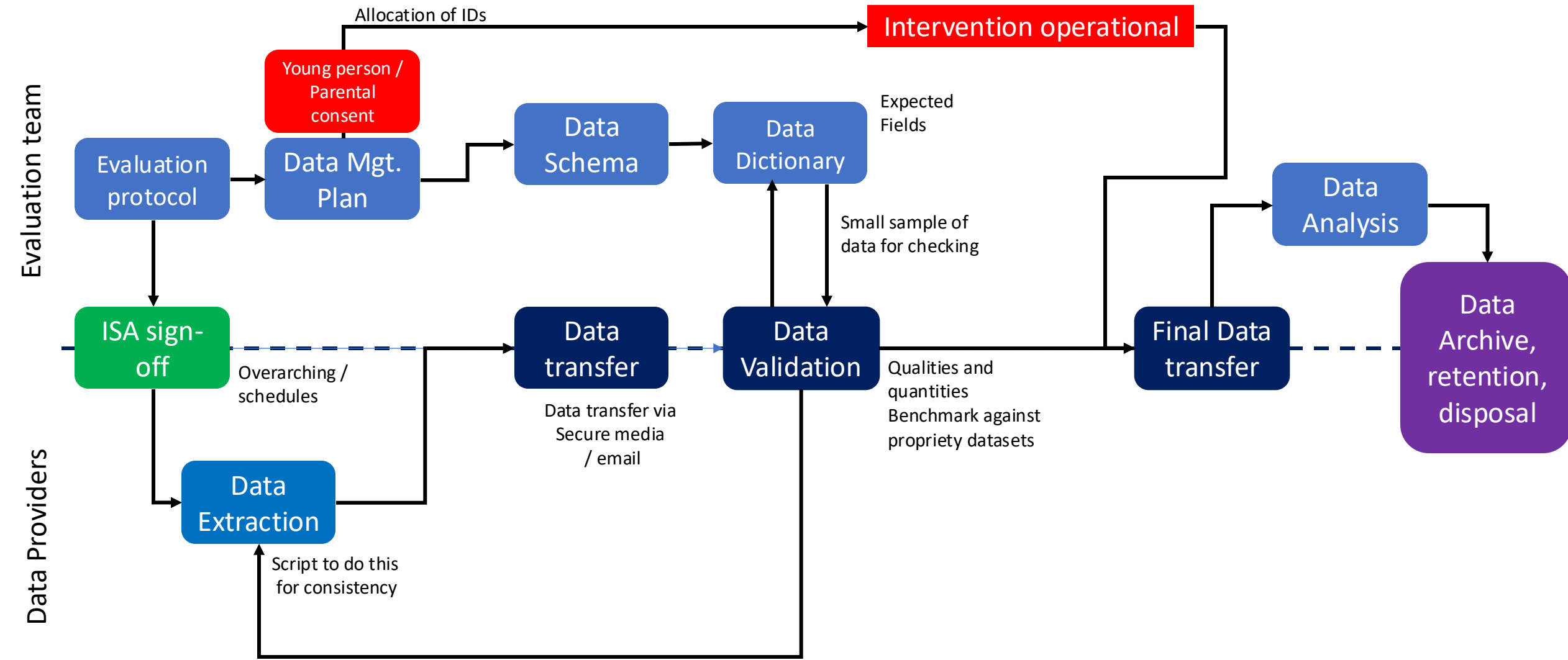


Data Access: the process [Local Police Data 1]



*Time limits on residency

Data Access: the process [Local Police Data 2]



Data Access procedures

- Who to contact | Information sharing procedures | Data Access Infrastructure
- **Local Police Data** – individual Police Forces
- **Police National Computer (PNC)**: MOJ extract data linking team datalinkingteam@justice.gov.uk
- Reconviction analysis provided by the **Justice Data Lab (JDL)**, contact justice.data@justice.gsi.gov.uk
- **Health Data** - For enquiries about health datasets, contact individual hospitals or foundation trusts
- **Home Office (HO)** recorded crime data email crimeandpolicestats@homeoffice.gov.uk
- **Data First datasets (including the MOJ PNC extract)**, contact either the Data First team (datafirst@justice.gov.uk) or the data linking team (datalinkingteam@justice.gov.uk).
- **National Pupil Database** – data Department for Education (DfE) at data.sharing@education.gov.uk
- **Single Online Service/police.uk** Crime datasets are available via the <https://data.police.uk/> website

