

Integrated Survey Data

Overview and conditions of access

Pierre Walthéry

UK Data Service

October2025

Plan of the presentation

1. Most common surveys with integrated data
2. Typical data integrated with surveys
3. Accessing secure integrated datasets

Integrated data

- When we add non survey data to survey data
 - Whether part of the original data collection or not
 - Whether primary or secondary
 - Whether same unit of analysis as the survey or not
 - Validation or enhancement (Benzeval et al 2020)
- Administrative, biometric, geographic, social media data
 - Accelerometer, genetic data, individual NHS/PAYE records
- This talk mostly deals with integrated data *available at the UK Data Service*

Part 1



**What are
the surveys
that include
integrated
data?**

Overview

- Depends on:
 - The topic covered by the data linked i.e. does it match common topics studied in surveys?
 - The survey itself (i.e. does it include the required linking information / user consent)
 - ... Scope of the surveys i.e. is linkage part of the original data collection, or is it a subsequent project?
- Major longitudinal studies:

Birth cohort studies

- Follow a sample of individuals over their whole life
- Born during a specific period of 1958(NCDS), 1970(BCS), 2000 (MCS), 2026 (?)
- Millenium Cohort Study (MCS)
 - ~ 19,000 children (born between June 2001 and Jan 2003)
 - 7 ‘sweeps’ 9 months then at 3, 5, 7, 11, 14, years old
 - Parent and child interviews
 - Focuses on education, skills and health, truancy, cognitive ability, biological measurements
 - ... In addition to traditional socio-economic and demographic data

Understanding Society (1)

- The largest longitudinal study representative of the UK population
- Initial sample size: 40K households, 100K individuals
- 14 waves so far: 2009-23; includes BHPS data 1991-2009
- Ethnic minority boost samples; Innovation Panel
- Very wide range of topics covered:
 - Employment, income, benefits, savings, debt, and assets
 - Health, well-being, and health behaviours
 - Housing, housing costs, and dwelling characteristics

Understanding Society (2)

- Further topics:
 - Family, partnerships, caring responsibilities,
 - Education, training
 - Expenditure, consumption, deprivation
 - Social attitudes, values, political opinions
 - Transport, mobility, and commuting patterns
 - Environmental behaviours, and related attitudes

Other studies

- Next Steps (formerly Longitudinal Study of Young People in England - LSYPE):
 - 16,000 people in England born 1980-90, from secondary school age (i.e. 13-14) onwards;
 - Initially set up by DfE to examine determinants of school achievement
- ELSA (English Longitudinal Study of Ageing):
 - Follows a sample of 19,000 people aged over 50 to understand all aspects of ageing in England.
 - Started in 2002, biennial waves.
 - Data on physical and mental health (incl. well-being), financial circumstances, and attitudes about ageing.

Part 2

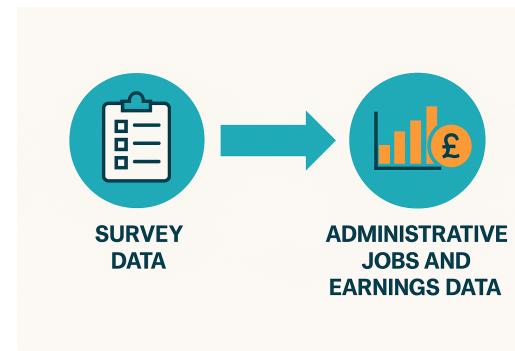


**What kind of
non-survey
data is
integrated
with UKDS
surveys?**

Overview

- Administrative records
 - ie *data collected by a public ie state controlled authority:* government department, the NHS
 - Health: NHS, SHS: medical records ie outpatient attendance, hospitalisation episodes, maternity
 - Education: National Pupil Database, school profile/teacher survey; student loan data, OFSTED data
 - Pollution; green space deciles; PAYE data
- Non survey measurement: energy consumption, health, behavioural
- Social media/digital trace

What is on offer: examples



1. Genetic risk data

- Polygenic scores (PGI) about health and social outcomes
- Probability of some outcomes given someone's genetic traits
 - A vector of probabilities attached to respondents' record according to their genetic information
 - 45 traits: ie health outcomes and behaviour; mental health and personality traits; Social outcomes
 - Available on the birth cohorts studies, ELSA & Next Steps
 - Subsamples limited to genetic 'Europeans'

