The Data First programme and opportunities for criminological and criminal justice

Andromachi Tseloni, Data First Academic Lead

andromachi.tseloni@justice.gov.uk or andromachi.tseloni@ntu.ac.uk

British Society of Criminology, Early Career Researchers Network
3 November 2021
Overview

✓ Research resource
✓ Academic research funding and MoJ collaboration
✓ Research Fellowships
✓ Data access and project approval
✓ Potential research topics
✓ Useful links – *get in touch* *(throughout the presentation)*
Data First - a research resource
Administrative to research ready data journey

ADR UK Strategic Hub

engages & facilitates

Govt. Dept. A provides

Dataset A

&

Govt. Dept. A provides

Dataset B

&

Areas of Research Interest

&

Areas of Research Interest

ONS & devolved data access partners

link & securely host

Administrative data researchers

analyse & produce

Government Departments

use to inform

Linked, de-identified, research-ready dataset

Research findings

Policy & practice decisions
Data First project 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 aims to unlock the potential of the wealth of data already created by MoJ, by linking 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.
Progress and updates

New datasets releases

The criminal courts and prisons linking dataset (May 2021) and the family justice dataset (October 2021) have been shared with the ONS SRS. Accredited researchers can now apply for the deidentified, deduplicated data. The family court dataset was published alongside detailed data documentation.

Refreshed publications

The criminal courts data catalogue has been refreshed and updated to reflect the extended version of the criminal courts data (now covering data up to December 2020) and to improve variable descriptions. The user guide has been updated to include guidance for new datasets.

Funding call & applicant workshop

A rolling programme of funding calls with ADR UK on linked datasets that are released: The successful applicants for the criminal courts funding call have started their research and presented at our academic seminar 20 October. An induction event was held on 21 October for the five successful Fellows of the MoJ/DfE funding call. Future calls will accompany the release of new linkages.

Internal research publications

The first MoJ-led research report based on the Data First criminal court datasets is due for publication shortly. The report is based on an exploratory analysis of returning defendants from 2011 to 2020 in England and Wales, and the potential of the linked criminal courts data.

User Representation Panel

We have held two quarterly meetings with the URP, alongside a workshop on our data shares and arranging a consultancy meeting for our research fellows.
Six datasets available so far and more data in the pipeline

- End to end justice system dataset by 2022 to provide a more complete picture of justice system users
- Annual refresh of datasets to add in further data as it is curated
- External data-linking programme will extend the range of questions that can be addressed by bringing in evidence about past experiences, complex needs and other government services
Data First – academic research funding and MoJ collaboration
Academic research funding

- **ADR UK funded Data First research** (e.g. feasibility studies, research fellowships)
- **Post-graduate research** (e.g. Doctorate-level studies, at AOC institutions MA/MSc/ MPhil dissertations)
- **UKRI funded research** (e.g. ESRC Secondary Data Analysis Initiative, mixed methods)
- **MoJ directly commissioned research projects** (e.g. data linking ethics, linked data representativeness, data users’ views on linked data)
- **HEI funded academic research** (e.g. “daily” independent research activity, internal /research centres interdisciplinary/seed-corn/ knowledge exchange funding)
Avenues for academic collaboration with MoJ

- Helping to shape our internal and directly commissioned research programme (e.g. identification of evidence gaps and feasible research questions, prioritising questions from ARI)

- Research project-based collaboration with MoJ (e.g. secondments, joint bids)

- Working alongside Data First Team on specific areas of research/analysis (e.g. helping with the analysis, interpretation and presentation of findings)

- Research-ready dataset design (e.g. what data should we share? How should we prepare and structure this? What decisions should we make on what to include/exclude, how to amend raw data and add derived or contextual data?)

- Collaboration with our User Representation Panel, a group formed of seven different representatives from Civil, Criminal and Family justice system user support organisations
Data First – research fellowships (based on the 2020 criminal courts call)
Requirements of Early Career Researchers (ECR)

• Description of why the researcher is eligible to apply as an Early Career Researcher;
• Articulation of how the award of an ADR UK Research Fellowship grant will have a demonstrable impact on their trajectory to becoming an independent researcher;
• Inclusion of a programme of skill development activities to support their continued development; and
• A nominated mentor from their host institution who should write a supporting statement describing the support that will be provided.

