# Managing, publishing and sharing research data

Veerle Van den Eynden
UK Data Service
University of Essex

Open Access week
Bangor, 23 October 2015

UK Data Service

# Overview

- Best practices for managing research data
- How to publish research data
- ReShare for ESRC data (and more)
- Discover and access data from UK Data Service

# What is the UK Data Service?

- a comprehensive resource funded by the ESRC
- a single point of access to a wide range of secondary social science data
- Supporting and training researchers and data creators of social science data and related disciplines

UK Data Service

The UK Data Service is a comprehensive resource funded by the ESRC. It is made up of the former services ESDS, Census.ac.uk, and the Secure Data Service and the survey question bank
We provide a single point of access to a wide range of social science data
As well as the data, we also provide support, training and guidance, webinars & a range of face to face events

# Who is it for?

- academic researchers and students
- government analysts
- charities and foundations
- business consultants
- independent research centres
- think tanks

ukdataservice.ac.uk

UK Data Service

Anyone & everyone can register

Most of our regular users are HE – staff, researchers, students

BUT contrary to popular beliefs, everyone can register & we have many users from other sectors such as local and national government departments, charities, etc.

23,000 users across the world – 60,000 data download ever year

# Research data management team

- Supporting researchers to make research data shareable
- UK Data Service helps materialise Data Policy for the Economic and Social Research Council (ESRC)
  - Data management planning advice & guidance
  - Data management guidance & training, esp. on confidentiality, security, ethics
  - Research data available for re-use to maximum extent possible, via:
    - ReShare repository
    - http://discover.dataservice.ac.uk

UK Data Service

# Why manage research data ?

- Data creation in research is often expensive
- Data = foundation of research
- Data underpin published findings
- Good quality data = good quality research
- Protect data from loss, destruction,…
- Compliance with ethical codes, data protection laws, journal requirements, funder policies

UK Data Service

# Why share research data

- Data can be re-used
  - new research
  - collaborative research
  - new forms of research (e.g. data mining)
  - learning & teaching, e.g. methods
- Evidence for published findings (verification, scrutiny, replication,…)
- Meta-analysis, augment sample sizes, develop time-series,….

- But: individual researchers reluctant to share their data (no time, no rewards, IP, fear competition,….)

UK Data Service

# The bigger picture: policies & drivers

- Many public research funders have data policies and expect:
  - data management planning
  - data management
  - data sharing

- Journal publishers increasingly expect accessible supplementary data as evidence and for verification

UK Data Service

# Driving data sharing: research funders

Research Councils UK Common Principles on Data Policy

- *Publicly funded research data are a public good, produced in the public interest, that should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.*
- in accordance with relevant standards and community best practice
- metadata to make research data discoverable
- legal, ethical, commercial constraints on release of research data
- recognition for collecting & analysing data; limited privileged use
- acknowledge sources of data, intellectual contributions, terms & conditions
- use public funds to support the management and sharing of publicly-funded research data

UK Data Service

Public funders are following the OECD principle that publicly funded research data should as far as possible be openly available to the scientific community.
RCUK adopted common broad principles.

# Driving data sharing: research funders

- Data sharing policy mandating or encouraging data sharing
- Data management / sharing planning required
- Grant holders responsible for managing & sharing data, except EPSRC – institutional responsibility
  - publish metadata online, with DOI (digital object identifier)
  - maintain data securely for 10 years after last access request
- Papers include statement on access to supporting data
- Fund data sharing support services and data infrastructure, e.g.
  - UK Data Service (ESRC)
  - NERC data centres (NERC)
  - Genbank (BBSRC, MRC)

UK Data Service

# Driving data sharing: research funders

European open access policies: Horizon 2020, European Research Council (ERC)

- communication & recommendation on access to / preservation of scientific information (July 2012) (publications & research data)
- pilot on open access to research data, primarily data underlying (open access) scientific publications for Horizon 2020
- data management guidelines for Horizon 2020 (~ policies)

generally based on *OECD Principles and Guidelines for Access to Research Data from Public Funding*

UK Data Service

And European Research Council

# Driving data sharing: journals & publishers

- data underpinning publication accessible
  - upon request from author
  - as supplement with publication
  - in public repository

Examples:

- BioMed Central open data statement
- Data journals: Scientific Data (Nature PG)
- PLOS ONE: "Publication is conditional upon the agreement of the authors to make freely available any materials and information described in their publication that may be reasonably requested by others." …/…. "will not consider a study if the conclusions depend solely on the analysis of proprietary data" … "the paper must include an analysis of public data that validates the conclusions so others can reproduce the analysis and build on the findings."

UK Data Service

Increasingly journals / publishers have data policies
Pre-publication vs post-publication monitoring of data access
Dryad = repository basic and applied biosciences
Treebase = repository phylogenetic data
JORD, Journal Research Data Policy Bank : JISC funded

# What can researchers do in practice ?

- Write a data management/sharing plan
- Make sure data are shareable and can be understood, e.g.
  - consent to share
  - disclose identities
  - open/standard formats
  - context & documentation
  - Well organised data

UK Data Service

# How to share data obtained from people in an ethical / legal way ?

- Obtain informed consent, also for data sharing and long-term preservation / curation
- Protect identities e.g. anonymisation, not collecting personal data
- Regulate access where needed (all or part of data) e.g. by group, use, time period
- Securely store personal or sensitive data (separately)

UK Data Service

# Consent needed across the data life cycle

- Engagement in the research process
  - decide who approves final versions of transcripts
- Dissemination in presentations, publications, the web
  - decide who approves research outputs
- Data sharing and archiving
  - consider future uses of data

Always dependent on the research context – special cases for covert research, verbal consent, etc.

UK Data Service

# Timing and form of consent

| | + | - |
|---|---|---|
| One-off | Simple<br>Least hassle to participants | Research outputs not known in advance<br>Participants will not know all info they will contribute |
| Process | Ensures 'active' consent | May not get all consent needed before losing contact<br>Repetitive, can annoy participants |

| | |
|---|---|
| Written | More solid legal ground, e.g. participant has agreed to disclose confidential info<br>Often required by Ethics Committees<br>Offers more protection for researcher<br>Not possible for some cases: infirm, illegal activities |
| Verbal | Can be difficult to make all issues clear verbally<br>Possibly greater risks for researcher<br>Best if recorded |

UK Data Service

# Aspects to consider

- Different forms of consent for different materials, e.g. audio recordings vs transcripts
- Right to withdraw – what to do with already collected data?
- Simple, without omitting key issues (like sharing!)
- Balancing of benefits and harms

UK Data Service

# In practice: wording in consent form / information sheet – interviews, photos

We expect to use your contributed information in various outputs, including a report and content for a website. Extracts of interviews and some photographs may both be used. We will get your permission before using a quote from you or a photograph of you.

