ADR England Flagship Datasets

Showcasing a collection of linked administrative datasets available for research
July 2023

Administrative data is
an invaluable resource
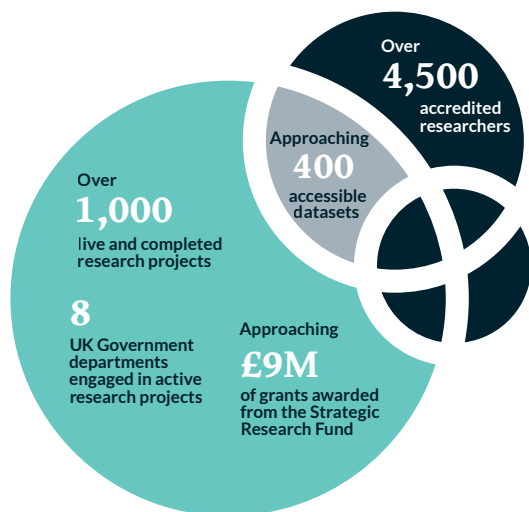for public good.
We're using it.

Contents

Introduction

This brochure showcases a collection of recently linked administrative datasets available for research. It is designed to introduce researchers to the wealth of linked public sector data available covering a wide range of topics and policy areas.

What is ADR UK?

ADR UK (Administrative Data Research UK) is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve lives.

Over **4,500** accredited researchers

Approaching **400** accessible datasets

Over **1,000** live and completed research projects

8 UK Government departments engaged in active research projects

Approaching **£9M** of grants awarded from the Strategic Research Fund

Administrative data is information created when people interact with public services, such as schools, the NHS, the courts or the benefits system, and is collected by government and other public sector organisations.

Education records

Court records

Health records

Employment & earnings

Census responses
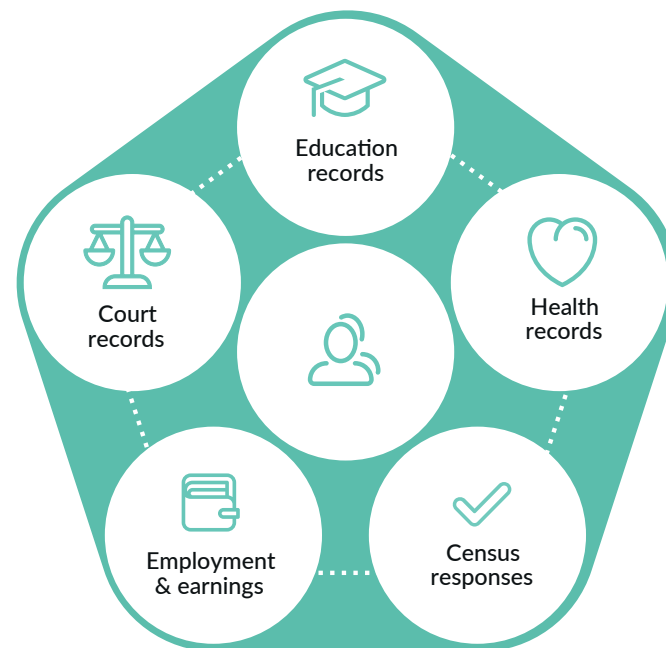
Over the past four years, ADR UK has made significant advances in making de-identified administrative data easier for researchers to access securely, and there is now an impressive collection of new flagship datasets available through our trusted research environments across the UK. In addition, ADR UK is developing further resources including explainer videos, training courses, synthetic data which can be used to develop and test code, and communities of practice in which researchers can share ideas.

The linked administrative datasets featured in this brochure form an exciting first collection of rich, curated data fit for use by researchers across all career levels. ADR UK's commitment to provide extra resources will add to the appeal of these datasets and pave the way for further collections to become available.

Read how ADR UK ensures data is used ethically and responsibly.

What are ADR England Flagship Datasets?

ADR England flagship datasets are a selection[1] of research-ready datasets[2] which ADR UK is coordinating activity and funding towards to drive up their research use and public value.

Our flagship datasets meet the following criteria:

1. Linked administrative data: Our flagship datasets must link administrative data or be linkable within a government department, or between two or more government departments. The linkage may be the result of an ADR UK-funded project or a government led-activity.

2. National datasets: Our flagship datasets should have coverage of at least one UK nation, or wider (e.g. England and Wales, GB or UK).

3. Of significant research value: Covering ADR UK's eight core strategic research themes - from environment and health to growing old - flagship datasets can address big societal issues and cross-cutting themes that transcend individual government department remits. These datasets have the potential to unlock considerable public good by helping to fill evidence gaps and address pressing policy needs of national importance.

4. Available to the wider academic community: All flagship datasets must be accessible to all accredited researchers via an ADR UK trusted research environment with approved legal gateways for research. This includes the datasets being findable in a catalogue with a minimum level of publicly available documentation including a minimum standard user guide with potential research questions; a data dictionary; and a route for seeking additional information.

5. Of wide appeal: We anticipate our flagship data being of notable interest to a wide diversity of researchers and research users who will be integral to ADR UK's mission to enable vital research that has the potential to lead to better informed policy decision and more effective public services.

---

[1] Other research-ready administrative datasets are available via ADR UK trusted research environments.

[2] ADR England's perspective on research-ready data is informed by this systematic review and roundtable discussion
   a) What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature
   b) What is research-ready data: A roundtable report

To drive up the research use and public value of our flagship data we aim to develop these research-ready datasets as meaningful and useable research resources and increase the number and capability of researchers using these complex linked administrative datasets.

ADR UK commits to working directly and in partnership with government and academia to provide a wrap-around offer for our flagship data.

Activities and support provided will be prioritised to address the specific needs for each dataset and could include:

- Awareness raising (e.g. signposting and promotion)
- Documentation (e.g. user guides, research tools, synthetic data[3], code sharing)
- Curation and enhancement (e.g. feedback loop between researchers and data owners to improve research utility over time)
- Training (e.g. courses on how to use the data)
- Research opportunities (e.g. ADR UK research funding including PhD studentships and secondments, stimulation and promotion of other research funding opportunities, developing research agendas)
- Public engagement (e.g. with user representation groups)
- Research and impact support (e.g. code sharing, citation, case studies, expert help)
- Community development (e.g. online user groups, themed engagement).

The current collection of ADR England flagship datasets are:

- Education and Child Health Insights from Linked Data – England (p8)
- Longitudinal Education Outcomes - England (p10)
- Growing Up in England (p12)
- Grading and Admissions Data for England (p14)
- Annual Survey of Hours and Earnings linked to 2011 Census - England and Wales (p16)
- Data First (p18).