More information: [Morris TT et al \(2025\)](#)

2. Hospital episodes data

- NHS data about all hospital admissions in England.
- Four datasets:
 - Episodes of using: Accident and Emergency ; Admitted Patient Care; Adult Critical Care; Outpatients
 - Mostly available for 2007/9-2023
- Data on diagnosis, maternity, mortality, mental health, treatment's length, deprivation etc.
- Available for the NCDS Birth Cohort

More information: [Kerry- Barnard et al \(2025\)](#)

3. School inspection data

- OFSTED ‘State of the nation’: anonymised data on latest schools inspections outcomes of 22,000 open schools
- Linked with the MCS, currently covers years 2005 to 2019
- Data on a wide range of topics i.e.:
 - Quality of teaching, learning and assessment
 - Effectiveness of leadership and management
 - Pupils’ achievement (aggregated) (2005-2015)
 - Behaviour and safety of pupils (2005-2015)

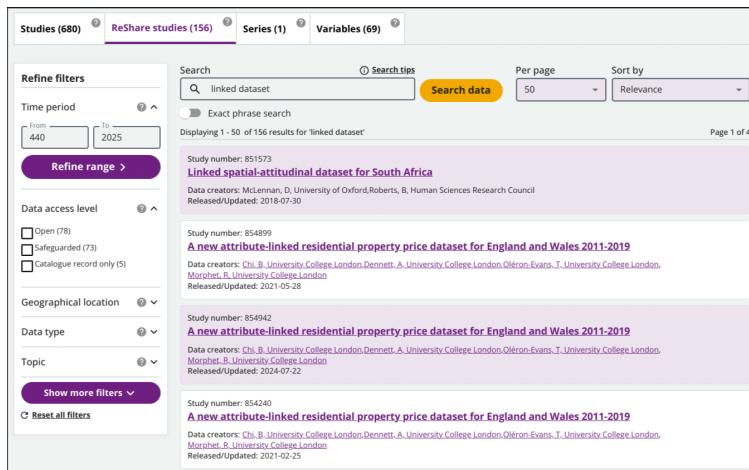
More information [Peters A et al \(2025\)](#)

4. NEST pension data

- (National Employment Saving Trust)
- Covers 1,000,000 employers, 11 millions employees
- Linked to consenting Understanding Wave 11 respondents (about 12,000)
- Data about:
 - Employer and employee characteristics
 - Current pension status
 - Pension contributions characteristics

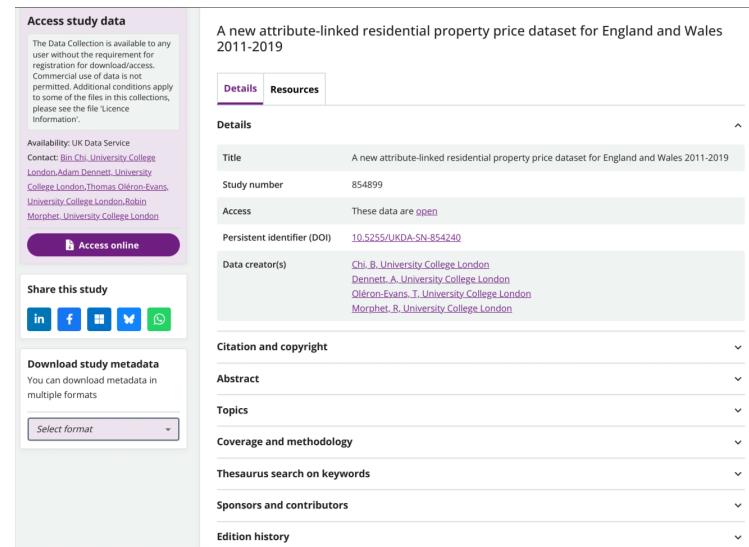
More information: [ISER \(2023\)](#)