After the project has ended, we intend to archive the interviews at …. Then the interview data can be disseminated for reuse by other researchers, for research and learning purposes.

The interviews will be archived at ……. and disseminated so other researchers can reuse this information for research and learning purposes:

- ❑ I agree for the audio recording of my interview to be archived and disseminated for reuse
- ❑ I agree for the transcript of my interview to be archived and disseminated for reuse
- ❑ I agree for any photographs of me taken during interview to be archived and disseminated for reuse

UK Data Service

# In practice: wording in consent form / information sheet – focus group

Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public.

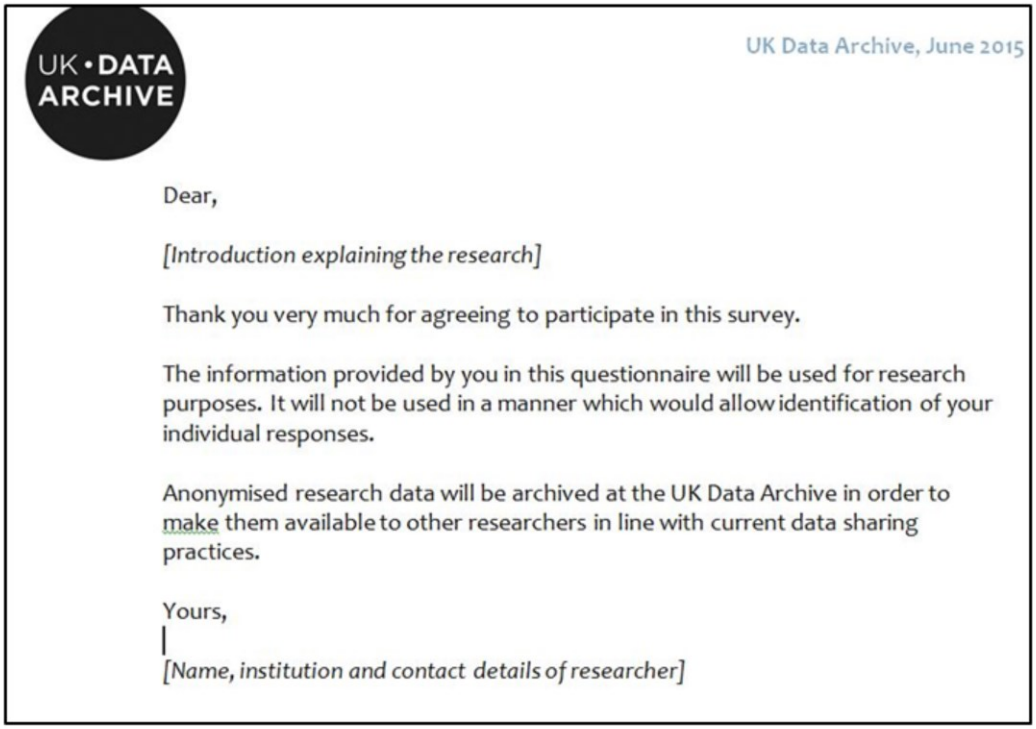
We ask you to consider the following points before agreeing to participate.

- Your contribution to the research will take the form of a focus group participant. This will be digitally video recorded and transcribed.
- **Your name and any information which may directly or indirectly identify you will be altered to protect your anonymity.**
- Any recordings of the discussions will be kept securely, and only authorised to other researchers on the condition they preserve your anonymity.
- The transcriptions (*excluding* names and other identifying details) will be retained by the researcher and analysed as part of the study. They will also be deposited with the UK Data Archive which has strict regulations about accessing data for research and protecting participant confidentiality.

ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms.aspx

UK Data Service

# In practice: wording in introductory letter – survey



UK Data Archive, June 2015

Dear,

*[Introduction explaining the research]*

Thank you very much for agreeing to participate in this survey.

The information provided by you in this questionnaire will be used for research purposes. It will not be used in a manner which would allow identification of your individual responses.

Anonymised research data will be archived at the UK Data Archive in order to make them available to other researchers in line with current data sharing practices.

Yours,

*[Name, institution and contact details of researcher]*

http://www.icpsr.umich.edu/icpsrweb/content/datamanage-ment/confientiality/conf-language.html

UK Data Service

# Anonymising quantitative data

- Direct identifiers – often not essential research info
- Indirect identifiers

- remove direct identifiers
  *e.g. names, address, institution, photo*
- reduce the precision/detail of a variable through aggregation
  *e.g. birth year vs. date of birth, occupational categories, area rather than village*
- generalise meaning of detailed text variable
  *e.g. occupational expertise*
- restrict upper lower ranges of a variable to hide outliers
  *e.g. income, age*

UK Data Service

# Anonymising qualitative data

- Remove direct identifiers, or replace with pseudonyms – often not essential research info
- Avoid blanking out; use pseudonyms or replacements
- Identify replacements, e.g. with [brackets]
- Plan or apply editing at time of transcription
- Avoid over-anonymising – removing information in text can distort data, make them unusable, unreliable or misleading; so balance anonymisation with the need to preserve context
- Consistency within research team and throughout project.
- Keep anonymisation log of all replacements or removals made – keep separate from anonymised data files

UK Data Service

# In practice: example anonymisation

**Ex 1. Health and Social Consequences of the Foot and Mouth Disease Epidemic in North Cumbria, 2001-2003** (study 5407 in UK Data Archive collection) by M. Mort, Lancaster University, Institute for Health Research.

Date of Interview: 21/02/02

Interview with Lucas Roberts, DEFRA field officer
Date of birth: 2 May 1965
Gender: Male
Occupation: Frontline worker
Location: Plumpton, North Cumbria

*Lucas was living at home with his parents, "but I'm hoping to move out soon" so we met at his parents' small neat house. We sat in a very comfortable sitting room with an open fire and Lucas made me coffee and offered shortbread. Although at first Lucas seemed a little nervous, quick to speech and very watchful he seemed to relax as we spoke and to forget abut the tape.*

**I will just start by asking you to tell me a little bit about yourself and your background.**

Well it is an agricultural background. I grew up on the farm where my brother is now. After I left school I did work on the farm but went to college and did exams, did land use recreation, sort of countryside/ environmental management course. So I obviously left agriculture, did the course and came back [to the farm] at weekends.