More ADR England flagship datasets, including **AD|ARC**, are under construction, and will be announced in due course.

---

[3] Low fidelity synthetic data will be developed in collaboration with data owners to support researchers to scope their research question or develop code prior to access to the real data, as well as to support the provision of training courses on how to us the data. This form of synthetic data does not preserve any statistical relationships between variables, and does not contain any real data.

Education and Child Health Insights from Linked Data - England

Summary

The Education and Child Health Insights from Linked Data (ECHILD) database includes linked records for around 20 million children and young people, which will be used to better understand how education affects children's health and how health affects children's education.

Overview of the component datasets and linkages

Timeframe:

This is a population-based cohort of children & young people born between 1 September 1984 and 31 August 2022.

Coverage: England only

Size: ~ 20 million records

Core documentation

User guide
Data dictionary

Education and Child Health Insights from Linked Data

Hospital Episode Statistics

- Admitted patient care
- A&E / emergency care dataset
- Outpatient care
- Mortality data
- Births notifications / registrations
- Maternity services dataset
- Community services dataset
- Mental health services dataset

National Pupil Database

Pupil characteristics
- School census (pupil level)
- Early years census
- Alternative provision census
- Pupil Referral Unit census

School characteristics
- School level census
- Education setting census

Educational outcomes
- Early years foundation stage profile
- Key stages 1-5
- Exclusions
- Absences
- National Client Caseload Information

Social care
- Children looked after return
- Children in need census

Please note that the included datasets have varying timeframes: see user guide for more detailed information about dates available

Examples of research questions the data could be used to address

The ECHILD database will facilitate research to improve policymaking for children's health, education and wellbeing across a range of health and social science disciplines. For example, the ECHILD database could help to answer questions such as:

- How can we improve health and education for children with chronic physical or mental health conditions in England?

- How can we better understand the health-related drivers of educational outcomes in England?

- How can we improve the quality and equity of health and education provision in England?

Examples of research done using the data:

- COVID-19-related school closures and patterns of hospital admissions with stress-related presentations in secondary school-aged adolescents: weekly time series

- Hospital admissions for stress-related presentations among school-aged adolescents during term time versus holidays in England: weekly time series and retrospective cross-sectional analysis

- Gestational age at birth, chronic conditions and school outcomes: a population-based data linkage study of children born in England

- Changes in adolescents' planned hospital care during the COVID-19 pandemic: analysis of linked administrative data

- What makes administrative data research-ready? A systematic review and thematic analysis of published literature

- Reductions in hospital care among clinically vulnerable children aged 0-4 years during the COVID-19 pandemic

Other supporting information and resources:

- ECHILD data visualisation: An overview of records and datasets - A data visual to explore the different types of de-identified records within the ECHILD database.

- Data Resource Profile

- Linking education and hospital data in England: Linkage process and quality

- ECHILD User Guide

- Data Resource: The National Pupil Database (NPD)

- Data Resource Profile: Children Looked After Return (CLA)

- Cohort profile: Children in Need Census (CIN)

- UK Government tool to explore data in the National Pupil Database

- Hospital Episode Statistics Data Dictionary

- Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC)

- Community Services Dataset Data Dictionary

- Mental Health Services Data Dictionary

- Maternity Services Data Dictionary

More information:

Find out more on the ECHILD website. You can contact the programme team at ich.echild@ucl.ac.uk

Longitudinal Education Outcomes - England

Summary

The Longitudinal Education Outcomes (LEO) dataset is a world-leading data product created by the Department for Education (DfE). It contains de-identified information on the characteristics, education, employment, benefits, and earnings of members of the British public. This is a unique source of information, with the potential to provide transformative insight and evidence on the longer-term labour market outcomes and educational pathways of (currently) around 38 million individuals. The purpose of the LEO dataset is to enable the UK to gain transformational insights about pathways within and after education, to enhance the life chances of current and future learners through the enhancement of evidence-based decision making.

Overview of the component datasets and linkage

LEO is a relational database comprised of a range of tables and access is restricted on a needs basis.

Timeframe: Jan 1995 – Aug 2021
Update frequency: At least annual
Coverage: England only
Size: ~38 million records

Longitudinal Education Outcomes (LEO) component datasets and linkage

UCAS data

National Pupil Database
School and early years data

Individual Learner Record Data
College and apprenticeship data

Inter-Departmental Business Register

Higher Education Statistics Agency data
Higher education data

LEO data (HMRC and DWP)
Earnings, benefits, and tax data

Covid-19 furlough data

Core documentation

User guide: The LEO user guide is available to [accredited researchers](#) within the [Office for National Statistics (ONS) Secure Research Service](#) and the [LEO Research Community](#).

Data dictionary: Information about variables within the current version of the LEO Standard Extract can be found in the [LEO variable request form](#).

Examples of research questions the data could be used to address

- Does attendance at a school with a "Good" Ofsted rating have an impact on a person's labour market outcomes?

- Are pupil referral units (for pupils who have been excluded from school) effective in improving outcomes for young people?

- Does being in the care system lead to poorer labour market outcomes after controlling for other demographic and education factors?

- Which education factors have a positive impact on labour market outcomes of school leavers with poor attainment?

- Does coming from a lower socioeconomic background lead to poorer labour market outcomes after controlling for other demographic and education factors?

- What is the link between location, labour mobility and education and labour market outcome? At similar education levels, do those who are more mobile have better outcomes?

Examples of research done using the data

The [LEO Research Community](#) library stores a list of published research. Examples from the broad and largely un-tapped field of potential LEO research include:

- Research into [post-16 pathways](#) (journeys of individuals following their state education).

- Research into social mobility and outcomes for students receiving free school meals in England.

Other supporting information and resources

Webinar: The Edge Foundation recently [published a recording](#) showcasing analyses using the LEO dataset.

Guidance & frequently asked questions on how to access the LEO dataset are available from the [ONS Secure Research Service](#).

Stakeholder report: a [recent report](#) covering key messages from stakeholders on the research priorities for LEO.