5. Studies deposited on ReShare



The screenshot shows the ReShare studies search interface. At the top, there are four tabs: 'Studies (680)', 'ReShare studies (156)' (which is selected), 'Series (1)', and 'Variables (69)'. Below the tabs are several filter options: 'Refine filters' (Time period from 1940 to 2025, Data access level Open (78), Geographical location, Data type, Topic), 'Search' (Search tips, search bar with placeholder 'linked dataset', 'Search data' button, 'Per page' dropdown set to 50, 'Sort by' dropdown set to 'Relevance'), and a note 'Displaying 1 - 50 of 156 results for "linked dataset"'. The main content area shows three study entries:

- Study number: 851573**
Linked spatial-attitudinal dataset for South Africa
Data creators: McLennan, D., University of Oxford, Roberts, B., Human Sciences Research Council
Released/Updated: 2018-07-30
- Study number: 854899**
A new attribute-linked residential property price dataset for England and Wales 2011-2019
Data creators: Chi, B., University College London, Dennett, A., University College London, Oléron-Evans, T., University College London, Morphet, R., University College London
Released/Updated: 2021-05-28
- Study number: 854942**
A new attribute-linked residential property price dataset for England and Wales 2011-2019
Data creators: Chi, B., University College London, Dennett, A., University College London, Oléron-Evans, T., University College London, Morphet, R., University College London
Released/Updated: 2024-07-22



The screenshot shows the detailed view of a study record. The title is 'A new attribute-linked residential property price dataset for England and Wales 2011-2019'. The 'Details' tab is selected. Key details include:

- Availability:** UK Data Service
- Contact:** Bin Chi, University College London, Adam Dennett, University College London, Thomas Oléron-Evans, University College London, Robin Morphet, University College London
- Title:** A new attribute-linked residential property price dataset for England and Wales 2011-2019
- Study number:** 854899
- Access:** These data are [open](#)
- Persistent identifier (DOI):** [10.5255/UKDA-SN-854240](https://doi.org/10.5255/UKDA-SN-854240)
- Data creator(s):** Chi, B., University College London, Dennett, A., University College London, Oléron-Evans, T., University College London, Morphet, R., University College London
- Citation and copyright:** (details collapsed)
- Abstract:** (details collapsed)
- Topics:** (details collapsed)
- Coverage and methodology:** (details collapsed)
- Thesaurus search on keywords:** (details collapsed)
- Sponsors and contributors:** (details collapsed)
- Edition history:** (details collapsed)

Other tabs available are 'Resources'.

