



Ministry
of Justice



Overview of Data First datasets available in SAIL

Kylie Hill, Ministry of Justice &

Georgina Eaton, Ministry of Justice

19 January 2023

What we will cover

- Overview of Data First
- Data First datasets and research potential
- How to engage with Data First
- Q&A

Data First overview



Data First overview

- Ambitious data-linking programme led by the Ministry of Justice and **funded by ADR UK** (Administrative Data Research UK), who in turn are funded by the Economic Social Research Council (ESRC).
- Data First is unlocking **the potential of the wealth of data already created by MoJ**. Administrative datasets from the courts, and prison and probation services, in England and Wales are being linked at person level to allow analysis of users' pathways through the criminal justice system and interactions with other jurisdictions. The project is also driving forwards the linking of justice data with that of other government departments, including the **MoJ-DfE data share** (linking the Police National Computer with the National Pupil Database).
- The programme is enabling accredited researchers, across government and academia, to access **deidentified, research-ready datasets in an ethical and responsible way** via the ONS Secure Research Service and the SAIL Databank. By working in **partnership with academics** to facilitate research using this rich new resource, and through the creation of focused research fellowships, are **providing new evidence to underpin the development of government policies** and drive real progress in tackling social and justice problems.



How is the programme adding value to data for research?¹

Accessible Setting up secure and ethical routes to make de-identified data available to researchers who can make best use of it for projects in the public interest via ONS Secure Research Service and SAIL databank

Getting approval from data owners (HMCTS, HMPPS) and buy-in from stakeholders (judiciary, third sector organisations) that data can be shared

Broad Bringing more data into scope through improved data pipelines and new sharing agreements between government departments and arms-length bodies

Joining up data using a cutting edge, open source, record linkage programme called Splink. This in-house solution is robust, accurate, and able to work at scale on large datasets

Curated Working in partnership with data experts inside and out of government to determine what to share and; ensure that data shared is relevant and useable; and building in feed-back mechanisms

Documented Providing documentation and resources to communicate the content and quality of the data and provide information about how to access it and link it. Releasing broad initial research findings to demonstrate the potential of the data.

What does this mean for you as a researcher?

- More administrative datasets **available for access** to accredited researchers
- **Better information** about what MoJ administrative data is available and how it can be used, including **data catalogues** and **synthetic data** for some sources (more to come)
- Ability to **link together data** from across criminal justice system and from other sectors to carry out novel analysis
- Clearer **processes for applying** to use data in a specific research projects
- Standardised **secure access routes** for working with data and approving outputs via **ONS SRS** and **SAIL**. Network of locations and options for accessing data is expanding
- Support from Data First and a **growing community of users** to share expertise and best practice on working with the data
- **New research funding** and collaboration opportunities including **ADR Fellowships**
- Invitation to **feed into future development** of data releases to ensure that data shared is relevant and useable

Data First datasets and research potential



Data First datasets available in SAIL

Criminal justice

- Ministry of Justice Data First magistrates' court defendant
 - Ministry of Justice Data First Crown Court defendant
 - Ministry of Justice Data First prisoner custodial journey
 - Ministry of Justice Data First probation
 - *Ministry of Justice Data First linking dataset(s) (using Splink)*
 - Ministry of Justice Data First Safety in Custody
 - Ministry of Justice Data First Offender Assessment
- can link on defendant and on case
- focus on the defendant / offender – can link on person
- only includes records that have been linked to an offender in the Data First prisoner custodial journey dataset

Other court jurisdictions

- Ministry of Justice Data First family court (*can be linked with Cafcass data in SAIL*)
 - Ministry of Justice Data First civil court (TBC 2023)
- include multiple parties – planned to link to CJS data on person using Splink TBC 2023

Understanding the research potential of Data First

Administrative data can only help with certain types of questions

- Broadly speaking, Data First helps address questions about **who** is interacting with the justice system **in what ways**, the **processes and timings** involved, and **'hard' outcomes** such as sentencing or repeat appearances.
- The kind of data we have cannot provide insight into **unmet needs, circumstances and experiences** of engaging with justice services, and (at this stage) longer-term outcomes.

Data First focuses on justice system users

- Management information systems in scope for Data First focuses on **justice system users** (defendants, prisoners, parties to family court cases).
- Will provide little information for questions about **the judiciary, court & prison staff and legal professionals**. Limited admin data available on some public users e.g. victims, witnesses.