LEO Library: The LEO Research Community supports data preparation and analysis and hosts a growing library of resources and information. To access the LEO Community you will need a Knowledge Hub account and UK Statistics Authority (UKSA) researcher accreditation.
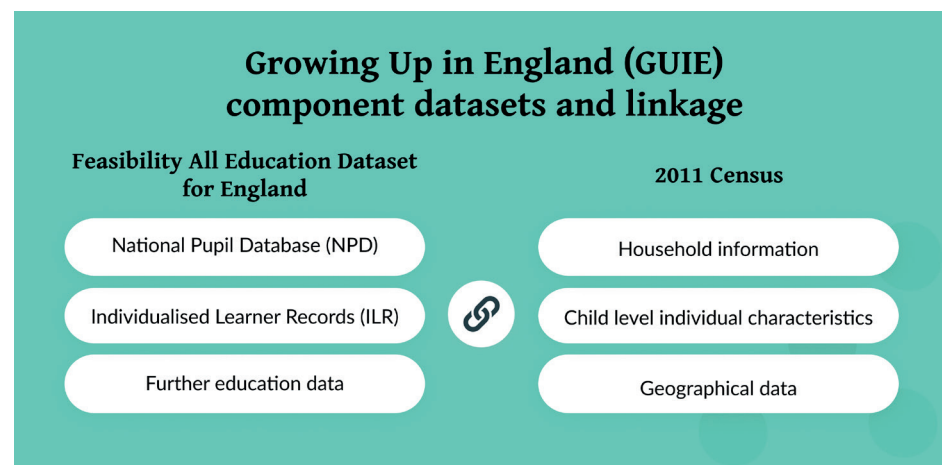
More information

Find more information about the [LEO data programme](#) and the [National Pupil Database](#) from gov.uk. The Higher Education Statistics Agency (HESA) provide information and [user guides for Higher Education data](#) on their website. For specific queries about the LEO dataset, please contact leo.programme@education.gov.uk.

Growing Up in England

Summary

ADR UK, the Office for National Statistics (ONS) and the Department for Education (DfE) are working together to make better use of existing data about children and young people in England. The Growing Up in England (GUIE) dataset will enable a more comprehensive understanding of how factors such as household composition and geography shape outcomes. This will allow researchers to shine light on what does and doesn't work in supporting children and young people to thrive and give policymakers valuable evidence for developing services that work better for all. There is intention for a future build to include 2021 Census, a review and update of the education component of the data, and linkage reconstruction.

Overview of the component datasets and Core linkage

Growing Up in England (GUIE) component datasets and linkage

Information contained in the Growing Up in England dataset
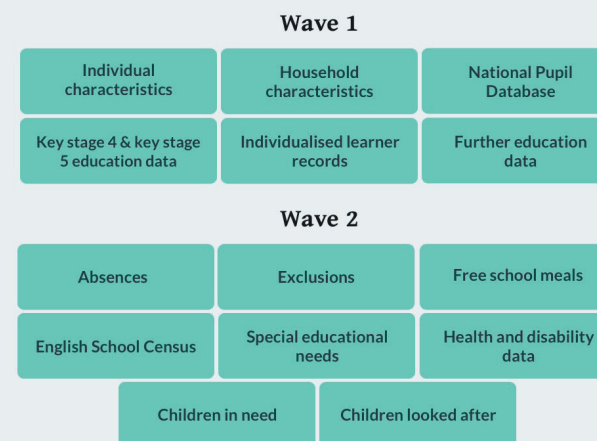
Timeframe:

Wave 1: Academic years 2001/02 – 2014/15
(additional wave 2 data has been requested back to 2001/01 where available. This is available in the ONS Secure Research Service).

Wave 2: Academic years 2010/11 – 2014/15

Update frequency: Static

Population: Level 1, 2 and 3 records for children and young people in predominantly state-funded schools and colleges (Census 2011-linked dataset includes individuals aged between 14 and 29 years on 31 August 2015).

Coverage: England only

Size: ~7 million records

Core documentation

User guide: A user guide will be made available and will be accessible through the ONS SRS Metadata Catalogue. The user guide will also signpost to an accompanying ONS methods publication for Growing Up in England.

Data dictionaries:

- Wave 1
- Wave 2 (Children in Need, Children Looked After)
- Wave 2 (Exclusions, Absences, and English School Census)

Growing up in England publication: the GUIE publication was released in October 2022 and provides information on the datasets that make up GUIE, and how the data was linked.

Other supporting resources

AEDE information: read more about the feasibility All Education Dataset for England (AEDE) Note that the AEDE data included in Wave 1 GUIE differs in some respects to what is described in this publication, e.g., HESA data within AEDE is not included in Wave 1 GUIE.
2011 Census information: find out more about the 2011 Census.

Examples of research questions the data could be used to address

- Is there a relationship between household characteristics and educational attainment?
- What are the underlying drivers of geographical differences in educational attainment?

Examples of research done using the data

AEDE feasibility information: Read more about the feasibility study that linked the All Education Dataset for England (AEDE) to the 2011 Census in this study from the ONS. Note that the AEDE data used in GUIE Wave 1 and 2 only differs in some respects to the data within this feasibility study.

Case study: ADR UK have funded a study to explore the potential of the data and address an evidence gap on educational experiences of Roma, Gypsy, and Traveller children.

Who are the children entering care in England?: ONS research which uses GUIE Wave 2 data to explore the characteristics of children entering care in England.

More information

You can find more information about the GUIE dataset on the ONS Secure Research Service Metadata Catalogue and the GUIE project page on the ADR UK website. Read more from the ONS about the potential for administrative data in understanding qualifications in a blog and a feasibility study.

Grading and Admissions Data for England

Summary

The Grading and Admissions Data for England (GRADE) is a joint open data initiative by Ofqual, Ofsted, the Department for Education (DfE), and UCAS. Its main objective is to provide accredited researchers with access to data to conduct independent research on the educational, assessment, and admission systems in England, and to facilitate internal analysis.

It will allow researchers to conduct independent high-quality evaluation of the judgements made in awarding grades in 2020, ensuring that lessons are learned. More broadly, the GRADE data will enable research to enhance the quality of the assessment system and produce evidence to inform future education policy, particularly around the fairness of methods for measuring.

Overview of the component datasets and linkage

Timeframe: Jan-2017 – Jan 2021
Update frequency: Every two years
Coverage: England only
Size: ~ 4 million records

Grading and Admissions data for England

Ofqual
Qualifications dataset
Centres dataset*
Grade boundaries dataset
Look-up tables*

Department for Education
Exam results dataset
Student dataset
Prior attainment dataset
Census dataset
Look-up tables*

UCAS
Apply qualifications dataset
Applications dataset
Applicant dataset
Look-up tables*

Data tables marked with * will not be shared for data protection reasons

GRADE data at a glance

| 2020 CAGs and calculated grades | GCSE and A level grades 2017-20 |

| Students' characteristics including prior attainment | Predicted grades for university applications |

Department for Education · ofqual · UCAS

ADR UK
Data-driven change

Core documentation

User guide: A data sharing Framework provides a high-level description of the datasets and sets out how to access the data. A more detailed user guide will be published soon.

