

ESDS Government Newsletter

The UK Statistics Authority (UKSA)



The UK Statistics Authority (UKSA) was established on 1 April 2008 by the Statistics and Registration Service Act 2007. It is an independent body operating as a non-ministerial department, directly accountable to Parliament. The UKSA's overall objective is to promote and safeguard the quality of official statistics that serve the public good. More information can be found at www.statisticsauthority.gov.uk.

The Office for National Statistics (ONS) is the executive office of the UKSA. Its web site www.ons.gov.uk/about contains details of its role at the centre of the UK's system of official statistics and its organizational structure.

What does this change mean for access to government surveys?

The Statistics and Registration Services Act 2007 resulted in changes in ONS policy that applies to all ONS Social Surveys lodged at the UK Data Archive.

For ESDS users, a key distinction is between data that are defined as 'personal' under the Act and data that are not so defined. Data are defined as 'personal' if an individual can be identified either directly from the data or by the auxiliary use of published information. Surveys which do not contain 'personal information' can be accessed under the UKDA's standard End User Licence (EUL). Surveys containing personal information may be downloaded from the UKDA as Special Licence datasets, which have had some degree of protection applied and which are made available to Approved Researchers. More detailed data may be accessed via the ONS Virtual Microdata Laboratory (VML).

ESDS users should expect little if any reduction in detail on government surveys currently supplied via the EUL. Crucially, single year of age on the QLFS, replaced with grouped-age in recent years, will be restored.

For most EUL microdata, the lowest level of geography will be the Government Office Region (GOR) and detailed versions of SOC and country of birth will be available. There will, however, be some top coding for very high income and some rounding for amounts won in lotteries.

Special Licence datasets, available to Approved Researchers, will give researchers access to more detailed data, such as with local authority level geography. They will also contain unprotected versions of some variables that may have been top-coded, for example, income. For the Annual Population Survey, which because of its large sample size has age banded in the EUL version, the SL version will include single year of age.

Applications for AR status are made through the UKDA. To apply for accreditation a user must provide evidence that that he/she is a 'fit and proper person' and also provide details about the purpose of the research. More information is available from www.esds.ac.uk/orderingData/specialLicence.asp

Users who obtain access to SL data are required to follow the '*ESDS Guide to good practice: Micro data handling and security*' that includes guidance on how to avoid inadvertent statistical disclosure and how to ensure that data is stored and accessed securely.

ESDS Government resources

ESDS variable search tool

A common problem when searching repositories for secondary microdata is finding useful data to meet specific requirements. This search tool allows you to search for information on variables contained in over 550 ESDS Government datasets. The search engine uses the text in question wording, variable and value labels. This tool can be used to identify whether a specific topic is covered, in which survey a particular question is asked, or if a given question is asked across a number of years of the same survey. It is available at www.ccsr.ac.uk/esds/variables/

Updated guides

The Guide to data sources for Scotland and the Employment and the labour market introductory user guide have been updated. Find them at www.esds.ac.uk/government/resources/themeguides.asp

Contents

New ESDS Government Guide: 'Using Hierarchical Data'.....	2
Focus on Eurostat resources: The EU-Labour Force Survey.....	4
Forthcoming events.....	4
Events held in the past six months.....	4

New ESDS Government Guide: 'Using Hierarchical Data'