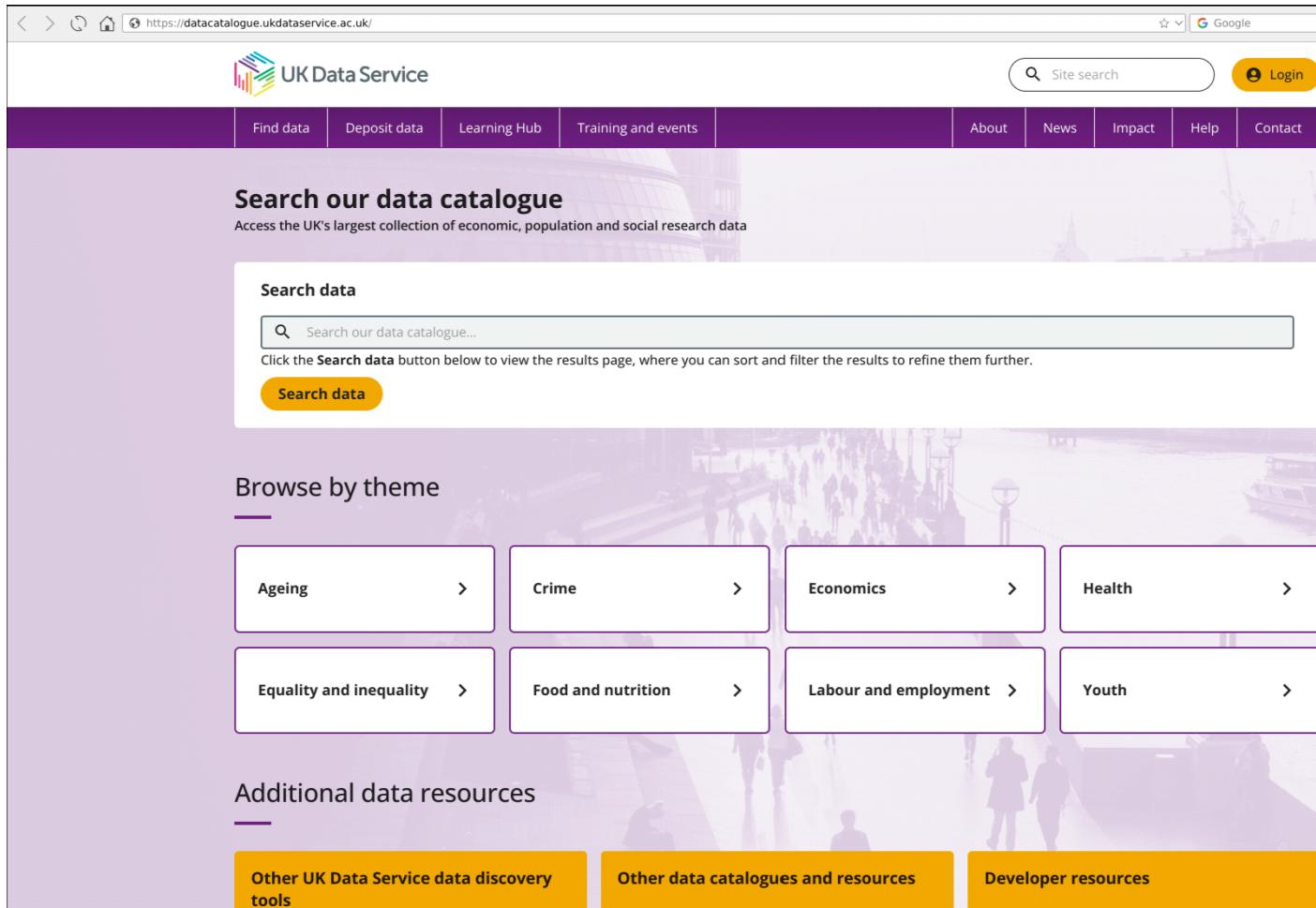
- Example: data about property characteristics linked to transaction ie price data
- 5,732,838 transactions in England and Wales, 79% of the total between 2011 and 2019,

