Ministry of Justice

ADR UK

The Data First programme and opportunities for criminological and criminal justice research

Andromachi Tseloni

Data First Academic Lead

Professor of Quantitative Criminology, NTU

Crime Surveys Users Conference

Tuesday 08 December 2020

Contributions to this presentation from:

Ministry of Justice, Data First

- Lead and Government Statistician,

Amy Summerfield and Kylie Hill; and

- Social Researchers

Caris Greyson, Thomas Jackson, Toby Ogun and Daisy Ward

ADR UK

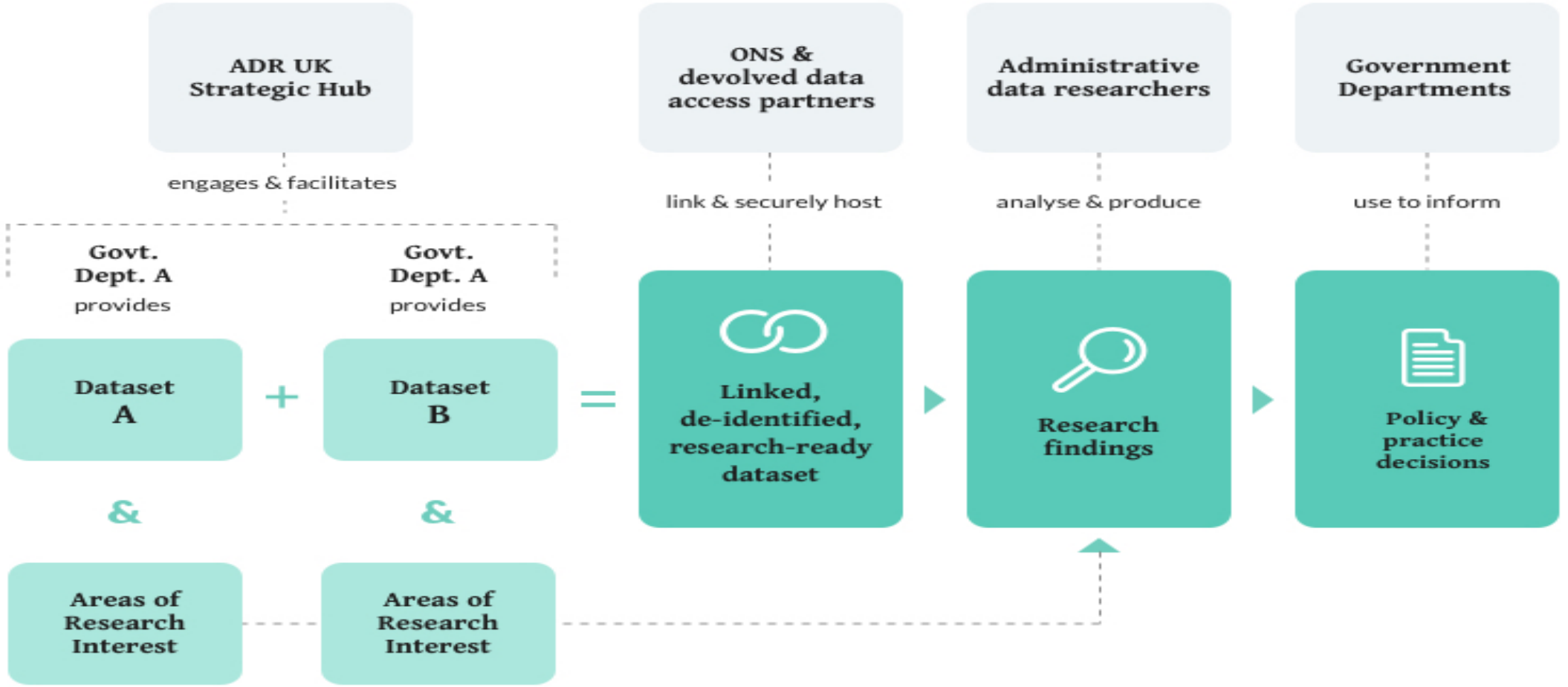
- Head of Research Strategy and Commissioning and Strategic Hub team members,