A good window into the big picture and small subsets

- Data on **all** defendants, cases, offenders in custody or under supervision etc. in England & Wales
- Can look at niche groups e.g. detailed ethnicities, specific offence types

Data criminal courts datasets overview

magistrates' court defendant dataset

Offender Characteristics	Offence Information	Proceedings and outcomes
<ul style="list-style-type: none"> • age • residence (local authority and lower super output area level) based on offenders postcode if applicable. • gender • ethnicity 	<ul style="list-style-type: none"> • principal offence (most serious disposal) • most serious offence (most severe regardless of conviction) • offence and arrest dates • police force 	<ul style="list-style-type: none"> • court • key dates (hearing, plea etc) • plea • mode of trial • finding • disposal • detail of penalty (sentence length, fine etc.)

Source: LIBRA, Period: 2011 to 2020

For further details of the variables included, please see the data catalogues on GOV.UK.

Crown Court defendant dataset

Offender Characteristics	Offence information	Proceedings and outcomes
<ul style="list-style-type: none"> • age • residence (local authority and lower super output area level) based on offenders postcode if applicable. • gender • ethnicity 	<ul style="list-style-type: none"> • principal offence (most serious disposal) 	<ul style="list-style-type: none"> • court • case type (e.g. trial, sentencing, appeal) • key dates (receipt, hearing, plea, disposal etc) • number of hearings • plea • mode of trial • finding • disposal • detail of penalty (sentence length, fine etc.) • appeal type and result

Source: Xhibit, Period: 2013 to 2020

Data First prisoner custodial journey dataset overview

prisoner custodial journey dataset	
Offender Characteristics	Sentencing and release
<ul style="list-style-type: none">• age• gender• ethnicity• marital status• nationality• religion• young offender flag	<ul style="list-style-type: none">• main offence (typically most severe convicted)• date sentenced• sentence length• number of sentences• imprisonment status• first and last movement dates• various release date types

Source: prisons NOMIS, Period: in custody from 2011 to September 2021

For further details of the variables included, please see the data catalogues on GOV.UK

Data First Safety in Custody dataset overview

Safety in Custody dataset

- Incident date and time
- Type of incident
- Incident location
- Participant role in incident
- Incident details, such as
 - Method of self harm
 - Type of assault
 - Injuries sustained
 - Treatment required
 - Delivery of treatment
 - Reason for incident

Source: prisons NOMIS, Period: in custody from January 2011 to September 2021
Data catalogue available on request

Data First Offender Assessment dataset overview

OASys dataset

- Social, Emotional and Mental Health
 - Psychological problems
 - Previous, current and pending treatment
 - Isolation
 - Attitude to self
 - Impulsivity
- General health
- Learning difficulties
- Relationships (current and previous)
 - Childhood
 - Family
 - Partner
 - DV experience
- Drugs and alcohol
 - Current and past usage
 - Impact on behaviour and offending