Part 3



**Accessing
secure
integrated
datasets**

Searching for data

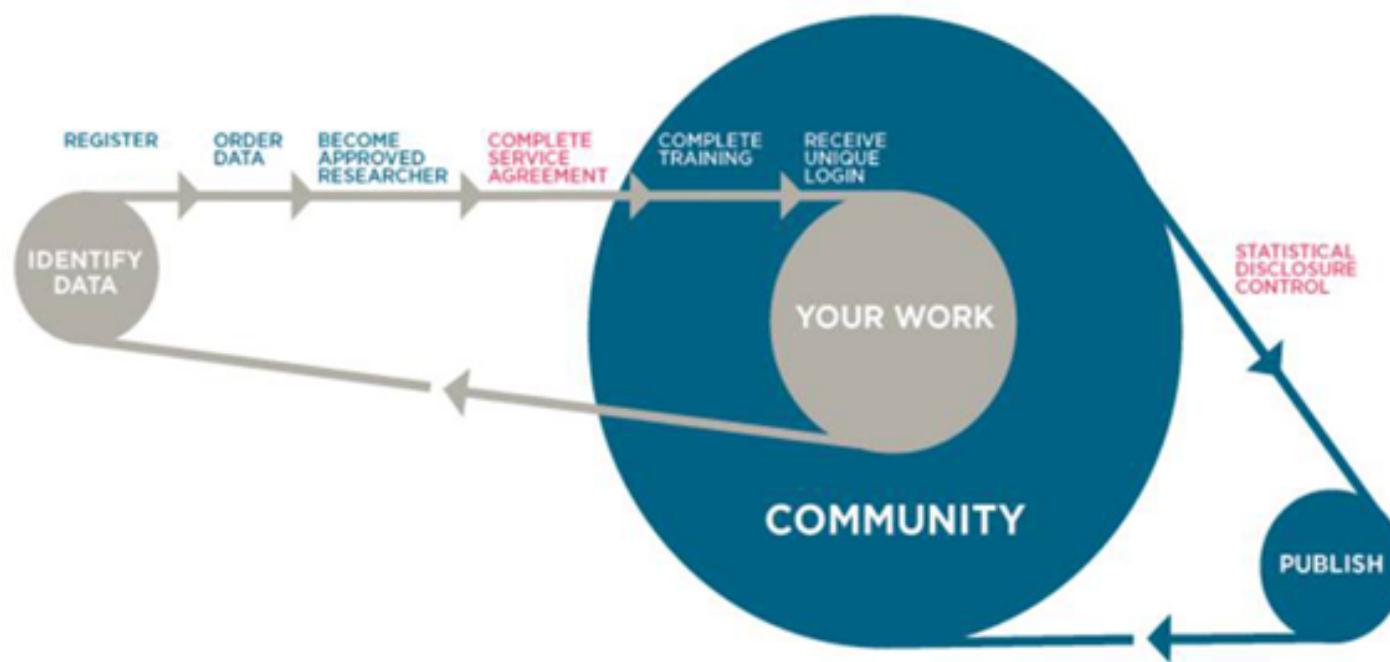


The screenshot shows the UK Data Service homepage. At the top, there is a navigation bar with links for 'Find data', 'Deposit data', 'Learning Hub', 'Training and events', 'About', 'News', 'Impact', 'Help', and 'Contact'. A 'Site search' input field and a 'Login' button are also in the top right. Below the navigation, a banner reads 'Search our data catalogue' with the subtext 'Access the UK's largest collection of economic, population and social research data'. A large search input field is centered below the banner. Below the search field, a button says 'Search data'. A descriptive text below the button reads: 'Click the **Search data** button below to view the results page, where you can sort and filter the results to refine them further.' Below this, a section titled 'Browse by theme' contains eight categories: 'Ageing', 'Crime', 'Economics', 'Health', 'Equality and inequality', 'Food and nutrition', 'Labour and employment', and 'Youth'. At the bottom, a section titled 'Additional data resources' includes links for 'Other UK Data Service data discovery tools', 'Other data catalogues and resources', and 'Developer resources'.

Secure datasets at the UK Data Service

- UKDS Secure Lab
- Access via encrypted web-based interface (<Citrix VPN technology)
- No data download, access only from organisational desktop, in the UKDA Safe Room, one of the SafePods across the UK, and abroad via partner organisations
- Outputs subject to Statistical Disclosure Control (SDC)

The application process



Application components

- Project application
- Accredited researcher
- Safe researcher training
- Secure access user agreement
- Secure Lab account setup

Safe researcher training

The course we recognise as valid for SecureLab access is the *Safe Researcher training (SRT) course* which covers:

- *Data security and personal responsibility* including legal background, security model, breaches and penalties.
- *Statistical Disclosure Control* – how to make statistical outputs safe and which principles are used.
- *Using the SecureLab* – how to use the interface and how to prepare and request data imports and suitable statistical outputs

Secure access user agreement

- Legally binding contract between the user, their organisation and the University of Essex
- It is a per person, per organisation agreement.
- Outlines the T&Cs of use of Secure Lab and includes:
 - The user has completed the mandatory SRT training.
 - Information about the user's security responsibilities.
 - Information about penalties and breaches
 - Our outputs release policy.
 - Our citation and copyright requirements.

Resources and links

- Polygenic scores:
 - [ELSA](#); [Next Steps](#); [Millenium Cohort Study](#)
- Hospital episodes:
 - [MCS](#); [BCS](#);
 - [NCDS](#)
- [School inspections data in the MCS](#)
- [NEST pension data in Understanding Society](#)
- [Linked housing data on ReShare](#)

Additional resources

- [UK Longitudinal Linkage Collaboration](#)
- [Administrative Data Research UK](#)
- [NHS Hospital Episode Statistics](#)
- [OFSTED](#)
- [SAIL: Secure Anonymised Information Linkage \(Wales\)](#)
- [Cohort surveys data linkage programme page](#)
- [Integrated Data Service](#)

Thank you for your attention

Any questions: help@ukdataservice.ac.uk