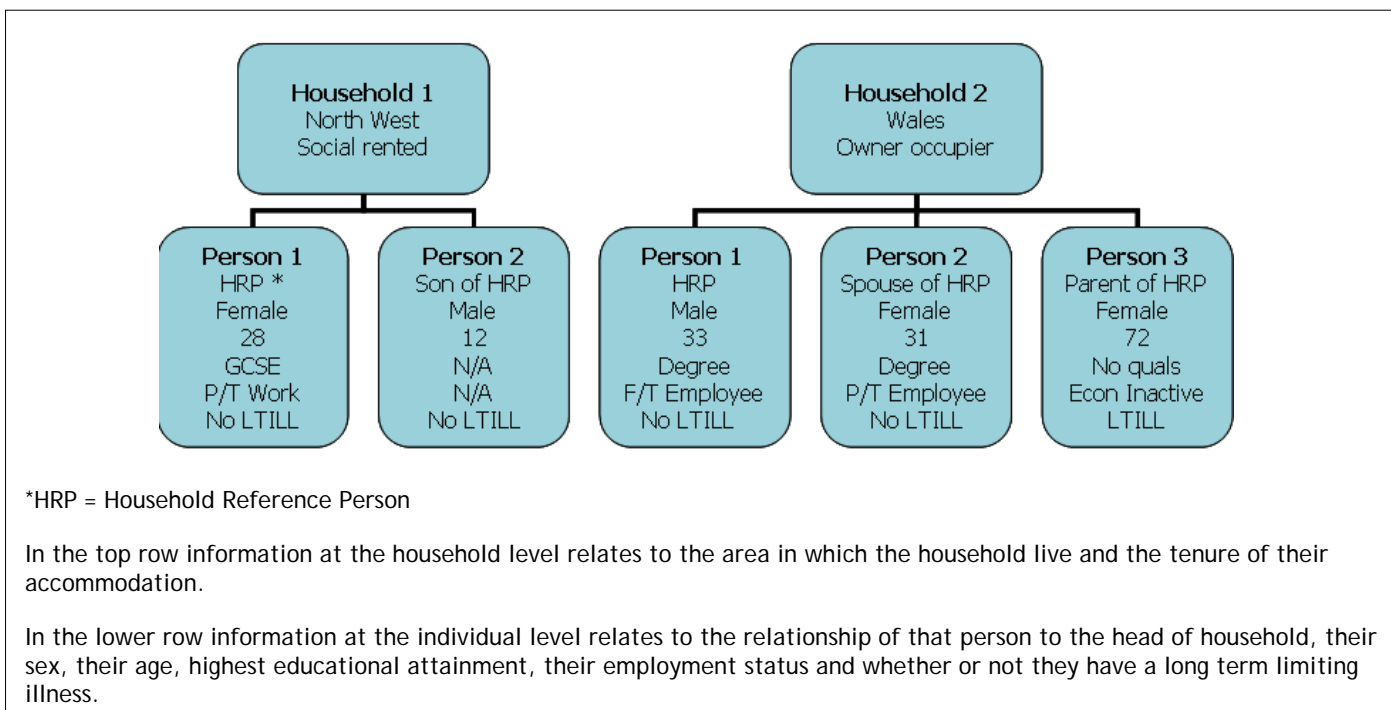
ESDS Government has produced a new introductory web guide, 'Working with survey files: Using hierarchical data, matching files and pooling data'. This guide offers an introduction to hierarchical data and units of analysis, demonstrating some useful data management procedures in SPSS and Stata.

Hierarchical data files are where data at a lower level unit are nested within a higher level so that smaller units within a larger unit can be identified. Examples of hierarchical data include:

- Individuals, households, or individuals within households (e.g. using the General Household Survey)
- Employees, businesses, or employees within business (e.g. using the Work Employee Relations Survey)

Hierarchical data has a variety of uses: It can be used to analyse the smaller unit; to analyse the larger unit; or information from one level can inform the analysis of another. Figure 1 shows a file containing information about both households and individuals. The figure shows data for two households - household 1 and household 2. Each household has some information relevant to the whole household (its ID number, region and tenancy). Information is also held on each individual within the household. It is possible to link each person to their household, and so link members of the same household to each other. You can therefore use the information about both individuals in Household 1 to discover that Person 1 in Household 1 has a 12 year old son living with her.

Fig 1. An example simple hierarchical file



Hierarchical data can be stored in data files in a number of different ways. Figure 2 shows hierarchical data in a 'rectangular' format, where data from both levels of the hierarchy are combined at one level. This means that the higher level information from the household level (such as the household level variable tenure) has been distributed (repeated) across all individuals at the lower level. The variable HHID in this example uniquely identifies households.