Source: prisons OASys, Period: in custody from January 2011 to September 2021

Data catalogue available on request

Data First Probation dataset overview

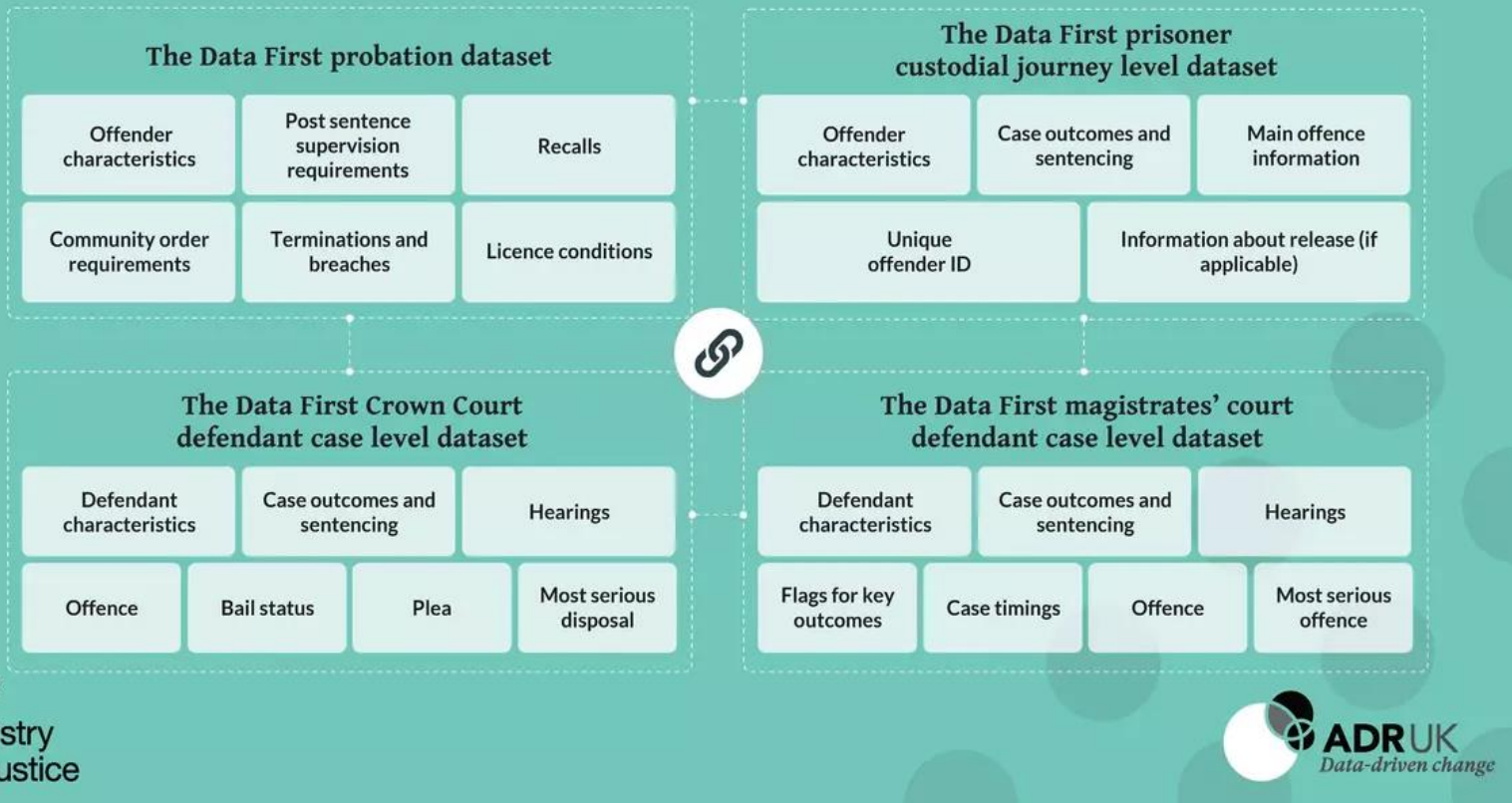
Offender and Event Information			Sentence Component Information			PSR Information
Offender Characteristics <ul style="list-style-type: none"> • Age • Residence (local authority and lower super output area level) based on offenders postcode if applicable. • Gender • Ethnicity 	Sentencing and Offence Information <ul style="list-style-type: none"> • Main offence (as picked by the person creating the 'event') • Main offence date • Number of additional offences • Disposal type • Date of disposal • Length of disposal 	Recalls <ul style="list-style-type: none"> • Earliest recall date • Earliest recall reason • Latest recall date • Latest recall reason • Recall count 	Community Order Requirements <ul style="list-style-type: none"> • Main type of requirement (e.g. unpaid work) • Sub type of requirement (e.g. regular) • Start dates and termination dates • Termination reasons 	Licence Conditions <ul style="list-style-type: none"> • Main type of licence condition (Residence at a specific place) • Sub type of requirement (Not to associate with any person linked with specific group/organisation) • Start dates and termination dates • Termination reasons 	Post Sentence Supervision Requirements <ul style="list-style-type: none"> • Main type of PSS (e.g. Standard 7 Conditions) • Sub type of requirement. • Start dates and termination dates • Termination reasons 	Pre-Sentence Reports <ul style="list-style-type: none"> • Type of court report (e.g. Pre-Sentence Report – Standard) • Type of proposal (e.g. ORA Community Order) • Proposed requirements • Note: where a report was not included in the data, does not necessarily mean the report was not carried out.

Source: nDelius, Period: Records with an event referral date from 2014 to 2020

For further details of the variables included in the Data First probation dataset, please see the [Data First probation data catalogue](#).

