The Data First Project:
Privacy and data protection

“Harnessing the potential of linked administrative data for the justice system.”
Contents

Summary 2
About Data First 2

How we use data 3
Adding value by linking and sharing 3
In line with legislation 3
For permitted research purposes 3
Taking wider views into account 4

How we protect personal information 5
Linking data 5
Considering risks and sensitivities 5
Limiting access to what is needed 5
Restricting and monitoring use and outputs 6

How we share data 7
Sharing with other government departments and agencies 7
Sharing with accredited processors 7
Sharing with external researchers 8
Approving access to data 9
Accessing your data 10
Reviewing the information 10
Contact us 10
Summary

This document has been produced to outline how we use and share data for research purposes as part of the Data First project and how we keep this data safe and secure.

About Data First

Data First is an ambitious data-linking programme led by the Ministry of Justice (MoJ) and funded by ADR UK (Administrative Data Research UK).

Data First aims to unlock the potential of the wealth of data already created by the MoJ by linking administrative datasets from across the justice system and beyond. Internal data linking work will link civil, family and criminal justice administrative datasets held by the MoJ. External data linking will seek to bring justice data together with data held by other large government departments.

As part of Data First, the MoJ are committed to sharing de-identified, research-ready datasets securely with accredited researchers across government and academia in an ethical and responsible way. By working in partnership with academics to facilitate and promote research in the justice space, we will create a sustainable body of knowledge on justice system users and their interactions across the family, civil and criminal courts, and their needs, pathways and outcomes across public services. This will provide evidence to underpin the development of government policies and drive real progress in tackling social and justice problems.
How we use data

Adding value by linking and sharing

Data First harnesses the potential of existing data; it does not involve the collection of any new data. This project reflects a drive for government to make better use of the information it routinely collects when users access public services, such as the courts.

Data First aims to enhance the potential of the MoJ’s existing administrative data by linking its internal data as well as linking externally with other government departments. Through making this data available to a wide range of accredited researchers for analysis via accredited processors we are enabling administrative datasets to reach their full potential and advance the quality of the justice system and the services it provides.

In line with legislation

The MoJ is permitted to process the data supplied by the police, the Crown Prosecution Service (CPS), courts and prisons by virtue of its common law powers for the administration of justice.

Data is only shared with other government departments where a suitable legal gateway exists and processed in compliance with all applicable data protection legislation including the General Data Protection Regulation (GDPR) and Data Protection Act 2018.

Data First has been developed within the framework established under the Digital Economy Act (DEA) (2017) which enables government to prepare administrative data for the purposes of research, and to provide de-identified versions of those data to researchers and projects accredited by the UK Statistics Authority (UKSA). This is the legal basis for the project and is the framework for ensuring that data is used responsibly, ethically and will not identify individuals.

In order for research projects to be approved they must comply with the Research Code of Practice and Accreditation Criteria which was approved by the UK Parliament in July 2018.

For permitted research purposes

Provisions for using data collected in the course of the department’s operations for research and statistical purposes are set out in the department’s Personal Information Charter and that of Her Majesty’s Courts and Tribunals Service (HMCTS).
Agreements to share data between government departments and to give access to researchers specify the permitted uses of the data and the safeguards in place, and restrict use to approved projects and purposes only.

These do not allow the administrative datasets that are part of Data First to be re-used for operational purposes or to investigate outcomes for individual users of justice services. MoJ will not make decisions that affect individuals as a result of the analysis of their individual data through this project. Additionally, no analysis of the performance of the courts, or of individual judicial decisions will be permitted. Any operational use of data linked and curated as part of Data First would not only be against our principles, it would also be illegal.

**Taking wider views into account**

The project has an external [Academic Advisory Group](#) made up of experts with wide-ranging technical and justice expertise, including on data linkage, and the safe and ethical use of administrative data for public policy. They provide advice and constructive challenge to the Data First team and help to ensure that the project follows best practice.