Karen Powell, Catriona Taylor, and Bogusia Wojciechowska

Project Overview

- Ambitious data-linking programme led by the Ministry of Justice (MoJ) and **funded by ADR UK** (Administrative Data Research UK), who in turn are funded by the Economic Social Research Council (ESRC).
- Data First aims to **unlock the potential of the wealth of data already created by MoJ**, by linking criminal, civil and family administrative datasets from across the justice system and beyond, and enabling accredited researchers, across government and academia, to access **anonymised, research-ready datasets in an ethical and responsible way**. The project will also enhance the linking of justice data with other government departments.
- By working in **partnership with academics** to facilitate research in the justice space, we will create a sustainable body of knowledge on justice system users 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 does it work?



How does it work?

The **ADR UK Strategic Hub** is responsible for coordinating work, such as **MoJ Data First**, across the partnership and manages a dedicated budget delivering UK-wide datasets and research

The **Office for National Statistics (ONS)** is an integral part of ADR UK. ONS is the UK's official statistical body with decades of experience processing and safeguarding governmental data. Data is made available to researchers through the **Secure Research Service (SRS)**.

What does the project involve?

Four different MoJ internal teams leading on different workstreams of the Data First project:

Internal data-linking

Data scientists and data engineers leading on the development of a robust, automated linking pipeline between criminal, family and civil justice datasets.

External data-linking

Statisticians and operational researchers leading on establishing data-shares with external partners and linking justice data with OGDs.

Data mapping and strategy

Social researchers leading on mapping data held across MoJ with a view to develop an externally-shareable list of research-ready datasets.

Research, academic engagement and communications

Social researchers and statisticians who are facilitating the link between Data First and the research and academic community, working to identify priority research questions to make best use of the linked datasets.

Data First De-identified Criminal Court Data Currently Available for Research

Data First Criminal Court Datasets

Magistrates' Court Data:

- 12.4 million records, deduplicated data.
- One record per-defendant per-case, giving details on characteristics, offence, proceedings, and outcomes in criminal cases dealt with by the magistrates' court in England and Wales (and Youth Courts)
- Coverage from January 2011 to December 2019

Crown Court Data:

- 0.9 million records, deduplicated data.
- One record per-defendant per-case, giving details on characteristics, offence, proceedings, and outcomes. Appeals data is less populated than for sentencing and trial data.
- Coverage from January 2013 to December 2019

Magistrates' and Crown Court Linking Data

- A lookup which allows users to join information from the individual magistrates' court and Crown Court datasets.
- It contain two tables, one as a lookup to group defendant records together (**defendant link table**), and the second as a lookup to link cases between the Crown Court and the magistrates' courts (**case link table**).
- The case link is nested with the person link, which means that ***a case is only determined the same if the person is first determined to be the same***.

Internal Data Share Timeline

June 2020:

- MoJ Data First magistrates' court defendant case level dataset

August 2020:

- MoJ Data First Crown Court defendant case level dataset

October 2020:

- Linkage of the criminal court shares i.e. magistrates' and Crown Court datasets

2021:

- Prison & Probation data to be linked with magistrates' and Crown Court data

2022:

- Linkage of all Data First datasets – criminal, civil and family

Achievements to date

MoJ – DfE data-share	The MoJ/DfE share provides data on childhood characteristics, educational outcomes and (re)-offending. The shared information consists of data on the educational characteristics of young people (from DfE), linked to data on their interactions with the criminal justice system (from MoJ).
MoJ Data First linked magistrates' and Crown Court dataset	This linking dataset will allow users to join up information in the Data First magistrates' and Crown Court datasets. It does not itself contain information about people, or their court appearances, but acts as a lookup to identify where records in the two criminal courts datasets refer to the same people and cases. There are two tables: one to identify which defendants are believed to be the same person (defendant table) and one to identify which records are for the same case (case table).
MoJ Data First magistrates' court defendant case level dataset	A de-duplicated share that provides data around defendant appearances in the magistrates' court between 2011-2019, extracted from the magistrates' court management information system- Libra.
MoJ Data First Crown Court defendant case level dataset	A de-duplicated share that provides data on defendant appearances in the Crown Court between January 2013 and December 2019, and is extracted from the new court management information system (Xhibit).
Development of the open source package 'Splink	**Splink** uses Python and Apache Spark to link and de-duplicate data flexibly, transparently and efficiently using the Fellegi-Sunter linkage model.
Academic engagement	We have established the **Academic Advisory Group** (AAG) to provide expert advice and challenge throughout the project and raise Data First's visibility through engagement with research networks and academia. We have also established a series of **academic seminars** with the wider academic community to promote the project.
User engagement	We have established the **User Representation Panel** (URP) to provide user perspectives on justice users' needs and policy priorities and enhance Data First's academic research impact through direct communication and potential collaboration between researchers and justice users.
Publications	We have created a data privacy statement, a user guide, and data catalogues for each dataset. These have been published on our Data First gov.uk webpages, alongside a new application form and supporting guidance for access to data. These publications will be continually updated with the incorporation of further datasets.

Technical Support or Data Enquiries

- **Please contact datafirst@justice.gov.uk for any questions regarding:**
  - Dataset enquiries,
  - If you would like access to the fake synthetic data,
  - Queries on the Application Form for Secure Access to Data

Thank you for your attention

andromachi.tseloni@justice.gov.uk

Magistrates' Court Data

- Information on defendants appearing in the magistrates' courts **in England and Wales**, from **January 2011 to December 2019**.
- **Defendant level** – each row shows a defendant on a case in the magistrates' courts.
- A very large dataset – **12.4 million rows**.
- **Deidentified data**, containing an *estimated defendant id* which is a unique identifier for the defendant. This is the key field to use for longitudinal analysis of defendant appearances.
    - The estimated defendant id is a derived field resulting from a process of probabilistic record deduplication using personal information not shared in this dataset. It therefore represents a statistical estimate of which defendants are the same person.
- **Deduplicated data**, to allow us to identify repeat users of the magistrates' court. A user who has entered the court multiple times will have the same *estimated defendant id*.
- Information on **cases**, *case id hash* presents a unique identifier for the case, and *defendant in case id hash* presents a unique identifier for a defendant within a case.
- Details of the **principle offence** (the offence with the most serious disposal) are given. If there was a murder and possession of cannabis – conviction was for drug possession, and acquittal of murder – then the principle offence is set to possession of cannabis.

Crown Court Data

- There are 0.9 million records in the deduplicated Crown court data and the data coverage is from **January 2013 to December 2019**.
- The Crown Court deals mainly with **appeals** against conviction and/or sentence regarding offences dealt with in the magistrates' court. Appeals have not been linked back to original case records or offences, compared to trial and sentencing cases and therefore data available for appeals is limited.
- **Most serious offence:** Refers to the offence within a case against the defendant with the maximum potential disposal penalty (e.g. longest sentence length), irrespective of outcome in this instance.
- **Most serious disposal:** Refers to the offence within a case against the defendant with the most serious disposal actually received. For example, the offence receiving the longest sentence or highest fine. This could differ from the 'most serious offence', for example if the defendant is only found guilty of a less serious offence.
- **Triable-either-way:** These are more serious than summary cases (heard by magistrates' courts) and can be dealt with either in the magistrates' court or before a judge and jury at the Crown Court. A defendant can invoke their right to trial in the Crown Court, or the magistrates can decide that a case is sufficiently serious that it should be dealt with in the Crown Court.

ONS SRS researcher journey