Fig 2. Rectangular hierarchical data

HHID	Region	Tenure	Person	Relation to HRP	Sex	Age	Education	Economic Activity	Long term limiting illness
1	NW	SocRent	1	HRP	F	28	Gcse	p/t	No
1	NW	SocRent	2	Son	M	12	n/a	n/a	No
2	Wales	OwnOcc	1	HRP	M	33	Degree	f/t	No
2	Wales	OwnOcc	2	Spouse	F	31	Degree	p/t	No
2	Wales	OwnOcc	3	Parent	F	72	No quals	inactive	yes

Alternatively, data from different levels of hierarchy can be stored in different files. If the example data in Figure 1 were to be held in separate files, it could look like Figure 3. Here the household and individual level data are held across a 'database' or set of 'data tables'. This format is slightly more efficient in terms of storage as there is less replication of data than in Figure 2. However, to use information about both levels, you must first link together the two tables using a key ID variable. In the current example, to link the household level information to the individual level, you would use the household ID variable HHID. Figure 4 lists a number of popular ESDS Government surveys and other data sources, showing main levels of hierarchy and storage format. The General Household Survey (GHS) gives an example of a survey where hierarchy is stored in an individual file, whereas the Family Resources Survey is a good example of a dataset where hierarchy is stored over multiple files.

Fig 3. Hierarchical data in separate files

Household level table		
HHID	Region	Tenure
1	N W	SocRent
2	Wales	OwnOcc

Person level table							
HHID	Person	Relation to HRP	Sex	Age	Education	Economic activity	Long term limited illness
1	1	HRP	F	28	Gcse	p/t	No
1	2	Son	M	12	n/a	n/a	No
2	1	HRP	M	33	Degree	f/t	No
2	2	Spouse	F	31	Degree	p/t	No
2	3	Parent	F	72	No quals	inactive	yes

Fig 4. Household and other hierarchy in popular datasets

Survey Hhd hierarchy?	Main Levels	Hierarchy Store as:
GHS ✓	Household, Family, Individual, Sub Individual	Single file
LFS ✓	Household, Family, Individual	Single file (QLFS/Hhd data)
FES ✓	Multiple, inc. household, person, family unit, benefit unit	Multiple files
FRS ✓	Household, Benefit Unit, Individual etc.	Multiple files
HSE ✓	Household, Individual (watch out for variable samples)	Single file (1 all inds, 1 all resps)
BSAS	Individual	Single file
BCS	Individual, Incident (Hhd context only)	Multiple files
BHPS ✓	Household, Individual (& below)	Multiple files
Household SARs ✓	Household, Family, Individual	Single file

The level of hierarchy most suitable to your analysis often depends on the question(s) you are trying to answer. Often the research question will indicate which unit to use. If you were interested in whether or not the introduction of Appropriate Pension Plans (APPs) in 1998 differentially affected men and women's uptake of pensions, this would suggest an individual level of analysis. If you wanted to look at whether the proportion of households where no one is in paid work has been changing, the household level of analysis would be more appropriate. It is also often possible to examine similar topics at different units of analysis. For example, you could use the individual level to consider the number of people who live in poverty, or the household level to consider how many households live in poverty.

Identifying the unit that you wish to use will help you ensure that you select an appropriate dataset and that you are using the dataset correctly. Results that have differing interpretations can be obtained on the same topics depending on which unit of analysis you use. Consider the example given in Figure 5. The percentage of households with one resident is more than twice the percentage of individuals who live alone. Similarly, individuals on average live in households of size 3, however because nearly a third of households are single person households the mean size of households is just over 2.

Fig 5. Individual and Household level questions and their answers

●	HOUSEHOLD LEVEL: 1 observation per household
—	What proportion of households contain only 1 person? 29.2%
—	What is the mean household size? 2.34
●	INDIVIDUAL LEVEL: 1 observation per person
—	What proportion of individuals live alone? 12.5%
—	What is the average household size for individuals in the sample? 3.05
<i>Source: QLFS March - May 2005 (authors analysis, unweighted)</i>	

Working with Survey Files will help you decide on the unit of analysis most useful for your research. It also illustrates procedures in SPSS and Stata for working with datasets with different levels of hierarchy within single files, as well as techniques for combining files where levels of hierarchy are stored across multiple tables. Exercises are given using the General Household Survey, Family Resources Survey, and the UK Labour Force Survey. Although the guide is designed for users of ESDS Government surveys, it is also relevant to new users of other data.