**Comment [v1]:** Replace: Ken
**Comment [v2]:** delete
**Comment [v3]:** delete
**Comment [v4]:** Replace: Ken
**Comment [v5]:** Replace: Ken
**Comment [v6]:** Replace: Ken

UK Data Service

# In practice: example anonymisation

Yeah. So is part of your job to look for funding bids and to write funding bits or is that separate?

No. That was what P3 used to do and then it sort of passed down to... really it's with P1 and P4. But I don't actually think there's anything out there at the moment. I think at the moment, because there's all this money saving and things, there's nothing to... there isn't actually anything to access.

M27UK

INT: Id like to start with your career history really. Perhaps you can start from when you finished school?

RES: First of all Id like to say that I do not think that my career is very traditional for someone who comes and works abroad. I graduated from mathematics in 'Country F'. That was equivalent to what is now BSc and MSc and then I did a separate MSc in Computer Science which at that time was called like Specialisation in computer science. And then I started working at the 'Country F' University as a teacher.

INT: Was that in 'City D'?

RES: No, my Specialisation in Computer Science was in 'City D' and my first degree, my mathematics degree was in another university in 'Country F' it is called [Text Cut]

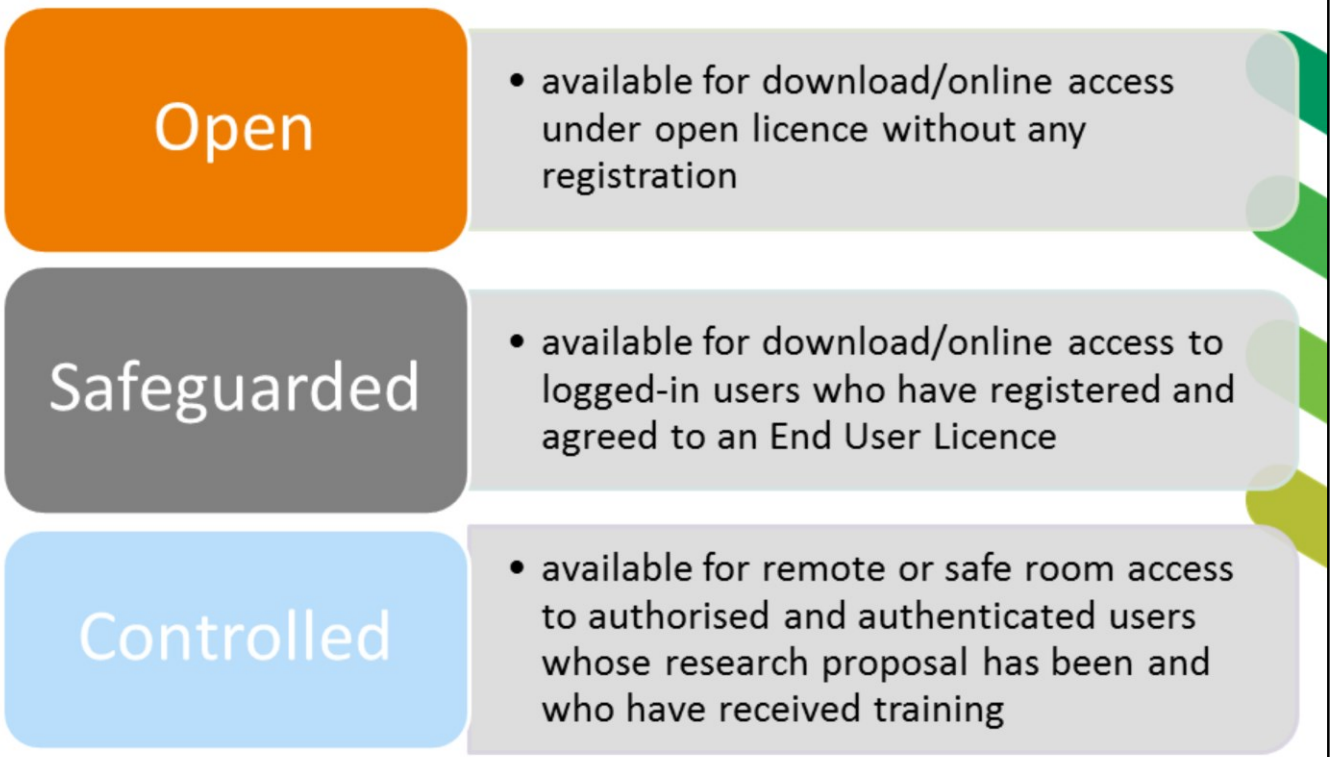
So after I finished in 'City D' my specialisation in computer science I went back to [Text Cut] where I did my first degree and I started working there as an assistant Professor.

UK Data Service

# What if anonymising is impossible?

- obtain consent for sharing non-anonymised data
- regulating/restricting user access, e.g. at UK Data Archive:
  - archived data NOT in public domain
  - use of data for specific purposes only after user registration
  - data users sign legally binding End User Licence – e.g. not identify any potentially identifiable individuals
  - stricter access regulations for confidential data (case by case basis):
    - access to approved researchers only
    - requiring data access authorisation from data owner prior to data release
    - confidential data under embargo for given time period
    - secure access to data
- researchers - consider access to data and safe storage

# Managing access to data



# Open about data with access restrictions

- Publish/advertise:
  - Which data exist
  - Where data are kept, e.g. which repository
  - Who can access them
  - For which purpose
  - Under which conditions

UK Data Service

# Organise your data

- Consistent/meaningful file names, no spaces/special characters
- Self-explanatory variable names, codes, abbreviations
- Label variables etc.
- Anonymise if needed (also file properties !)
- Check data for completeness

UK Data Service

# File formats

Digital data can take countless different form(at)s…

**A file format is a specific way of structuring information so that a machine, and therefore a person, can understand it**

- should be readable by as many types of system as possible
- without compromising the purpose of the data

UK Data Service

# File formats

Choice of software format for digital data:

- planned data analyses
- software availability/cost
- hardware used – e.g. audio capture
- discipline-specific standards and customs

Digital data is software dependent, so endangered by obsolescence of software/ hardware

Best formats for long-term preservation:

- standard, interchangeable, open
- *e.g. tab-delimited, comma-delimited (CSV), ASCII, RTF, PDF/A, OpenDocument format, XML*
- UK Data Archive optimal file formats for various data types
- Digital Preservation Coalition guidance on preservation formats

UK Data Service

# Document your data

- Sufficient information for future researchers to understand and use the data
- Data collection methodology and processes: sampling, sampling size, experiment protocol, instruments used, questionnaire, ….
- Links to reports, lab books, publications
- Readme file describing the dataset, dataset structure, content, how to use it,…

UK Data Service

# Key metadata

- Describe data collection
  - Title, abstract, keywords, coverage, collection method,….
- Find a data collection (discovery)
  - Keywords, subjects, controlled vocabularies, standardisation of metadata across repositories & disciplines (e.g. data registry)
- Assess use potential of data
  - Collection methods, sample size, data content,….
- Know how to access data
  - Access conditions, licence
- Citation of data use
  - Author/creator, publisher, dates
- Understand / interpret / use data
  - Variable labels and codes, file-level data content,…

UK Data Service

# Where can I publish my data?

**Discipline specific repositories and data centres**

- dedicated to archiving, preserving and disseminating discipline specific digital data
- e.g. UK Data Archive/UK Data Service, Environmental Information Data Centre, Visual Arts Data Service, Endangered Language Archive.. (the list goes on)

- UK Data Service has many components:
  - main collection is actively curated but reserved for high impact data, primarily from government departments
  - ReShare = repository for self-deposit for most research data offered to UK Data Service

UK Data Service

# Where can I publish my data?

**Generic repositories**

- Focus on sharing / publishing research data, often to support published papers
- e.g. Dryad, figshare, Zenodo
- Find a repository: Re3data

UK Data Service

# Where can I publish my data?

**Institutional repositories**

- Many universities now have an institutional repository where researchers can put their research outputs
- Generally designed for articles/publications, but increasingly providing for datasets too
- Bangor: Data eBangor Data http://edata.bangor.ac.uk/
- Jisc UK Research Data Discovery Service project: develop a national registry for research data, which will include records from institutional repositories and national data centres

UK Data Service

# Where can I publish my data?

**Data journals**

- a fairly new phenomenon, but growing
- publish a detailed journal style data article describing the data, how it was collected, reuse potential
- recommends or provides a place of deposit (repository)
- academic credit

Examples:

- Journal of Open Public Health Data
- Journal of Open Psychology Data
- Scientific Data

UK Data Service

# ReShare (UK Data Service)

- self-deposit data repository for research data
- part of UK Data Service
- remit: "*data resources of value to research and teaching communities across and beyond the social science discipline*" (Collections Development Policy)
- ESRC award holders: contractual requirement to archive and share their research data
- customised version of EPrints, open-source digital repository software
- developed via ESRC funding
- reshare.ukdataservice.ac.uk

# ReShare features

- UK Federation Access authentication of data depositors and data users
- Pull related project information from RCUK Gateway to Research (auto-fill 11 metadata elements: project-level, people, dates)
- Publish metadata records to UK Data Service Discover portal
- DOI minted
- UKDS staff review data collections before publishing:
  - confidentiality
  - copyright
  - documentation sufficient
  - long-term file formats

UK Data Service

# ReShare access and licence options

- Access options for individual files or zip bundles:
  - open access – CC licence
  - safeguarded access - UK Data Service End User licence
  - permission access
  - embargo (up to 12 months)

UK Data Service

# Deposit process

- Log in
- Create metadata record for data collection
- Upload data and documentation files
- Set access and licence conditions
- Submit

researcher

- Review
- Publish

UK Data Service

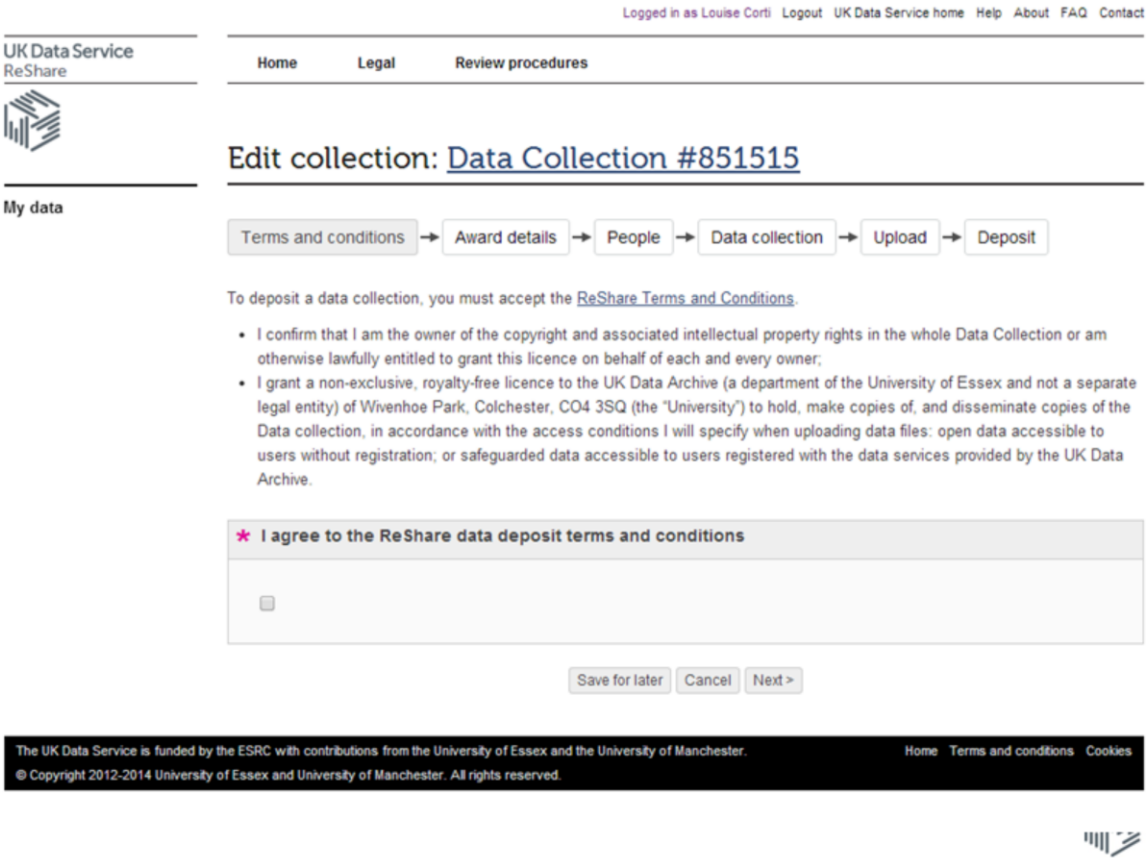
UK Data Service

# ReShare deposit process demo

A video tutorial of the ReShare deposit process is available at: https://www.youtube.com/watch?v=focv1z3lpPI