Data dictionaries: Data specifications are available from Ofqual, UCAS, and DfE (NPD) on .gov.uk.

Other supporting information and resources

Synthetic data: A low-fidelity synthetic dataset with the same format and structure of the real data, but containing only made-up information based on simulated data, is available by request. Please email data.sharing@ofqual.gov.uk.

Examples of research questions the data could be used to address

- How does the standardisation model introduced in summer 2020 stand up to scrutiny?

- What was the impact of awarding approaches on students and schools and/or colleges?

- What is the relationship between the 2020 grading data and other measures of attainment (validity related questions)?

Examples of research done using the data

Non-journal publications: Two research reports looking at the awarding of centre assessment grades for GCSEs, AS and A levels in summer 2020.

ADR UK Research Fellowship: ADR UK is funding a Research Fellow to carry out pathfinder research using the GRADE data linkage (call now closed). Keep posted about their upcoming research activities and outputs on the ADR UK website or sign up for updates.
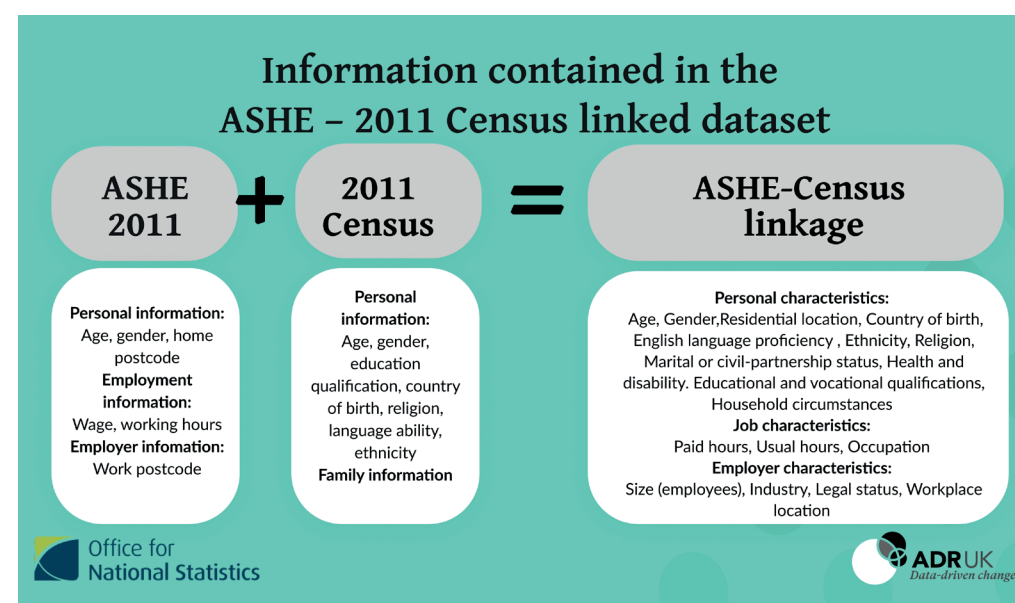
More information

Find more information about the data and its research potential in the ADR UK Fellowship funding call specification (note that this opportunity has now closed but that the information provided about the data is still relevant), the ONS Metadata Catalogue, and the GRADE website.

Annual Survey of Hours and Earnings linked to 2011 Census - England and Wales

Summary

The **Wage and Employment Dynamics (WED)** data linkage project is a collaboration led by researchers from the University of the West of England, aiming to better understand wage inequalities in Britain. Phase one of the project is made up of two parts: 1) Linkage of the Annual Survey of Hours and Earnings (ASHE) data to 2011 Census; and 2) the development of Stata code to enrich the ASHE data. This will allow for insight into the dynamics of wage and employment issues, and how characteristics such as gender, disability, and ethnicity influence these. The ASHE – 2011 Census linkage is currently available for access through the ONS Secure Research Service.

Information contained in the ASHE – 2011 Census linked dataset

ASHE 2011 **+** 2011 Census **=** ASHE-Census linkage

Personal information: Age, gender, home postcode
Employment information: Wage, working hours
Employer information: Work postcode

Personal information: Age, gender, education qualification, country of birth, religion, language ability, ethnicity
Family information

Personal characteristics: Age, Gender, Residential location, Country of birth, English language proficiency , Ethnicity, Religion, Marital or civil-partnership status, Health and disability. Educational and vocational qualifications, Household circumstances
Job characteristics: Paid hours, Usual hours, Occupation
Employer characteristics: Size (employees), Industry, Legal status, Workplace location

Office for National Statistics

ADR UK
Data-driven change

Overview of the component datasets and linkage

1) ASHE – 2011 Census linked dataset
2) Enriched ASHE code

The Wage and Employment Dynamics team has enriched the core ASHE dataset by adding new variables such as minimum wage rates and survey dates and carrying out methodological reviews on the data and sampling. They have also created new weights which account for attrition and selection effects. These are available from 2004. Access to the code will enable research into the dynamics of wage and employment issues, from labour market entry, through job mobility and career progression to retirement. The code created by the WED team will be available within the ONS Secure Research Service on a shared accessible drive. Please note that this code is not part of the ONS ASHE official Statistics output.

Timeframe: 1997 - 2021

Update frequency: Annual

Population: ASHE population spine (attained sample is approximately 0.66% of the entire workforce)

Coverage: England and Wales*

Size: ~175,000 records per year

* Northern Ireland: The Earnings and Employees Study (EES) 2011 for Northern Ireland links together variables from the Annual Survey of Hours and Earnings (ASHE) 2011 with variables from the Census of Population and Housing 2011, and Capital Value data from the Land and Property Services. The dataset takes the form of one pre-linked table.

Core documentation

User guide for ASHE-2011 Census and Code for Enriched ASHE Quick Start Guide: Both these documents can be found on the **Wage and Employment Dynamics website**.

Data dictionary: Information about the variables will be available from the **ONS Secure Research Service Metadata Catalogue**.

Other supporting information and resources

Training to use the data: See the **upcoming training opportunities on the data**, as well as recorded webinars from previous training sessions.

> **Tim Butcher, Chief Economist, Low Pay Commission:**
>
> *"This project addresses weaknesses in our evidence base – improving the quality of longitudinal earnings data and extending coverage to a broader range of characteristics – that should enable researchers to give new and innovative insights into the wage and employment dynamics of the lowest paid."*

Examples of research questions the data could be used to address

- How do people's earnings change throughout their career? And how does this differ depending on characteristics such as gender, disability, or ethnicity?