Data Access: MOJ PNC

- Variables needed:
 - URN: Unique local identifier – to link back to original dataset
 - Forename / Surname
 - Date of Birth
 - Gender
 - Postcode
 - PNC ID (if available)
- Data returned: URN, case type, court of conviction date, offence (ID, codes, descriptions, primary), disposal date (duration/amount/rank)
- MOJ JDL – intervention cohort matched to a similar cohort and a binary reconviction analysis conducted

Strengths and Limitations of Datasets (1)

Dataset	Route	Strengths	Limitations
Local Police Datasets (LPD) (recorded crime)	Local access	<ul style="list-style-type: none"> • Comprehensive: all incidents (calls for service and crimes that are reported to police) • Crime events and nominals (suspects, offenders and victims) • Disaggregated or individual events • Geographical detail (x-y coordinates) enables the ability to link nominals to crime events at locations 	<ul style="list-style-type: none"> • Resource required to identify, match and extract datasets • Only available at the police force level. Therefore, data collection on a force-by-force basis (x43). • Potential for duplicate nominals in police data (if not assigned a PNC number)
Police National Computer (PNC)	MOJ access	<ul style="list-style-type: none"> • National dataset • Most comprehensive dataset for CJS contact or (re)offending • Data extract provided (with pseudo-anonymised data for linking) • Individualised records • Offending history 	<ul style="list-style-type: none"> • Data is subject to some inaccuracies inherent in any large-scale data recording system (e.g. mistyped data entries). • Only covers offenders • Long lead time for access (12m+) • May include a risk of bias by focusing on different offender/offence types over time and space
	Local access (e.g. police, prisons and hospitals)	<ul style="list-style-type: none"> • National dataset • Shorter lead time than MOJ data access request 	<ul style="list-style-type: none"> • Resource required to identify, match and extract data prior to analysis
	Justice Data Lab (JDL)	<ul style="list-style-type: none"> • A matched comparison group based on the characteristics of the intervention cohort 	<ul style="list-style-type: none"> • Long lead time for a report to be provided • Limited detail: only headline re-offending figures are produced for an intervention group against a comparison group.

Strengths and Limitations of Datasets (2)

Dataset	Route	Strengths	Limitations
Health	Local access	<ul style="list-style-type: none"> Hospital admissions or ambulance service data will supplement police datasets 	<ul style="list-style-type: none"> Local datasets at the hospital trust level Some hospital admissions due to crime may not be identified (e.g. domestic violence).
MOJ Data First	MOJ – ONS SRS MOJ – Secure Anonymised Information Linkage (SAIL) Databank (not MOJ-DfE Share)	<ul style="list-style-type: none"> Defined access route via MOJ Other Data First datasets available 	<ul style="list-style-type: none"> Currently not timely (up to 2021) for PNC outcomes in the MOJ/DfE dataset
National Pupil Database (NPD)	DfE – ONS SRS	<ul style="list-style-type: none"> Comprehensive: a census dataset that covers everyone educated in state schools 	<ul style="list-style-type: none"> Crime outcomes are limited. Absence and exclusions outcomes are not crime outcomes (but are predictive). NCCIS only records those in custody at the time of data collection.
So-IS	Police.uk	<ul style="list-style-type: none"> National coverage Monthly release Useful for benchmarking activities Suitable for LSOA-level analysis 	<ul style="list-style-type: none"> Only crime events with limited outcome information Highly aggregated (crime groups, monthly and to street centroids) and the anonymisation process may exclude some crimes Not all forces are submitting data to So-IS.

Recommendations

- Consider the **bias that exists in the data** - biases may lead certain groups to be under- or over-identified as being involved in crime and violence relative to their true level of involvement.
- **Relationships** are important and, to an extent, determine the speed of access.
- **Build in sufficient lead times** for accessing administrative datasets from police forces (3-6mths) and MOJ (12mths).
- **Ensure that consent is collected** from intervention participants to access police and PNC data and that appropriate ethical considerations are in place to undertake the evaluation.
- **Security standards with IT / researchers** (cyber essentials, vetting, DPIAs and ISAs)
- **YEF-orientated recommendation: Evaluation funding envelopes/time scales** provided through YEF commissions are too short to effectively evaluate the delivery of interventions, and the appropriate follow-up periods for re-offending measures (12 months + 6 months = 18 months) are standard evaluation timelines.



***Thank you for Listening
Any Questions***

Mark Ellison

Associate Head of Department, Sociology, MMU

m.ellison@mmu.ac.uk