Early Career Researchers are also encouraged to include wider programmes of training and capacity-building that develop their understanding or use of relevant data resources, methods or topic areas. This might be through seminars, workshops or other mechanisms, and should take account of training and events already being offered through other ESRC investments such as the National Centre for Research Methods (NCRM), UK Data Service and Doctoral Training Centres.
Key Outputs/ Expectations from the first fellowships

At least 2 accessible/user-friendly outputs for the ADR UK website and MoJ Data First landing page on gov.uk

Maximise research impact by giving busy policy leads short and easy to digest summaries
1 x 'Data Insight' (template to be provided)
1 x Your choice (findings; visualisations; policy briefings)
Plus a blog

At least 2 'Data Explained' (template to be provided)

Summarise and record your learnings about the data for the benefit of others including
1 x shorter summary within the funding period
1 x detailed document to be deposited within ONS SRS with 1 year of the funding period
Source of information to update the official User Guide and associated meta data

Academic publications

Supported but not at the expense of other research outputs
Commitment to Dual publication: easily digestible outputs for government policy makers and wider society AND papers submitted to journals

Public and stakeholder engagement

Method will depend on your research question and your own preference and expertise
Minimum: Data First User Representation Panel and ADR UK stakeholder events

ADR UK Communications and Engagement Team will be on hand to support Fellows

NOTE
All researcher outputs will go through a statistical disclosure control
The Applicant Journey of the first fellowships

- **AWARENESS**
- **APPLICATION**
- **FEASIBILITY & DATA OWNER APPROVAL**
- **PANEL DECISIONS**
- **PROJECT START**

**FORMS**
- Je-S
- DAP
- SAD

Look out for applicant workshops on future Data First research fellowship calls
This is a pilot project to enable researcher access to innovative datasets. At this point, three different forms are required to support three different decision making panels, each evaluating the same research project according to different criteria:

1. ADR UK’s Je-S form – to enable the Funders to make a robust decision on sponsoring the research, using public money wisely. Five criteria (see slide 17) are considered, from demonstrating use of administrative data, to scientific merit and value for money.

2. MOJ’s Secure Access to Data (SAD) form – to assure the Data Controller (HMCTS) that there is a legal basis to share the data, and that the request complies with all applicable data protection legislation [e.g. Data Protection Act 2018 and General Data Protection Regulation (GDPR)] to ensure safe, necessary and proportionate handling processing of data.

3. ONS/Research Accreditation Panel’s (RAP) form – to assure the UKSA that the requested access to de-identified data to accredited researchers under the Digital Economy Act is independent, consistent and transparent, and consists of three elements.
   i. ONS Feasibility Screening: RAP readiness assessment with feedback provided to applicant to increase chance of success at RAP (nb: Data Owner project approval must be sought separately – see SAD form above)
   ii. Research Accreditation Panel: final approval to access the data
   iii. NOTE: Applicants will need to be Accredited Researchers by the time of the RAP meeting (but outside the ADR UK research fellowship calls at the time of submitting RAP, see slide 20)

Note: Reuse of content between these forms is encouraged where possible
Je-S: The Funding Application

• The following are mandatory Je-S attachments for this call:
  o Case for support
  o Justification of resources
  o CV
  o Head of Department statement (including commitment to register for secure remote access – AOC)
  o Workplan
  o Mentor’s statement and CV (captured under ‘Other’ attachment type) – for ECRs only
  o Statement of Eligibility and Skill Development Programme – for ECRs only
  o Letters of Support from Project Partners (if applicable)
  o Final/Interim report of existing UKRI awards (if applicable)
  o List of Publications where references are cited in the proposal (if applicable).
  o Additional Letters of Support (optional)

NOTE: All proposals must be made on the Je-S Proposal Form – further information can be found in the Je-S proposal guidance document.
Je-S: The Funding Decision Criteria

- **Criteria 1**: Likelihood of demonstrating the potential of administrative data research for public policy impact;

- **Criteria 2**: Scientific Merit;

- **Criteria 3**: Feasibility to deliver in the timeframe;

- **Criteria 4**: Knowledge and experience; and

- **Criteria 5**: Value for money.

**FOR FUTURE CALLS PLEASE VISIT THE FUNDING FINDER UKRI PAGES:**
[https://www.ukri.org/opportunity/](https://www.ukri.org/opportunity/)
Data First – data access and research project approval

*for either funded or unfunded research projects*
ONS SRS researcher journey

Researcher contact the RSDA team

RSDA team provides training

Researcher completes the assessment

Successful applicants become DEA Accredited Researchers

Data Owner support for project (and agreement to supply) confirmed

Data linked, De-identified and ingested into ONS SRS

Project proposal reviewed by DEA Research Accreditation Panel

---

Safe People

Safe Projects

Safe Data

Safe Setting

Safe Output

Research serving the public good is published

RSDA team support the researchers when drafting their project proposal, with advice from experts across the ONS

SRS check outputs to safeguard the confidentiality of data subjects

RSDA team monitor the physical and technical security of the SRS

RSDA team support the researcher throughout his research proposal

Data linked, De-identified and ingested into ONS SRS

Project approved

---

Research serving the public good is published

SRS check outputs to safeguard the confidentiality of data subjects

RSDA team monitor the physical and technical security of the SRS

RSDA team support the researcher throughout his research proposal

Data linked, De-identified and ingested into ONS SRS

Project approved
Applying to use the data

- Researchers must be accredited under the Digital Economy Act 2017 (DEA).
- To become accredited you must have either:
  - an undergraduate degree (or higher) including a significant proportion of mathematics or statistics; or,
  - be able to demonstrate at least 3 years quantitative research experience.

