

Datasets

Access, uses and management

Introduction

Social science research findings are important because they affect everyone. They are an invaluable resource for policymakers in every area of government, from childcare and education to employment, health, welfare and pensions.

Social scientists rely on high quality collections of data as a major tool to analyse what is happening in our lives today. Data collections are usually described as datasets, in which digital information collected from people or organisations is stored in tabular form and where each column represents a different attribute or variable. An example of a variable would be the age of a person or what kind of work they do. These datasets are used to track how our behaviour is changing within the family, in the workplace, the street and in our leisure time. They also provide invaluable feedback about how people respond to changes in policy.



Questions that might interest social scientists include:

- What prompted the street unrest of Summer 2011?
- Why do so many people say they don't trust politicians?
- How have social media changed the relationship between young people?

How are datasets used?

One of the major collections of data in the UK is the Economic and Social Data Services (ESDS) Catalogue www.esds.ac.uk/findingData/aboutCat.asp which contains over 5,000 datasets, including 35 tailor-made social survey datasets for teaching purposes.

The ESDS studies range from broad overviews of our changing society to more specific studies of particular groups, such as children or residents of a particular region. One of the broadest and biggest studies is Understanding Society, formerly known as the United Kingdom Household Longitudinal Study, which follows the socio-economic circumstances and attitudes of 100,000 individuals in 40,000 British households.

This major dataset, which is funded by ESRC, is led by the Institute for Social and Economic Research (ISER), at the University of Essex.

Understanding Society is a form of longitudinal study, which means that interviewers return at regular intervals to the same people and collect detailed information about their lives and how they are changing. As well as providing information about the composition of households, Understanding Society data can be used to measure phenomena such as patterns of poverty, unemployment, and marriage or cohabitation, and analyse the factors that affect them.

A snapshot view of each household

The first large data hoard from the survey is now available to researchers. It gives a detailed (anonymous) snapshot of all the individuals in a household including their gender, date of birth and employment status. Questions about housing, mortgage or rent payments, material deprivation, and consumer durables and cars are also included.

Datasets: Access, uses and management

Close up view of individuals

The survey also conducts individual interviews with every person in the household aged 16 or over, which provides further details such as:

- family background, ethnicity and language;
- migration, partnership and fertility histories;
- health, disability and caring;
- current employment and earnings;
- employment status, parenting and childcare arrangements;
- family networks; benefit payments;
- political party identification;
- household finances;
- environmental behaviours;

The results from the survey provide an invaluable source of information for academic research and policymakers and is often linked to the results of other data collections held by organisations such as the Medical Research Council (MRC) and the Office for National Statistics (ONS).



International datasets

One of the most significant research trends in recent years has been the growth of comparative and international studies, often in collaboration with universities and research groups across the world. The datasets generated by these studies allow researchers to compare the attitudes and behaviour of people in different parts of the world. There has been a parallel growth of interest in international datasets, which enable data from international sources to be shared and used.

One example is the ESRC-funded European Social Survey, (the ESS) www.europeansocialsurvey.org which is run by a team headed by the Centre for Comparative Social Surveys, City University, London working with institutions in Norway, Germany, the Netherlands, Spain, Belgium and Slovenia. The central team is funded by the European Commission, with part funding from the ESRC (who also fund the UK's participation in the project).

The ESS is a social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. The survey, which began in 2002, covers more than 30 nations (UK sample size: 2400) and a core set of topics, which include: political engagement;

trust in institutions; moral and social values; social capital; social exclusion; national, ethnic and religious identity; well-being, health and security.

In addition to the core module, each round includes specific topics in particular detail. For example, the 2010 round focuses on the areas of:

- Work, family and well-being including: the impact of the recession on households and work; job security; housework; wellbeing; unemployment; work-life balance.
- Trust in criminal justice including: confidence in the police and the courts; cooperation with the police and the courts; contact with the police; attitudes towards punishment.

Why should we care what people in other countries think?

Comparative surveys help us put our own experience in a wider context. For instance, if we know how voter turnout in the UK compares with that across Europe, we can explore why Britain is in line with - or different from - other countries. We can also look at how attitudes to education, employment and health vary across different welfare regimes. This helps our understanding of why Britain is how it is.

Datasets

Using ESDS data for teaching

Introduction

The use of data in teaching is an invaluable way for learners to engage with real-life research. By investigating the collections held at ESDS, learners can find out how real data is originally conceived, collected, analysed and managed. Data based on well-known studies or surveys can bring both substantive and methodological topics alive.

The ESDS produces 35 tailor-made survey datasets, which are available for teaching purposes, which are easier for students to handle. Both the study methods and data can be downloaded, free of charge at www.esds.ac.uk/findingData/subjectResults.asp?subcat=IV\B The subjects areas covered include crime, health, general household surveys and young people.