UK Data Service

# Easy to publish and upload data



# Example collection 1

- http://reshare.ukdataservice.ac.uk/851233/

UK Data Service

Home Legal Review procedures

# Global Uncertainties: Security In an Africa of Networked, Multi-Level Governance

**Leonard, David** 2014. Global Uncertainties: Security In an Africa of Networked, Multi-Level Governance. [Data Collection]. Colchester, Essex: Economic and Social Research Council. http://reshare.ukdataservice.ac.uk/id/eprint/851233

The programme of research centres on how the various institutions responsible for the production of security and the management of conflict in Sub-Saharan African societies do, could and should evolve in response to the presence of violent conflict. These institutions include: the armed forces, police, courts (civil as well as criminal), 'traditional' institutions of local governance, elections (and other methods of choosing leadership), international organisatoins (such as the United Nations, African Union, UN High Commission for Refugees, the World Court,), and the community of international donor nations. The programme is built on the observation that all governance (especially in Africa) is multi-leveled and networked - from the village to the international organisation, and well beyond what is specified in formal government structures. Thus the focus will be not only on the ways in which key conflict-management institutions evolve themselves but also on the changing ways in which the networks in which they are embedded actually operate. This leading edge research challenges theories about the state, state formation, and sovereignty as well as the dynamics of violent conflict as presented in the dominant current work on Africa. The primary methods of research will be network analysis and qualitative interviewing.

| Creators: | Creator Name | Email | Affiliation | ORCID |
|---|---|---|---|---|
| | Leonard, David | Unspecified | Institute of Development Studies | Unspecified |

| | |
|---|---|
| **Research funders:** | Economic and Social Research Council |
| **Grant reference:** | RES-071-27-0048 |
| **Subjects:** | Law, crime and legal systems<br>Politics |
| **Date deposited:** | 31 Jan 2014 19:00 |
| **Last modified:** | 22 Oct 2014 19:27 |

+ Coverage and Methodology

+ Access and Administration

UK Data Service

## Coverage and Methodology

| | | |
|---|---|---|
| **Collection period:** | **Date from:** | **Date to:** |
| | 1 April 2009 | 31 March 2013 |
| **Country:** | Sierra Leone | |
| **Data collection method:** | Group interviews in 39 villages and towns in rural Sierra Leone | |
| **Observation unit:** | Groups | |
| **Kind of data:** | Alpha-numeric, Textual | |
| **Type of data:** | Qualitative and mixed methods data | |
| **Resource language:** | English | |

## Access and Administration

| | | | | |
|---|---|---|---|---|
| **Data sourcing, processing and preparation:** | Citizen interviews | | | |
| **Copyright holders:** | **Name** | **Email** | **Affiliation** | **ORCID** |
| | Leonard, David | Unspecified | Institute of Development Studies | Unspecified |
| | , | Unspecified | Unspecified | Unspecified |
| **Contact:** | **Name** | **Email** | **Affiliation** | **ORCID** |
| | Leonard, David | leonard@berkeley.edu | Institute of Development Studies | Unspecified |
| **Notes on access:** | The names of individual respondents have been removed from the data set. | | | |
| **Publisher:** | Economic and Social Research Council | | | |
| **Last modified:** | 22 Oct 2014 19:27 | | | |

**AVAILABLE FILES**

**Data**

− Security_%26_Networks.xlsx

| | |
|---|---|
| **Accessible to:** | Registered users only (safeguarded data) |
| **File or bundle content:** | Data |
| **File or bundle description:** | Security & Networks |
| **File format:** | application/octet-stream |
| **License:** | UK Data Service End User Licence |
| **File size:** | 10Kb |

+ Sierra_Leone_Security_%26_Networks__Coded.xls

+ Copy_of_Somalia_SC_peace_initiatives.xlsx

**Documentation**

− Sierra_Leone_methods.doc

| | |
|---|---|
| **Accessible to:** | Anyone (open data) |
| **File or bundle content:** | Documentation |
| **File or bundle description:** | Sierra Leone methods |
| **File format:** | application/msword |
| **License:** | UK Data Service End User Licence |
| **File size:** | 64Kb |

**RELATED RESOURCES**

**Website**

Project outputs and publications

UK Data Service

# Preparing data files

- Consistent/meaningful file names, no spaces/special characters
- Self-explanatory variable names, codes, abbreviations
- Label all variables
- Remove temporary/dummy/admin variables
- Anonymise sufficiently (also file properties !)
- Check textual variables are OK to disseminate
- Missing values – label and consistency
- Check best file format for longevity
- Check data for completeness
- Bundle multiple files into zip bundles

UK Data Service

# Prepare documentation

- Sufficient information for future researchers to understand and use the data
- Data collection methodology and processes: sampling, sampling size, fieldwork protocol, interviewer instructions
- Information sheet / consent form
- Questionnaire, show-cards
- Links to reports, publications

UK Data Service

# Our data management guidance

- Online best practice guidance: [ukdataservice.ac.uk/manage-data.aspx](ukdataservice.ac.uk/manage-data.aspx)
- [Managing and Sharing Research Data – a Guide to Good Practice](): (Sage Publications Ltd)
- Helpdesk for queries: [ukdataservice.ac.uk/help/get-in-touch.aspx](ukdataservice.ac.uk/help/get-in-touch.aspx)
- Training: [www.data-archive.ac.uk/create-manage/advice-training/events](www.data-archive.ac.uk/create-manage/advice-training/events)