We will consult our internal ethics advisory group at MoJ and external ethicists as the programme develops to ensure that it, and the research projects that it enables, follow ethical guidelines and take concerns about impact on data subjects and data protection into account appropriately.

We are committed to ensuring that the needs and interests of the different groups who use the justice system are represented and protected throughout Data First. We will assemble a “Representative User Panel” made up of representations of third sector organisations who work closely with the users of the justice system to help consider questions around user and public acceptability.
How we protect personal information

Linking data

We link data relating to the same person by using identifying information, such as names, date of birth and addresses, that are held in the source data. This information will not be shared with researchers and will be replaced in the data linking process by a meaningless identifier (one that has been generated for these datasets and is not used in any existing operational systems). Other identifiers used within the justice system, such as case IDs, will also be replaced.

See our user guide for more information on data linking and identifiers.

Considering risks and sensitivities

The MoJ recognises the risks associated with sharing data. As well as only allowing researchers to access data after identifiers have been removed, we have rigorous safeguards in place to ensure data cannot be accessed by any unauthorised persons, or for any reason other than approved research projects.

Information classed as ‘special category’ data under GDPR are sensitive, as is information about criminal histories. The MoJ recognises the risk of this information being disclosed and we require additional justifications for processing before sharing special category data, in line with GDPR.

We may share personal descriptors such as sex or gender, ethnicity or age that we think is insufficient to identify someone. Some fields or combinations of data do not identify a person directly but increase the risk of identification. This could occur, for example, if very few people match a set of characteristics or activities recorded in the data. To protect individuals, our application process will consider the sensitivity of both the individual data items and the full combination of data being requested. The ethical issues around the use of data are factored into the decision-making process. All research done using Data First must be approved on the basis that it is in the public interest and offers real potential for public good.

Limiting access to what is needed

The MoJ will take steps to minimise the amount of data used for the purposes of achieving the specified objective. Researchers will have to justify why they require the use of each specified field and the coverage of the dataset asked for in their research application.
For more sensitive data, greater justification is needed. For example, where two fields give different levels of details (such as offence categorisations) researchers would need to justify that their research question requires use of the more extensive version.

**Restricting and monitoring use and outputs**

The users will access data in a secure setting without internet access and with limited external resources. This limits the risk of identification as researchers will not be able to combine the data with information outside of the dataset and use this to infer information about individuals.

The current primary processor for accessing the data is the Office for National Statistics (ONS) Secure Research Service (SRS) (see below: How we share data). Users’ activity is closely monitored in the ONS SRS and access will be revoked if they go beyond the permitted uses of the data. Attempting to identify individuals is strictly against the terms and conditions which researchers must agree.

All outputs are checked to ensure they remain within the scope of the project that has been approved and that personal information is protected as far as is reasonable, for example, through minimum cell counts and statistical disclosure control.

Data is only kept as long as necessary and is then securely deleted.
How we share data

Sharing with other government departments and agencies

Data shared between the MoJ and other government departments will always be governed by a Data Sharing Agreement (DSA). This establishes a framework for appropriate processing, including the contents and duration of the share, permitted uses of the data, the legal basis and justification for proceeding, and the protections in place throughout its lifecycle.

A data protection impact assessment (DPIA) will also have been completed prior to data sharing, to help identify risks to individuals’ privacy and ensure appropriate protections are in place to minimise them. As the share evolves, so too will the details in the DPIA and DSA report.

For data linking to take place, personal identifiers (such as name, address and date of birth) are transferred between departments. Access to this information is strictly limited to those who need it in order to carry out the linking process. These identifiers are transferred separately from all other information about individuals and deleted immediately following linking, to minimize the risk of identification. New, meaningless IDs assigned to individuals throughout linkage are the only IDs provided for use in analysis. Only the minimum of information (both fields and coverage) justified for sharing is supplied. All data are transferred using an agreed mechanism and stored in a setting that is compliant with HMG Security Policy Framework.

Sharing with accredited processors

To facilitate sharing with researchers outside of government, the MoJ will share deidentified, research-ready data with accredited processors under the Digital Economy Act (2017) to make access securely available for research purposes.