And

- Undertake training
- Agree to be included on the public record
- Sign a declaration
Accessing the data

1. ONS Safe Rooms (Titchfield, London, Newport)

2. Safe Rooms hosted by other organisations
   (NISRA, Scottish Government, Perinatal Institute, Universities of Leeds, Liverpool and West of England)


4. Assured Organisational Connectivity (over 40 organisations including government organisations and universities)
   • Safe settings set up in academic or business premises after agreement with ONS, complying with ONS security standards and receiving annual certification
Technical Support or Data Enquiries

• Please contact datafirst@justice.gov.uk if you would like access to the fake synthetic data.

• To establish AOC agreement for your own organisation please contact: SRS.Connectivity@ons.gov.uk

• Please contact srs.customer.support@ons.gov.uk for any questions regarding:
  o Dataset enquiries,
  o Queries on the Application Form for Secure Access to Data.
How to Obtain Project Approval from Data Owners

• To obtain project approval from Data Owners (in this case it is HMCTS), you must complete our Application Form for Secure Access to Data.
  
  o We ask applicants to fill in six sections of the form providing information on applicant details, your project, and the data you wish to request. Guidance on how to complete is available on gov.uk

  o Applications are assessed against their feasibility, methodology, ethics, and public benefit.

  o There will be overlap with the ONS RAP form, and therefore we ask that you only submit your Application Form for Secure Access to Data once your RAP form has been screened.

• Your application will be considered by the HMCTS Data Access Panel (DAP).
Applying for accreditation

- Researchers can apply for accreditation through the Research Accreditation Service (RAS).

- This is also where you will fill in your RAP project form.
Part 1: Code of Practice

About the Code of Practice and Accreditation Criteria

1.1 Through Chapter 5 of Part 5 of the Digital Economy Act 2017 ("the Act") the UK Parliament has enacted legislation, applicable across the UK, that facilitates the linking and sharing of datasets held by public authorities (as defined in the Act) for research purposes. The provisions aim to broaden the capacity of research to deliver a number of direct and indirect public benefits, including the production of valuable new research insights about UK society and the economy.
Clearly justify why you require access to legally protected data. If applying for access to multiple datasets, explain how these will be used together to address your research aims.

Indicate the starting point for the research and/or modelling process. Include statistical tests that are likely to be used and the types of models to be employed.

The data from the small business surveys that we request will complement our use of administrative tax records and firm accounts data. Administrative records are informative because they contain information on the population (rather than a sample) of interest, and well-researched information on incomes and the use of tax allowances. However, they do not measure other useful information contained in surveys, such as type of financing used, plans for future investments, reasons why certain decisions are taken etc. By combining information from the two sources, we will be able to get a richer picture of UK small business behaviour.

We will bring analysis from the different data sources by matching aggregate research outputs through the use of an econometric model, and methods of data linkage. We are in the process of building a dynamic model of business owner decisions, including whether to start a business, whether to incorporate, and how much to invest, and when to exit. This methodology will allow us to explicitly account for the dynamic aspects of individual’s decisions.

We will estimate the model by matching moments in the model with the same moments observed in the data. Moments are features of the data, for example, the mean, standard deviation, or other transformations of the data. In this application, we will use moments that are informative about the behaviour that we are trying to analyze e.g. the amount of investment, propensity to incorporate, distributions of profit and taxable income. We can use moments from multiple data sources (including the business survey data requested here) to estimate the model, without having to link the data at the observation level. Moments that we anticipate will be useful from the business survey data include: plans to launch new products, R&D expenditure, type of finance used, among others.

We also plan to exploit policy features and reforms over our period of study – for example, the introduction of the additional rate of tax on incomes above £150,000, the withdrawal of the personal allowance for those earning more than £100,000, changes to investment allowances, and corporate tax reforms. We will use this variation to help identify how small business owners respond to various parts of the tax system.

Once we have estimated the model, we can perform counterfactual analysis to look at the effect of various policy reforms. For instance, what would be the implications of taxing capital income at the same rates as labour income, or removing the flexibility for business owners to shield taxable income across years. This will allow us to provide important evidence on the potential costs and benefits of various policies that target small business owners.