UK Data Service

# Our guidance

- plan to share research data
- legal and ethical aspects of data sharing and reuse
- data copyright
- documentation and metadata to understand and use data
- file formats, organising, versioning and quality control
- storage, backup, encryption and security of data and files
- strategies for collaborative research

UK Data Service

# Other resources

- RDM guidance at Bangor: http://www.bangor.ac.uk/library/info_staff/ResearchDataManagement.php.en
- Research Data Management Training MANTRA (Edinburgh) – online learning units, also e.g. RDM guidance for N-Vivo http://datalib.edina.ac.uk/mantra/
- NASA Distributed Active Archive Centre: http://daac.ornl.gov/PI/pi_info.shtml
- Digital Curation Centre:
  - Data management planning http://www.dcc.ac.uk/resources/data-management-plans
  - Data management training / courses http://www.dcc.ac.uk/training/data-management-courses-and-training

UK Data Service

# Overview UK Data Service – some statistics

Holdings: data for research and teaching purposes, used in all sectors and for many different disciplines

- **6,000** datasets in the collection
- **25,000** registered users
- **60,000** downloads worldwide p.a.
- **4000+** user support queries p.a.

UK Data Service

# Overview service – data sources

- Official agencies - mainly central government
- International statistical time series
- Individual academics - research grants
- Market research agencies
- Public records/historical sources
- Access to international data via links with other data archives worldwide

UK Data Service

# What do users do with the data ?

- Comparative research, restudy or follow-up study
- Re-analysis/secondary analysis
- Research design and methodological advancement
- Replication of published statistics
- Teaching and learning

UK Data Service

# Types of data collections

- Survey microdata
  - cross-sectional
  - Panel/ Longitudinal
  - International microdata
- Aggregate statistics
  - International macrodata
- Census data
  - Aggregate data for 1971 -2011
  - Microdata for 1991and 2001 (2011 data are forthcoming)
- Qualitative and mixed methods data
  - in-depth interview transcripts, diaries, anthropological field notes, answers to open-ended survey questions, audio-visual recordings and images (mixed methods: combined with numeric data)

UK Data Service

# Overview service – key data

**Government**

Large-scale government funded surveys

**Longitudinal**

Major UK surveys following individuals over time

**International**

Multi-nation aggregate databanks and survey data

**Census**

Census data 1971 - 2011

**Business**

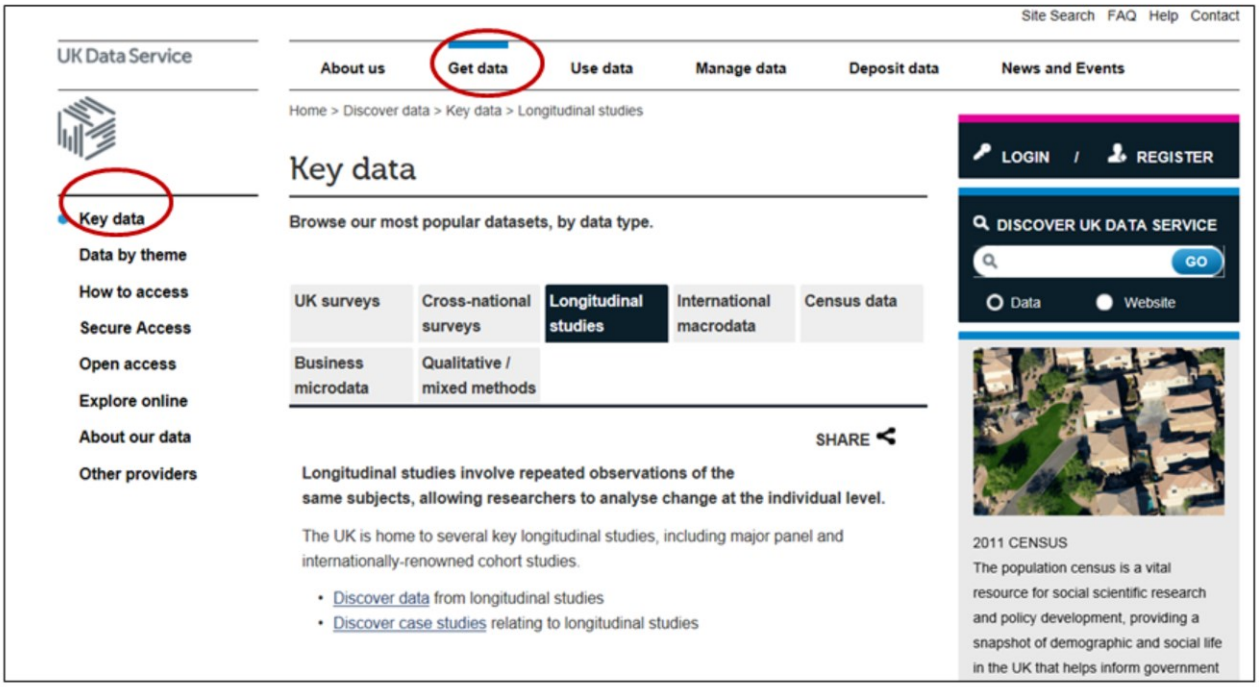
Sensitive data requiring secure access systems

**Qualitative**

Range of multi-media qualitative data sources

UK Data Service

# Key data




UK Data Service

# Survey Microdata

- Individual or household level data which have been anonymised to prevent disclosure
- Needs to be analysed using an analyses package such as SPSS or Stata
- Very flexible as you can produce your own populations and tables.

UK Data Service

# Large-scale UK surveys

- General Lifestyle Survey (General Household Survey)
- Labour Force Survey
- Health Survey for England/Wales/Scotland
- Living Costs and Food Survey (Expenditure and Food Survey)
- Crime Survey for England and Wales (British Crime Survey)
- Family Resources Survey
- Opinions and Lifestyle Survey (ONS Opinions Survey, ONS Omnibus Survey)
- English Housing Survey (Survey of English Housing)
- British Social Attitudes
- National Travel Survey

UK Data Service

# Longitudinal data

- Longitudinal surveys involve repeated surveys of the same individuals at different points in time
- Allow researchers to analyse change at the individual level
- More complex to manage and analyse

UK Data Service

# Longitudinal data

Range of longitudinal data collections, but 9 'core' collections in particular:

- National Child Development Study (NCDS)
- 1970 British Cohort Study (BCS70)
- Millennium Cohort Study (MCS)
- British Household Panel Survey (BHPS)
- Understanding Society (UKHLS)
- Families and Children Study (FACS)
- English Longitudinal Study of Ageing (ELSA)
- Longitudinal Study of Young people in England (LSYPE)
- Growing Up in Scotland (GUS)

UK Data Service

# The UK Household Longitudinal Study

- Started in 2009: UKHLS or 'Understanding Society' ; First data: 2010, 3 waves so far
- Study of the socio-economic circumstances and attitudes of 100,000 individuals in 40,000 British households, largest household panel survey in the world
- UKLHS replaces and incorporates the BHPS (1991-2009 , 5500 household's), retaining the latter's longevity whilst adding
  - to the sample size
  - to the scope of the study (new components: innovation panel, greater detail on ethnic minority groups, qualitative and biomedical data collections)
- coverage includes:
  current employment and earnings; employment status; parenting and childcare arrangements; family networks; benefit payments; political party identification; household finances; environmental behaviours; consents to administrative data linkage (health and education).

UK Data Service

# International data

International data is split into two types of data:

- Aggregate or macrodata
  … data that has been aggregated to a country or regional level. This type of data is typically produced by inter-governmental organisations like the UN or World Bank. All the aggregate databanks contain time series data.

- Survey or microdata
  … data about individuals. This type of data is typically produced by an organisation in the academic sector and looks at values and attitudes.

UK Data Service

# International macrodata

- time series data aggregated to country/region
- regularly updated
- international governmental organisations (IMF, OECD, IEA, World Bank)
- wide range of socio-economic topics
- only available to staff and students from UK institutions of higher and further education (World Bank data is open access)

UK Data Service

The first set of data we will tell you about is the international macrodata.
This type of data has been aggregated to a country or regional level. It is time series data, depending on the database it will be available annually, quarterly or monthly.
We update the data regularly, some databases we updated every month
It is typically produced by international governmental organisations, and we host data from the IMF, OECD, World Bank and the IEA.
The databases cover a range of socio-economic themes such as economics, trade, industry, employment, human development, demography, and many more!
Much of the data is only available to staff and students from UK higher/further education institutions. We pay for the licence cost centrally to provide access to this community, so it is free at the point of use. However, data such as the World Bank databases are now open access – so anyone can access these through us.

# International data themes

Databanks cover:

- Economic performance and development
- Trade, industry and markets
- Employment
- Demography, migration and health
- Governance
- Human development
- Social expenditure
- Education
- Science and technology
- Land use and the environment

UK Data Service

# Links with other data archives worldwide

CESSDA - Council of European Social Science Data Archives

Non-European: ICPSR, ASSDA

UK Data Service

# Qualitative data

The UK Data Service collections contain examples from a range of qualitative methods including:

- interview transcripts
- visual data
- focus groups
- essays
- diaries
- online data
- observation notes
- documents
- audio data
- newspaper clippings

etc.

Interview transcripts are the most commonly deposited qualitative data with the UK Data Archive.

Most contemporary data is deposited in a digital form, but we also hold (and sometimes digitise) legacy data which is a paper (handwritten or typewritten) format.

UK Data Service

# Qualitative data

**Example**

Ray Pahl, SN 4867: School Leavers Study, 1978

Teachers at a comprehensive school on the Isle of Sheppey were asked to set a particular essay to those pupils who were students in English lessons about ten days before they were due to leave school. The students were asked to imagine that they were nearing the end of their life, and that something had made them think back to the time when they left school. They were then asked to write an imaginary account of their life over the next 30 or 40 years.

**The resulting data:** 142 handwritten essays in 1978 by school leavers aged 15 and 16 years old.

UK Data Service

# UK census data

- 1971-2011 census data
- gold standard
- used as baseline for other statistics
- detailed combinations of characteristics
- small geographies
- census outputs
  - Aggregate data
  - Boundary data
  - Flow data
  - Microdata
- restricted to UK higher/further education staff and students, except for the microdata

UK Data Service

Next I'll talk about the UK census data, the census has a long history going back to 1801, it takes place every 10 years. We provide access to and user support for the last 5 censuses.
Census data is considered the gold standard, aims to cover 100% of the population
Used as a baseline for other statistics
Contains detailed combinations of characteristics, and there is a range of topics, which I'll mention on my next slide
The data is available in many geographies, right down to small geographies. This makes it possible to reliably compare different areas which makes it so unique
Census data can be divided into different outputs, such as aggregate, boundary, flow, and microdata – more about this later
Much of the census data we hold is restricted to UK higher/further education staff and students, except for the microdata, details on the website