- What are the patterns or characteristics of those who do and do not progress out of low paid employment?

- What is the relationship between migration and the labour market?

- What role do employers play in wage inequality?

Examples of research done using the data

Wage and Employment Dynamics **publish research and reports on their website**.

More information

Find out more about the **Wage and Employment Dynamics** project and data. Keep posted about their upcoming events, outputs and activities on the **ADR UK website** or **sign up for updates**. Or **get in touch** with the team directly via their website.

Ministry of Justice Data First: Datasets and Linkages

Summary

Data First is a ground-breaking data linking, academic engagement and research programme led by the Ministry of Justice (MoJ) and funded by ADR UK. The programme is unlocking the potential of the wealth of data already created by MoJ, by linking administrative datasets from across the justice system and beyond and enabling researchers within government as well as approved academics to access the data in an ethical and responsible way.

Data First datasets, key features, and linkages

- **Magistrates' courts defendant data:** Adults and young people appearing as defendants before the magistrates' courts in England and Wales

- **Crown Court defendant data:** Defendants appearing before the Crown Court in England and Wales

- **Prisoner custodial journey data:** People in custody in prisons and young offenders' institutions in England and Wales, including on remand

- **Probation data:** Adults under supervision of the probation service in England and Wales while serving a community sentence, or released from prison on licence or on parole

- **Family court data:** Adults and children involved in cases to do with child custody, care arrangements, divorce or adoption

- **Civil court data:** People (and companies) involved in civil cases as claimants or defendants, including landlords and tenants in housing cases.

Information about research priorities

The MoJ Areas of Research Interest 2020 sets out the department's critical evidence gaps that could be explored using the Data First datasets.

Data use register

Find out how researchers are using the Data First datasets in the MoJ data requests log, which records all approved external data requests.

More information

More information about the Data First datasets and linkages can be found on pages 14-19 of this brochure, as well as on the ONS Metadata Catalogue, the ADR UK website, and the Data First website. For more specific data, linkage, or metadata queries you can also email datafirst@justice.gov.uk.

> **Lord Justice Fulford, Vice President of the Court of Appeal (Criminal Division) and previous Judicial Lead on Data:**
>
> *"I consider this to be an extremely important project, which has been established according to all of the principles that necessarily apply to work of this kind. In a sentence, it will enable critical research to be undertaken by accredited professionals, whilst protecting the identities of those involved in the individual cases."*
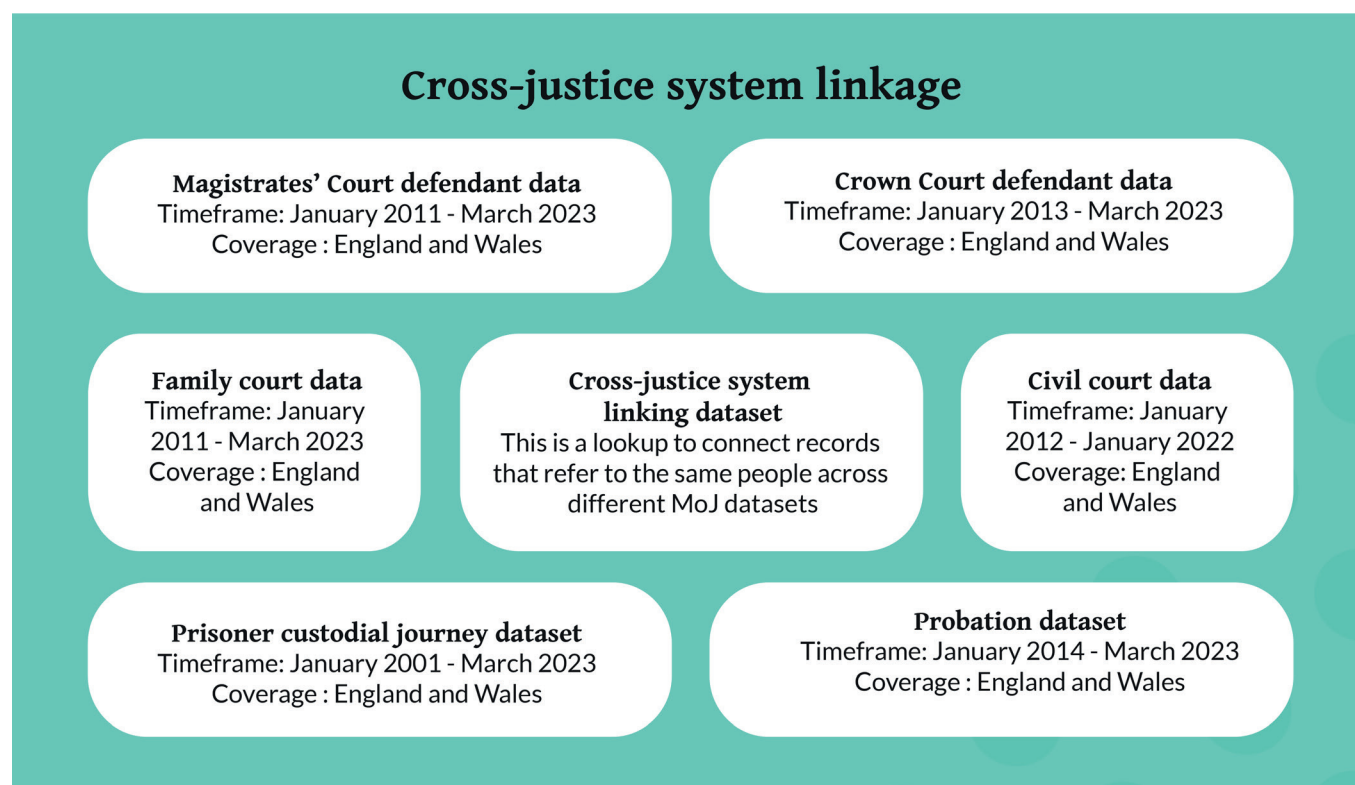
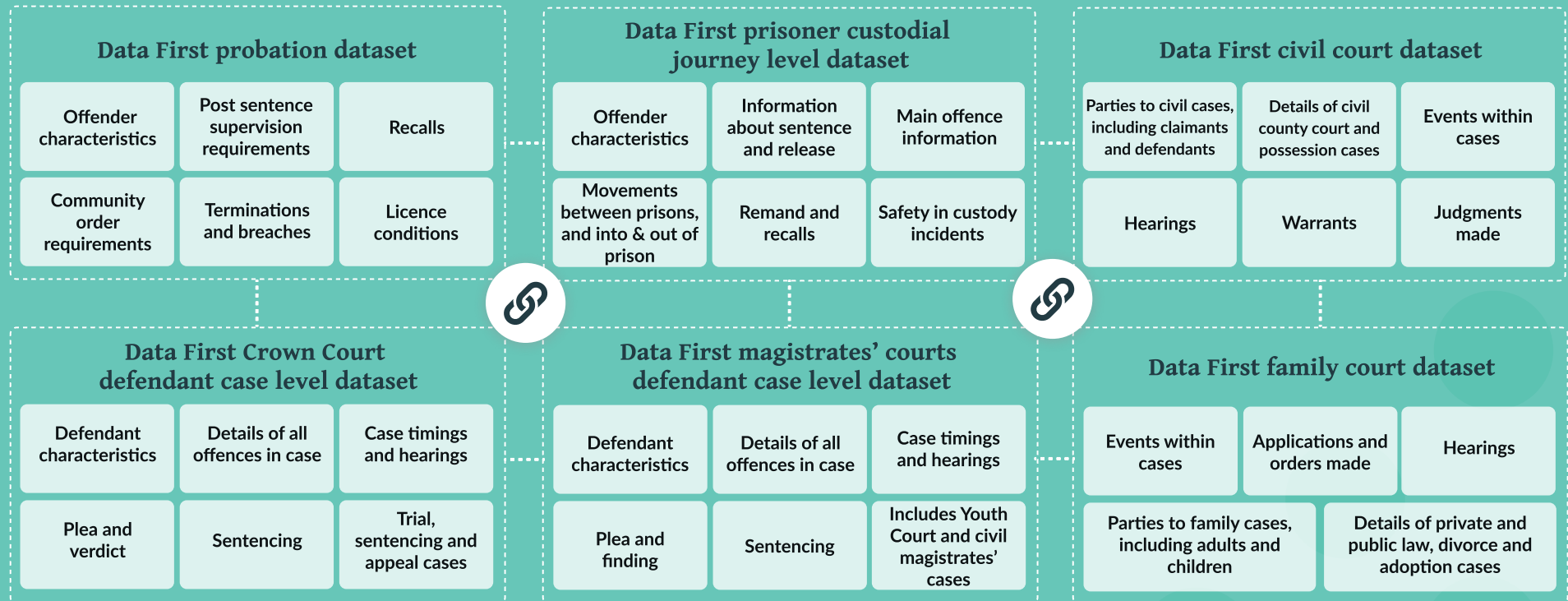# Data First: Cross-Justice System - England and Wales