Currently, data is shared with one accredited processor: the ONS SRS. The SRS is designed to deliver full compliance with the statutory Conditions and Code of Practice of the DEA (2017) and has substantial data expertise, especially in data management, metadata, and the checking of outputs before they leave the centre.

Each dataset shared between the MoJ and ONS will be subject to a DSA and DPIA.

The de-identified extract of the data will be transferred using an agreed mechanism and stored in a setting that is compliant with HMG Security Policy Framework.
Sharing with external researchers

Researchers will be given use of the deidentified data according to the needs of their accredited research project only.

User access to the shared data is to be governed by a project application process whereby users must specify exactly what data they require for analysis and for what purpose. Access to the data is limited to a fixed time period agreed by the MoJ. Research outputs will consist of aggregate data only and users are monitored and exports checked to ensure that their analyses remain within the scope of the project that has been approved.

Data First uses the ONS ‘Five Safes’ framework to ensure information is kept safe and secure:

**Safe People** – Trained and accredited researchers are trusted to use data appropriately.

Data First allow access to data extracts only to accredited researchers, who have met the conditions published by the UKSA in the Research Code of Practice and Accreditation criteria.

We also seek assurance in the application process that researchers have appropriate skills and expertise in working with data and have undertaken data protection training.

**Safe Projects** – Data are only used for valuable, ethical research that delivers clear public benefits.

All research applications to access Data First datasets are scrutinised by data access panels at MoJ, HMCTS and/or partner departments (where applicable) to consider whether the research proposal is ethical and its potential impact on data subjects; whether the data access sought is necessary to carry out the analysis; whether there are any data protection concerns; the research proposal’s methodology; and benefit to the public good.

After approval, to access data in the ONS SRS, the research application must also be accredited by the UKSA Research Accreditation Panel (RAP). Projects must meet the accreditation criteria published by the UKSA in the Research Code of Practice and Accreditation criteria.

**Safe Settings** – Access to data is only possible using secure technology systems.
Access to the SRS is through a safe setting in safe rooms on ONS sites, in safe rooms on other certified sites, or through an organisation which has an Assured Organisational Connectivity Agreement with ONS and which maintains current certification.

**Safe Outputs** – All research outputs are checked to ensure they cannot identify data subjects.

Research outputs will consist of aggregate data only and will be checked thoroughly to ensure that they are in line with the stated purposes of the project and that personal information is protected.

In line with the Digital Economy Act (2017; Bar on further disclosure of personal information), all research outputs from the SRS will undergo thorough Statistical Disclosure Control (SDC) prior to publication to protect the identity of individual persons and/or businesses in those releases.

**Safe Data** – Researchers can only use data that have been de-identified.

All data provided through Data First will be de-identified and risks and sensitivities around sharing the combination of data requested will be considered before access is granted.

Please see [how we protect personal information](#) for more details.

**Approving access to data**

Applications by external researchers will be scrutinised by internal governance processes based on their individual merit.

The [application form and guidance](#) can be found on gov.uk.

- MoJ governance processes require consideration by the Data Access Group (DAG) and approval from the Data Access Governance Board (DAGB) chaired by MoJ’s Chief Statistician.
- If access to data on courts and tribunals is requested HMCTS processes apply. These require consideration by their Data Access Panel (DAP). Applications must have an HMCTS business sponsor to proceed to assessment by the Panel.
- If access to data owned by other government departments is requested they also require consideration by relevant processes at partner organisations.
A brief summary of successful applications to access the data will be published on gov.uk. There may be exceptions where we do not publish information – for example, to protect the privacy of researchers working on particularly sensitive topics.

**Accessing your data**

Information on how to request access to copies of your personal information can be found in the MoJ’s personal information charter.

**Reviewing the information**

The data will be reviewed and refreshed annually to include new data. This privacy statement and our data protection processes will be updated to reflect any relevant changes.

**Contact us**

If you would like more information or have any questions, please contact us at: datafirst@justice.gov.uk