

FAQs: Data First: Cross-Justice System dataset briefing ADR UK Research Fellowships 2025 funding opportunity

The following frequently asked questions were compiled from questions raised during a dataset briefing webinar on the Data First: Cross-Justice System – England and Wales dataset, which was held to support the applications for the 2025 round of ADR UK Research Fellowships.

- [Find out more about the ADR UK Research Fellowships funding opportunity](#) – open for applications until 26 February 2026.
- [Find out more about the Cross-Justice System dataset](#).
- Watch the [recorded webinar](#) about the dataset and access the [slide deck](#).

Dataset Content & Detail

1. **How detailed are the *Safety in Custody* variables? Do they include subcategories such as self-harm or drug-related incidents?**

The safety in custody table contained in our Data First prisoner custodial journey dataset includes the subcategories assault and self-harm. Drug-related incidents are not currently included in this dataset. For a more detailed look at what is contained in this dataset we publish a [data catalogue](#) and make available a low-fidelity [synthetic dataset](#).

2. **Does the dataset allow identification of cases involving mental health conditions or homelessness?**

The Data First offender assessment dataset includes a range of variables from the accommodation section of the full assessment, which aims to identify an offender's accommodation related needs. It also includes mental health related variables, but information on specific conditions is limited. For specific questions on this data we recommend getting in touch via our mailbox (datafirst@justice.gov.uk).

3. **How is geographical location recorded? Does it include court names, local authority codes, or more granular identifiers?**

Variables relating to the geographical location can vary according to the dataset, depending on what information is relevant. For a more detailed look at what geographical information is contained in the datasets we publish [data catalogues](#) and make available low-fidelity [synthetic datasets](#).

Linkage & Integration

4. Can the Data First dataset be linked to other ADR UK flagship datasets (e.g. RAPID or ECHILD) or external data such as ONS mortality data?

Unfortunately, it is not possible to link individuals in the Data First cross-justice system datasets to persons from other datasets held on the ONS Secure Research Service. However, it is possible to use geographic markers, such as Lower Layer Super Output Areas (LSOA) or Local Authority to join information on deprivation, for example by ingesting the index of multiple deprivation.

When using the SAIL Databank, it is possible for researchers to use the justice data shared through Data First, linked to other datasets held in SAIL.

5. Is it possible to link to health data via SAIL, and what types of health data are available?

Yes, it is possible to link justice datasets to health data held in SAIL which includes data from GP, A&E and substance misuse datasets. A full list of data available via SAIL is provided via the [Health Data Research Gateway](#), and the SAIL databank will be best placed to provide information on each datasets linkability.

6. If a project analyses licence supervision alongside post-release socioeconomic factors using other ADR UK datasets, is there a recommended order or timeline for securing cross-dataset approvals?

Specific orders will depend on the unique circumstances of a project, but the earlier we receive a project application, the earlier we can advise on areas such as feasibility and data availability, and ensure sufficient time is left to receive all necessary approvals.

Synthetic Data & Testing

7. Is synthetic data available without accreditation, and how can researchers access it?

You do not need to be an accredited researcher to access the synthetic data. The Data First synthetic datasets can be found on [the UK Data Service](#) and users are required to agree to and sign the End User Agreement in order to access the datasets.

8. Should researchers test cohort definitions and variable joins using synthetic data before submitting a proposal?

Data First synthetic datasets have been published to support researchers to develop their research proposals. It can be valuable to access and interact with the synthetic datasets while developing your research proposals to strengthen your application, and in conjunction with the data catalogues (available alongside the synthetic datasets at the UK Data Service). This can help researchers better understand what data is available, what cohorts can be defined, and how the datasets link together.

9. If a researcher uses synthetic data only (e.g. for a methodological case study), do outputs need approval, and can they be published if clearly labelled as synthetic?

Researchers accessing the Data First synthetic datasets must agree to the End User Agreement, which outlines how the data can be used. It is important to be very clear that the synthetic datasets available via the UK Data Service are low-fidelity synthetic datasets and therefore any analytical results obtained using these synthetic datasets will be meaningless. The tables have been designed to include joining fields as per the underlying, de-identified datasets that are made available to approved researchers via trusted research environments. The created synthetic tables contain low-fidelity synthetic data, with only within-concept correlations (such as matching ethnicity descriptions and codes) retained. No between-concept correlations have not been retained. If researchers wish to use the data for methodological research, we would encourage them to contact the Ministry of Justice to discuss.

Ethics & Approvals

10. Does ethical approval need to be in place before submitting an application, or can it be obtained later?

Ethical approval does not need to be in place when submitting an application, but it is needed before the application can receive data owner approval.

11. What are the key requirements for a strong application (e.g., variable justification, methodology clarity, sensitivity considerations)?

A strong project application is characterised first and foremost by clarity. Being clear in what research questions you want to answer, and how you are planning to do so is crucial. We recommend identifying your cohort of interest and stating how you will identify these individuals in the data, and explaining clearly and concisely how the selected methodology will allow you to robustly answer your research questions. Furthermore, genuine consideration of ethics and impact on data subjects, limitations and risks, as well as data minimisation principles are vital to create a well-rounded application which can receive approval from data owners, the trusted research environment, and the research accreditation panel.