Summary

This dataset connects Ministry of Justice data from the civil and family courts with different areas of the criminal justice system. This presents an unprecedented opportunity to understand cross-cutting questions about the overlaps and intersections between users of services in different justice jurisdictions.

Data is available on an individual person- and case- level, from the start of criminal prosecutions in the magistrates' courts and Crown Court, through to periods spent in prison custody or under supervision of the probation service. Additionally, data is available on adults and children involved in family court cases to do with child custody, care arrangements, divorce or adoption, as well as people (and companies) involved in civil cases as claimants or defendants. This will enable insights on end-to-end user journeys across the criminal, family and civil justice systems.

Overview of the component datasets and linkage

Cross-justice system linkage

Magistrates' Court defendant data
Timeframe: January 2011 - March 2023
Coverage : England and Wales

Crown Court defendant data
Timeframe: January 2013 - March 2023
Coverage : England and Wales

Family court data
Timeframe: January 2011 - March 2023
Coverage : England and Wales

Cross-justice system linking dataset
This is a lookup to connect records that refer to the same people across different MoJ datasets

Civil court data
Timeframe: January 2012 - January 2022
Coverage: England and Wales

Prisoner custodial journey dataset
Timeframe: January 2001 - March 2023
Coverage : England and Wales

Probation dataset
Timeframe: January 2014 - March 2023
Coverage : England and Wales

Information contained in the Ministry of Justice
Data First: Cross-Justice System - England and Wales
linked dataset

Data First probation dataset

| Offender characteristics | Post sentence supervision requirements | Recalls |
| Community order requirements | Terminations and breaches | Licence conditions |

Data First prisoner custodial journey level dataset

| Offender characteristics | Information about sentence and release | Main offence information |
| Movements between prisons, and into & out of prison | Remand and recalls | Safety in custody incidents |

Data First civil court dataset

| Parties to civil cases, including claimants and defendants | Details of civil county court and possession cases | Events within cases |
| Hearings | Warrants | Judgments made |

Data First Crown Court defendant case level dataset

| Defendant characteristics | Details of all offences in case | Case timings and hearings |
| Plea and verdict | Sentencing | Trial, sentencing and appeal cases |

Data First magistrates' courts defendant case level dataset

| Defendant characteristics | Details of all offences in case | Case timings and hearings |
| Plea and finding | Sentencing | Includes Youth Court and civil magistrates' cases |

Data First family court dataset

| Events within cases | Applications and orders made | Hearings |
| Parties to family cases, including adults and children | Details of private and public law, divorce and adoption cases | |

Ministry of Justice

ADR UK
Data-driven change

Core documentation

The Data First user guide and data catalogues can be found on the Data First gov.uk webpage. The datasets can also be found in the ONS Secure Research Service Metadata Catalogue.

Other supporting information and resources

Synthetic data: 1000 row, low-fidelity synthetic datasets are available for the magistrates' court and Crown Court datasets. These can be requested by emailing datafirst@justice.gov.uk.

'**Data Explained**': Data Explained are produced by ADR UK Research Fellows discussing the first-hand experiences on the utility and limitations of the data for their research projects, for example, on racial bias in court cases. These can be found on the ADR UK project page, and more will be published as they become available.

Information about research priorities

Areas of Research Interest (ARI): The Ministry of Justice ARI 2020 sets out the department's critical evidence gaps.

User Representation Panel report: A report on the Data First User Representation Panel roundtable held 31 January 2022 on the research interest and potential of the criminal justice data. A User Representation Panel report on the cross-justice system linked dataset will be available soon, check the project page for updates.

Examples of research questions the data could be used to address

- What links are there between people's cases in the civil and family courts, and their interactions with the criminal justice system?

- How can we better understand how problems interact and reinforce each other, and how people move through different jurisdictions across the system as they attempt to resolve them?