Specify the methods of analysis you plan to use. If you are proposing a non-standard methodology, provide methodology references. The level of detail provided should enable an experienced researcher to understand the type of model, how bias is being addressed and what is likely to constitute an important effect.

Include an explanation of how your methodological approach will answer the research questions set out in the project.
Data First - potential research topics
Understanding the 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.
Criminal Courts data

What we can (at least partially) address:
• How do individuals in the courts and tribunals system vary by, for example, protected characteristics, socio-economic or demographic background? How do these characteristics compare across jurisdiction and case type?
• Are there any differences in outcomes based on these factors?

What we hope to be able to address:
• How do individuals within the criminal court system move through it, interacting with multiple (often disparate) agencies, services, interventions and policies? How we can develop, target, sequence, and measure the outcomes of, interventions and support? (multiple datasets)

What we can’t address:
• Which groups struggle most to resolve their justice problems, either through inaction or difficulty accessing the justice system and wider support services?
• How prevalent is repeat victimisation and who does it affect?
• How can we better understand flows into the courts and tribunals system, reasons for entry, and the impact of external organisations and their activities?
• How can we better understand the opportunities and impact of the use of digital technology on those engaging with the courts and tribunals system?
Prisoner custodial journey data

What we can (at least partially) address:

• What is the characteristic profile of repeat occupiers of the prison system?
• Are certain release types more likely to deter an offender from re-entering the system?
• Are offender characteristics such as ethnicity and sex associated with a variation in custodial reconviction?

What we hope to be able to address:

• The link between imprisonment types and the likelihood to reoffend (link to courts and probation data).
• How interaction with other government services impact the likelihood of an individual entering the prison system.

What we can’t address:

• How effective are different prison types, categories, functions, such as reception in meeting their core objectives?
• How can services and therapeutic interventions, such as training, peer and family relationship support, drug and addition services, improve post release outcomes?
• What works to reduce levels of self harm and self inflicted deaths in prison, for different individuals and groups?
MoJ-DfE data

What we can (at least partially) address:
- Educational characteristics and offending, for example:
  - Risk/protective factors and predictors of involvement in different types of criminal activity
  - Links between educational characteristics and outcomes in terms of disposals
  - Associations between particular interactions with the education system and offending
- Early interventions and reduced reoffending, for example:
  - Evaluating early intervention projects
  - Identifying the characteristics of children at risk of offending at an early stage
  - Evaluating ‘what works’ for different groups in the justice system

What we hope to be able to address:
- The influence of early education/child social care experiences, for example:
  - Links between spending time in the care system as a child and involvement in the youth/criminal justice system
  - Effects of repeated private law proceedings on school performance
- Impact of interactions with criminal justice system on educational attainment, for example:
  - What happens to young people’s educational engagement post-release from youth justice system?

We can’t address:
- Identifying individual needs
- Assessing quality of experiences
- Experiences of victimisation
Cross-cutting research enabled with further (currently out of scope) data shares

Sources that could be used in combination with Data First data to address questions and support multidisciplinary approaches:

- Linking with administrative data of other government departments and agencies (Individual and Contextual levels)
- Linking data on interventions for the purposes of evaluation (Individual and Contextual levels)
- Matching with Census and/or Ordnance Survey data (Contextual - geographical level)
- Matching with Survey data (Contextual - geographical or other group level)
- Matching with Digital Open Access (Big) data (Contextual - geographical or other group level)
- Mixed methods approaches (whilst safeguarding individual confidentiality)

Potential cross-cutting research questions:

- How does housing and employment stability of ex-prisoners impact upon reoffending – length of time from release and offence seriousness?
- How, defendants’ protected characteristics, historic and current contextual socio-economic differences impact upon age of onset, frequency, seriousness, progression and duration of interactions with the justice system?
- How, family circumstances and family justice outcomes impact upon children’s and young people’s health and education outcomes. Do these, and if so how, in turn impact upon children’s and young people’s likelihood and frequency of victimisation and/or offending? Can we map the association and inter-connections of these mediating factors?
Join MoJ Academic Network and look out for opportunities to get involved with the broader engagement programme – seminars, workshops, knowledge exchange events. Look out for the next Data First academic seminar.

In preparation for research bids:

- Request synthetic (fake) criminal courts datasets: datafirst@justice.gov.uk

Find out about future Data First Research Fellowships Scheme and/or UKRI Funding Finder.

Contact MoJ if you have research findings that address any of MoJ ARI questions, or any queries about research engagement/collaboration: Evidence_partnerships@justice.gov.uk