There are also over 15 teaching datasets in the ESDS Nesstar Catalogue where a selected range of social and economic data can be browsed and visualised online without the need for special software. These include material from the UK longitudinal studies <http://nesstar.esds.ac.uk/webview/index.jsp>

Online documentation in the ESDS Catalogue, such as the actual survey questionnaires and interviewer instructions, can be accessed without registration, but in order to download data as part of a taught course teachers need to register an account. Details can be found at www.esds.ac.uk/support/newuser.asp or www.esds.ac.uk/resources/teaching.asp Data used for teaching should be kept in a secure place and access restricted to students on the course.

The ESDS offers the following additional resources for classroom use:

- methods and data analysis guides providing introductions to re-using data using particular methods or software; www.esds.ac.uk/support/statguides.asp
- case studies showing how ESDS data have been used for research and teaching; www.esds.ac.uk/resources/datainuse
- useful guidance and training materials on data sharing and data management www.data-archive.ac.uk/create-manage including confidentiality, consent and ethics; www.data-archive.ac.uk/create-manage/consent-ethics
- thematic pages which provide an introduction to available material on themes like housing, health or crime; www.esds.ac.uk/resources/themes.asp
- guides to popular datasets and data series www.esds.ac.uk/support/datasetguides

Other dedicated ESDS resources for teachers and students include:

- Teaching material on interview methods and non-interview methods such as focus groups, diaries and visual methods;
- An online resource for teachers who want to explore data on measuring psychological distress to demonstrate research principles and methods;
- Resources supplying key steps in analysing attitudes to same sex relations, fear of crime and health differences;
- Teaching materials demonstrating how qualitative data can be re-used with activities for use in the classroom or as self-paced learning activities

Examples of case studies

Summaries of 80 social science research projects may be accessed at www.esds.ac.uk/resources/datainuse/casestudies.asp. The following may be of special interest to young people:

Do comprehensive schools reduce social mobility?

This study by Vikki Boliver, University of Durham, in collaboration with Adam Swift at the Politics Department and Balliol College, University of Oxford, used data from the National Child Development Study.

When comparing young people with the same level of ability, but from different types of school, the study found that the selective system did not lead to increased mobility of children from any particular background.

Do smarter children avoid drug use later in life?

Using data from the 1970 British Cohort Study, James White and others at University of Cardiff and University College, London found that children with higher IQs measured at five or ten years of age were more likely to use illegal drugs such as cannabis, cocaine and ecstasy later in life. There was no link with social class and education. They also found a stronger association amongst women than men.

Electoral participation of young people in Europe

In elections between 1999 and 2002 the overall turnout rate for 22 European countries was 70 per cent, compared to 51 per cent for electors aged less than 25. Using data from the European Social Survey, Edward Fieldhouse, Mark Tranmer and Andrew Russell, University of Manchester, looked at non-voting among the young.

They used multilevel logistic regression, a statistical model used to analyse clustered data in medical, public health, epidemiological, and educational research, to test the extent to which variations in turnout were - on the one hand - attributable to the characteristics of young people and - on the other hand - attributable to the electoral context in each country. They found that variations in turnout among young people have both individual-level and contextual-level explanations.

SOCIAL SCIENCE FOR SCHOOLS

This resource is one part of a series of resources produced for Social Science for Schools. The Social Science for Schools website, www.socialscienceforschools.org.uk, is home to a number of teaching resources created to bring current and engaging social science research to your classroom.

Feedback and enquiries can be sent to schools@esrc.ac.uk

Datasets

Key points to remember when creating and managing data

Collecting data through surveys and questionnaires is the only the first step in a complex research process. Analysis would be meaningless without strict practices for creating, preparing, storing and sharing data. For this purpose international standards have been set to ensure that comparisons between various data sources are valid. The UK Data Archive provides guidance to researchers on how to collect and look after data well.

Sharing data

Much research data about people—even sensitive data—can be shared ethically and legally if researchers employ strategies of informed consent, anonymisation and controlling access to data. Ethical guidance is provided by professional bodies, host institutions and funding organisations

Describing data

Raw data cannot be used without some accompanying context. For example, columns in spreadsheet should be meaningfully labelled, and details about how a survey sample was chosen and how data were gathered should be accurately described.

Data security

Making back-ups of files is an essential element of data management. Regular back-ups protect against accidental or malicious data loss and can be used to restore originals if loss of data does occur.

Physical security, network security and security of computer systems and files all need to be considered to ensure security of data and prevent unauthorised access, changes to data, disclosure or destruction of data.

Transferring data

Transmitting data between locations or within research teams can be challenging for data management infrastructure. Data encryption will maintain data security during transmission and should be used when transmitting sensitive or personal data. Collaborative research can pose challenges for facilitating data sharing, transfer and storage, and providing access to data across partners or institutions. Virtual research environments may offer solutions.

Disposing of data

Having a strategy for reliably erasing data files is a critical component of managing data securely and is relevant at various stages in the data cycle.

Deleting files and reformatting a hard drive will not prevent the possible recovery of data that have previously been on that hard drive. At the conclusion of research, data files which are not to be preserved need to be disposed of securely.