- Who are the 'repeat' users across the justice system? Does involvement in other areas of the justice system influence criminal (re-)offending? What works to enable effective outcomes across the justice system? For example, are some children in family court cases at an increased risk of (early) criminal offending compared to the general population?

- How do individuals in the justice system vary and compare across jurisdictions in the system?

Examples of research done using the data

Non-journal article publications: An exploratory analysis of the magistrates' and Crown Court defendant datasets has been published by the Ministry of Justice.

ADR UK Research Fellows' outputs: Four ADR UK Research Fellows have undertaken pathfinder research using the linked magistrates' and Crown Court defendant datasets. Their publications including blogs and Data Insights can be accessed via the ADR UK project page. Two ADR UK Research Fellows are using the probation and criminal justice system linked datasets. Sign up to our newsletter for updates.

More information

More information about the cross-justice system datasets and linkages can be found on the ADR UK website and on the MoJ Data First website.

Data First: Family Court linked to Cafcass and Census 2021 - England and Wales

Summary

The Data First family court - Cafcass linked dataset came about from a data sharing agreement between the [Secure Anonymised Information Linkage (SAIL) Databank](#), Ministry of Justice and Cafcass (Children and Family Court Advisory and Support Service). The de-identified dataset contains information on adoption, divorce, family law act, public law and private law cases, and their legal outcomes, in England and Wales. Public law cases include local authority intervention to protect children, such as care or supervision proceedings, and private law cases include parental disputes concerning the upbringing of children, for example arrangements for contact, access or custody.

Census 2021 results contain data and supporting commentary on the population and households in England and Wales. This has been made available in the SAIL Databank for standalone access and linkage to other datasets. The Family Court - Cafcass linked dataset is the first ADR England flagship dataset to be made available with Census 2021 data.

Overview of the component datasets and linkage

Information contained in the Data First family court-Cafcass linked dataset

- Unique ID for party
- Divorce, private law and Family Law Act cases
- Public law and adoption cases
- Family law applications and legal outputs
- Marriage and divorce characteristics
- Personal-level information
- Case events, timeliness and number of hearings
- Legal representation

Timeframe:

Family court data: January 2011 – March 2023

Cafcass: April 2007 – January 2023

Coverage: England and Wales

NB: Cafcass, Cafcass Cymru, and the linked Cafcass – family court data are currently only available to access in the SAIL databank.

Core documentation

User guide: Data First Introductory User Guide

Data catalogues:

- Family court
- Cafcass England

The data catalogue for Census 2021 will be available soon.

Other supporting resources

Webinar: This explainer webinar gives an overview of the Family Justice system and how the family court - Cafcass linked dataset can be used to address key evidence gaps.

Information about research priorities

Areas of Research Interest (ARI): The Ministry of Justice ARI 2020 sets out the department's critical evidence gaps.

Infographic: The Nuffield Family Justice Observatory children in the family justice system infographic summarises evidence and evidence gaps.

Webinar: Research evidence gaps on children and families in the family justice system.

Examples of research questions the data could be used to address

- What is the household composition of families in the family justice system and how does this change over time?
- How do socio-demographic characteristics; education and qualifications, or employment of family court users compare with the general population? Does this vary by case or role type?
- Who are the repeat users of family courts and which characteristics are more common among repeat users?
- Which proportion of private cases have had previous public law proceedings and vice versa?
- What variation (by local authority and/or court area) is seen in the way family courts are dealing with cases?

More information

More information about this linked dataset can be found on the ADR UK website.

> **Sir Andrew McFarlane, President of the Family Division of the judiciary in England and Wales:**
>
> *"The Data First project is likely to be the most significant step forward in the provision of hard and reliable information about the family justice system for a generation. It breaks entirely new ground and the fruits of the data that will be provided are likely to inform policy and practice and identify important trends for decades to come. Its advent could not be more welcome."*
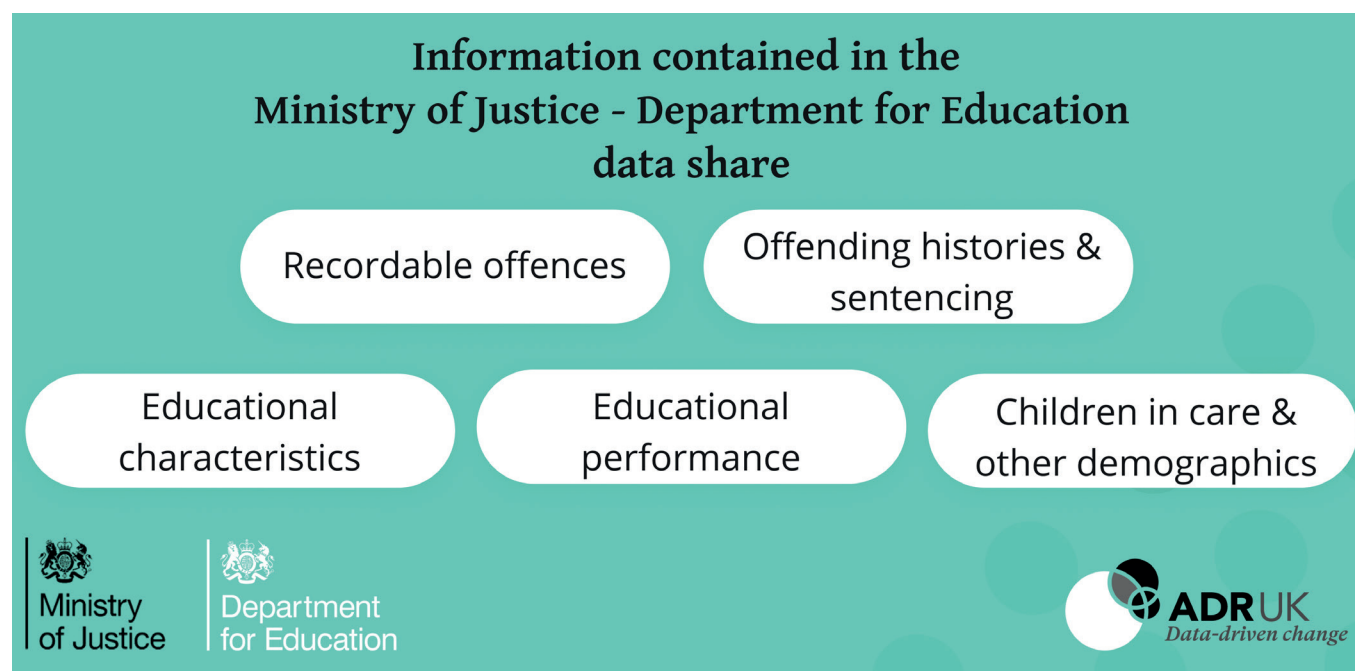
Ministry of Justice & Department for Education linked dataset - England

Summary

The MoJ-DfE linked dataset allows analysis to understand links between childhood characteristics, educational outcomes and (re)offending. The dataset includes data from Prison, Courts, Police National Computer, National Pupil Database, Looked-After-Children and Children in Need and covers variables such as demographics, offending data, school exclusions, and children in care. Analysis of this data will help develop the evidence base on understanding the relationships between educational and criminal justice outcomes and the drivers of offending. It will assist in identifying the population that requires support through early intervention and evaluating interventions to understand whether they are effective.