Data First linking dataset (covering CJS)

Information contained in the Data First probation and criminal justice system linked dataset



- A linking dataset connects records at person level across criminal courts, prisons & probation. This allows for end-to-end analysis of CJS journeys. (will cover civil and family in future)
- Probabilistic data linking carried out by MoJ using the [open-source Splink package](#). Using this link will join same records as in research at ONS Secure Research Service and MoJ Data First publications.
- Cannot be used to join to other datasets in SAIL – datasets also linked to SAIL spine using Anonymised Linkage Field (ALF)

Data First Family Court dataset overview

- Source: FamilyMan
- Geographical coverage: England & Wales
- Period covered: 1 January 2011 to 31 January 2021 (based on the date an event was entered into the FamilyMan system).
- Deduplicated and de-identified dataset containing around 16.3 million records
- 76 variables spread across 3 tables – cases, events & people

Cases

Events

People

- For further details of the variables included in the Data First Family Court dataset, please view the [Data First Family Court data catalogue](#) on GOV.UK

Data First Family Court dataset overview – Cases table

Cases

Case information

- Case type
- Number of hearings
- Court name
- Fee paid or exemption

Key dates

- Case creation
- Receipt
- Issue
- Case closed

Divorce & marriage information

- Marriage date
- Marriage type
- Divorce grounds
- Age

Flags

- Legal representation
- Contact activity provider
- Litigation friend
- Unpaid work facilitator

- Can join between the events & people table

Data First Family Court dataset overview – Events table

Events

Case and event information

- Case type
- Event description
- Event court
- Number of children involved

Key dates

- Receipt
- Entry

Hearing, application and order information

- Hearing type
- Application type
- Withdrawn application type
- Order type

Court information

- Court name
- Tier of judiciary
- Event heard at high court
- Fee paid

- Can join to the cases table

Data First Family Court dataset overview – People table

People

Personal information

- Gender
- Month/year of birth (for children only)
- LSOA of residence
- Local authority of residence
- Unique party ID

Role information

- Role name
- Role model

Case information

- Case type

- Can join to the cases table

Using linkage to address novel questions

Repeat appearances: over timeframe of datasets can see all instances where individuals appear

- identify those with frequent court appearances, multiple short prison sentences etc.
- patterns of offending – escalation, specialisation

Longitudinal analysis: follow appearances of individuals over time and build up picture of pathways and interactions

- arrests, court appearances, custody, release, recall, reappearance etc)

Filling in gaps: where data in one dataset is limited, can find detail in others

- e.g. time on remand courts/prisons
- data on people (characteristics) collected at different points in system

Contextual data / counterfactual: can look at population who don't appear in a certain dataset

- e.g. those convicted in court who did not receive custodial sentence

Cross-cutting questions: addressing different parts of justice system

Cross-cutting research enabled with further linkage

Linkage through SAIL Anonymised Linkage Field:

- CAFCASS / family court data (England and Wales)
- England and Wales Census
- Welsh population, health & social care data

Future linkages planned by Data First

- Cross-justice system (using Data First linking dataset)
- Other government departments

Linkage at aggregate level:

- Higher level factors linked on e.g. court ID, region (size, caseload etc.)
- Contextual /proxy data linked on geographic variables (LSOA) e.g. indices of deprivation
- Linking data on interventions for evaluation where known roll-out, pilot, comparison etc.

Mixed methods approaches

- Supplementing administrative findings with other sources and techniques e.g. qualitative, surveys

How to engage with Data First



Engaging via Data First

Academic Seminars

We have established a series of seminars with the academic and research community to raise awareness of Data First, identify priority research questions and facilitate use of the data. To register your interest in attending one of these seminars or to find out more about previous events, please email datafirst@justice.gov.uk.

Research Fellowships

In partnership with ADR UK, we have announced a rolling programme of funding calls seeking proposals from Research Fellows to use the newly available data. Future opportunities will be announced via [ADR UK's website](#).

Apply for data for research

Accredited researchers can apply to access available datasets for research via the ONS Secure Research Service & SAIL Databank. More information on how to apply is available in our Data First [user guide](#) and the [MoJ Data First application form](#).

For research applications using the SAIL Databank, submit information about your proposed research through SAIL's scoping form and feel free to get in touch with the Data First team at datafirst@justice.gov.uk

Get in touch:

datafirst@justice.gov.uk

Find out more:

Data catalogues, user guide & privacy statement:

<https://www.gov.uk/guidance/ministry-of-justice-data-first>

Application form and approved projects:

<https://www.gov.uk/government/publications/moj-data-first-application-form-for-secure-access-to-data>

Splink data linking package: <https://github.com/moj-analytical-services/splink>

Administrative Data Research UK: <https://www.adruk.org/>

Q&A