Working with Survey Files is available online at www.esds.ac.uk/government/resources/analysis/

Footnote 1: This guide is in part based on the contents of a one-day workshop on data linking held at CCSR, University of Manchester in October 2007. See www.ccsr.ac.uk/esds/events/past.shtml#datamanipulation for the slides and workbooks.



Issues of cross-national comparability represent a common challenge to comparative research, particularly where research involves secondary data sources collected in different countries. Often, such problems arise because of methodological differences between countries. In the case of survey data, these can include differences in survey design, survey questions, variable definitions, or in how variables are derived. Harmonisation provides a potential remedy to such problems. This can involve standardisation, where a single protocol for survey design and questions are agreed prior to fieldwork. Data can also be harmonised once collected, where possible, such as through recoding variables that vary between countries so that their categories match, or by using information from individual countries to derive internationally agreed classifications.

At this year's Labour Force Survey User Meeting in January, Frank Espelage (Eurostat) gave a presentation on the EU Labour Force Survey (EU-LFS). Derived from national Labour Force Surveys, the EU-LFS offers harmonised data across the 27 EU Member States, and for Norway and Iceland.

The EU-LFS provides a primary source of labour market information across the European Union, offering a high degree of cross-national comparability. Detailed regulations are implemented by Eurostat to ensure harmonisation on the recording of labour market characteristics across countries. Harmonisation is achieved through the use of identical variable definitions for all countries, such as the ILO definition of unemployment, and through the use of common classifications such as NACE, ISCO, ISCED, and NUTS.

The EU-LFS data were first released in 2003. Data for years spanning back to 1983 are available with varying levels of coverage across countries, although the most complete data is from 2002 onwards. The current available release (2007) covers core survey data up to 2006, and additional ad hoc module data from 2002 to 2005. For Germany, due to legal regulations, anonymised data is only produced for 2002 onwards. Data for Malta is not present in anonymised datasets.

In addition to the EU-LFS, the following survey datasets are available from Eurostat:

- European Community Household Panel;
- Community Innovation Survey;
- Continuing Vocational Training Survey;
- Structure of Earnings Survey;
- European Union Statistics on Income and Living Conditions;
- Adult Education Survey.

For a fee, data is available to a variety of types of research organisations. These include researchers of universities and other higher education organisations; other organisations or institutions for scientific research; and the national statistical institutes of the Member States. Access to Eurostat microdata is granted through anonymised versions distributed by CD, or through an in-house secure 'safe centre' at Eurostat (Luxembourg). The type of access available depends on the survey. In the case of the EU-LFS, anonymised CD versions are obtainable. Applications for data are made at the level of host research institute/organisation, for specific research projects, rather than at the level of individual researchers. For further details on the Eurostat microdata and information on access see <http://epp.eurostat.ec.europa.eu/>.

Forthcoming events from ESDS Government

www.ccsr.ac.uk/esds/events/

Introduction to the Expenditure and Food Survey, 19 June 2008, Manchester

Health surveys user meeting, 8 July 2008, London

Other dates for your diary

Census Microdata: findings and futures, 1-3 September 2008, University of Manchester,

www.ccsr.ac.uk/sars/conference

Events held in the past six months

Family Resources Survey user meeting, Friday 25 April 2008, London

General Household Survey user meeting, Thursday 6 March 2008, London

ESDS Government meeting on changes in access to the government surveys and Labour Force Survey/Annual Population Survey user meeting, Thursday 17 January 2008, London

British Crime Survey User Meeting, Tuesday 11 December 2007, London

Slides from these events are available at www.ccsr.ac.uk/esds/events/past.shtml.

ESDS Government

Email:
govsurveys@esds.ac.uk

Telephone:
+44 (0)161 275 1980

Fax:
+44 (0)161 275 4722

ESDS Government
CCSR
University of Manchester
Humanities Bridgeford Street
Manchester
M13 9PL