Overview of the component datasets and linkage

Information contained in the Ministry of Justice - Department for Education data share

- Recordable offences
- Offending histories & sentencing
- Educational characteristics
- Educational performance
- Children in care & other demographics

Timeframe:

Police National Computer extract (PNC): 2000 – 2021

National Pupil Database (NPD): Children born after 31 August 1985

Update frequency: Annually

Coverage: England

Size: PNC: 1.65 million records

NPD: 15 million records

Core documentation

User guide: Data First Introductory User Guide

Data catalogues: These can be requested by emailing datalinkingteam@justice. gov.uk and data.sharing@education.gov.uk.

Information about research priorities

Areas of Research Interest (ARI): The Ministry of Justice ARI 2020 sets out the department's critical evidence gaps.

Examples of research questions the data could be used to address

- Is there an association between particular interactions with the education system and offending, and if so, is one of these factors typically the driver?

- How are the relationships between educational and criminal justice outcomes impacted by demographic factors?

- Can education policy have unintended consequences for crime?

Examples of research done using the data

Feasibility study: This study explores the use of the MoJ-DfE linked dataset to evaluate early interventions to prevent violent crime.

Non-journal publications using the data:

- The Ministry of Justice published an examination of the education and social care background of children cautioned or sentenced for an offence.

- The Office for National Statistics published an analysis of the education and social care background of young people who interact with the criminal justice system.

- **ADR UK funded Research Fellows' outputs**: Five ADR UK Research Fellows are undertaking pathfinder research using this dataset. Their publications including blogs and Data Insights will be available via the ADR UK project page. Sign up to our newsletter for updates.

More information

More information about this linked dataset can be found on the ADR UK website and on the MoJ Data First website.

"

Dr Matteo Sandi, Research Economist, London School of Economics:

"I have done research on education policy and youth crime for a few years now using data from a number of countries, and this is probably the best data source that exists today in the world to study youth crime and links between education and crime."

"

# Accessing Data from the ADR UK Partnership's Trusted Research Environments

Much of the linked administrative data referred to in this brochure uses the Digital Economy Act 2017 (DEA) as the legal gateway for access. For secure data to be disclosed for public good research purposes under the DEA, a research project must be accredited to confirm that it meets the legal requirements for safe and secure data access. This is assessed by the UK Statistics Authority (UKSA) which accredits research projects for data accessed under the DEA.

The diagram below depicts the basic process for accessing data under the DEA. Where a different legal gateway is needed, the process for access also might be different. Please check with the relevant trusted research environment for more detailed information or see the ADR UK website.

Applicants need to become an accredited researcher in order to access the data. Safe researcher training is provided for researchers who need it, and this can be done through any of the ADR UK partnership's trusted research environments. Researcher accreditation is needed before the research proposal reaches the UKSA accreditation approval stage (see below), so a researcher could submit a research project application to UKSA (via the appropriate trusted research environment) to request access to the data before they become an accredited researcher but will only be granted access following accreditation.
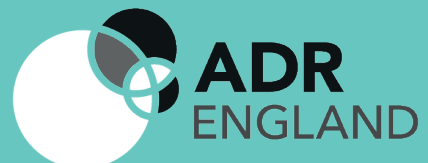
Most ADR England flagship datasets are available via the ONS Metadata Catalogue. Other ADR UK datasets are available via:

- NISRA statistics catalogue
- Research Data Scotland catalogue
- SAIL Databank catalogue
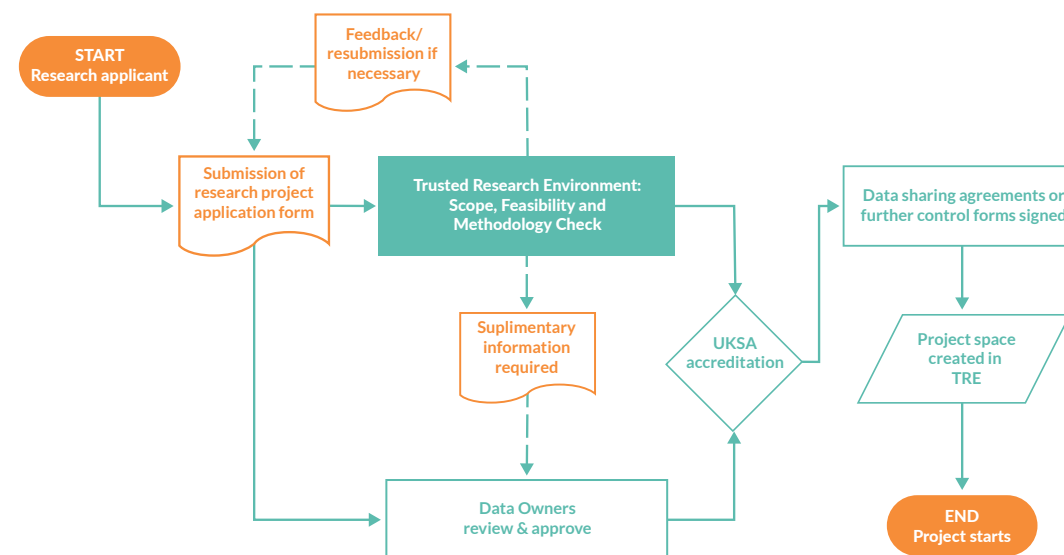
The ADR UK Partnership

ADR England is a portfolio of data linking and research projects, which enable policy-relevant insights across England and the UK. This work is delivered by a wide range of partners across academia and government, using data held by UK Government departments and public bodies. ADR England is managed by the ADR UK Strategic Hub, embedded within the Economic and Social Research Council.

To find out more, visit adruk.org

Get in touch with us via:

✉ hub@adruk.org
🐦 @adr_uk

ADR UK
Economic and Social Research Council
Polaris House
North Star Avenue
Swindon SN2 1UJ

9497