# Finding and accessing data

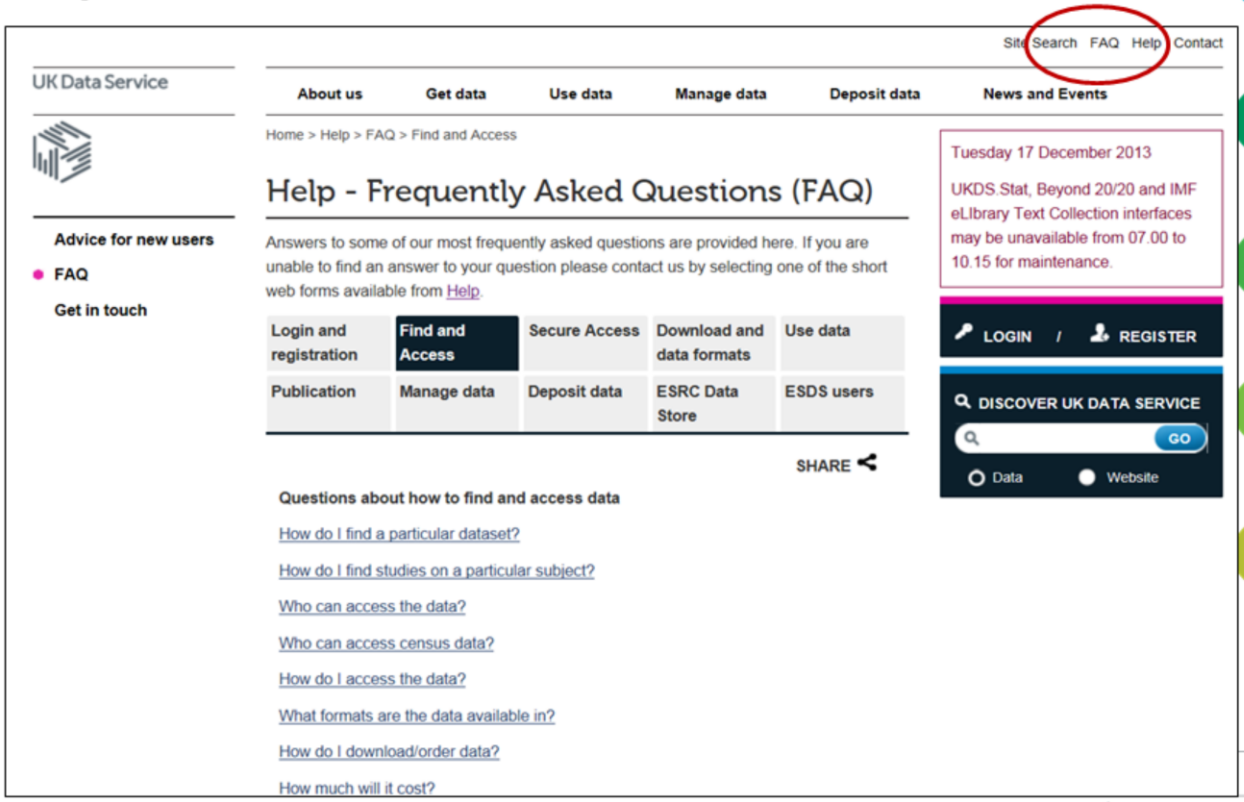
UK Data Service

# Data access

- Web access to data and metadata via the UK Data Service web pages
- Documentation (codebooks, questionnaires) freely available to anyone browsing the web pages
- Data are freely available for use in higher education institutions
- Data supplied in a variety of formats
  - statistical package formats (e.g. SPSS, STATA)
  - databases and spreadsheets
  - word processed documents, PDF documents etc.

UK Data Service

# Help


Site Search FAQ Help Contact
UK Data Service
About us
Get data
Use data
Manage data
Deposit data
News and Events
Home > Help > FAQ > Find and Access
Help - Frequently Asked Questions (FAQ)
Advice for new users
FAQ
Get in touch
Answers to some of our most frequently asked questions are provided here. If you are unable to find an answer to your question please contact us by selecting one of the short web forms available from Help.
Login and registration
Find and Access
Secure Access
Download and data formats
Use data
Publication
Manage data
Deposit data
ESRC Data Store
ESDS users
SHARE
Questions about how to find and access data
How do I find a particular dataset?
How do I find studies on a particular subject?
Who can access the data?
Who can access census data?
How do I access the data?
What formats are the data available in?
How do I download/order data?
How much will it cost?
Tuesday 17 December 2013
UKDS.Stat, Beyond 20/20 and IMF eLIbrary Text Collection interfaces may be unavailable from 07.00 to 10.15 for maintenance.
LOGIN / REGISTER
DISCOVER UK DATA SERVICE
GO
Data
Website

# UK Surveys

- data about individuals or households
- large sample sizes
- nationally-representative
- repeated cross-sectional surveys
- many surveys are repeated every year
- well documented
- cover many topics including health, work, crime, social attitudes, family expenditure, living costs, housing etc.

UK Data Service

• Some of the most commonly used of our micro data are the UK Surveys

•These are mostly about individuals or households.
The surveys produce very high quality data: They are produced by experienced research organisations such as the ONS and the National Centre for Social Research who have had many years of experience in sampling, data collection, analyses etc. They have their own methodologists, researchers and teams of interviewers.

•They are mostly nationally representative at the UK level, or the countries within the UK level depending on the particular survey. And they tend to have large sample sizes, For example, Labour Force Survey interviews 60,000 people every quarter.
These datasets are repeated cross-sectional surveys – many of them are repeated every or most years but each time they are run they use a new sample of people.
They often use the same questions each time they are run so they can be used to track trends in the population over time.

# Census data

UK Data Service

So we're going to move on to a different type of macro data, and this is data that comes from the census
This is the Census homepage

# UK census data

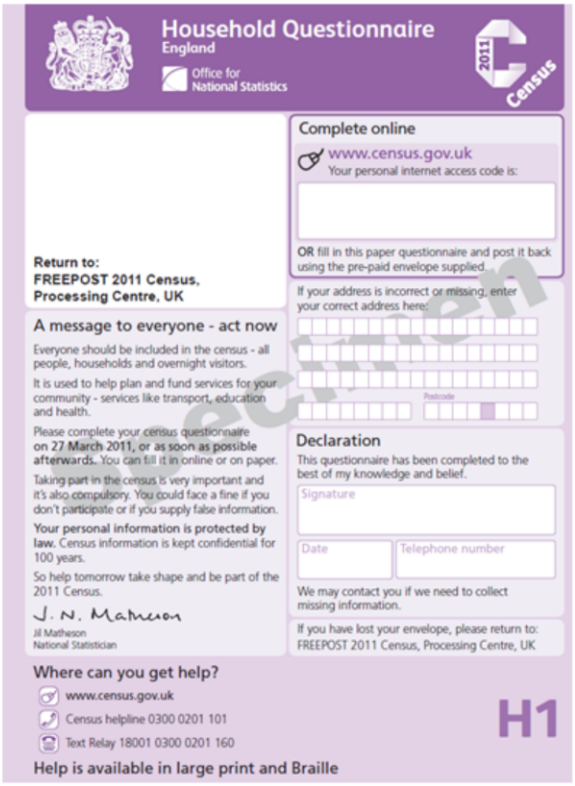
- 1971-2011 census data
- gold standard
- used as baseline for other statistics
- detailed combinations of characteristics
- small geographies
- census outputs
  - Aggregate data
  - Boundary data
  - Flow data
  - Microdata
- Mostly open (some restrictions may apply)

UK Data Service

the census has a long history going back to 1801, it takes place every 10 years. We provide access to and user support for the last 5 censuses.
Census data is considered the gold standard, aims to cover 100% of the population
Used as a baseline for other statistics
Contains detailed combinations of characteristics, and there is a range of topics, which I'll mention on my next slide
The data is available in many geographies, right down to small geographies. This makes it possible to reliably compare different areas which makes it so unique
Census data can be divided into different outputs, such as aggregate, boundary, flow, and microdata – more about this later

Most of the data are now open, although there may be some restrictions on some elements of the census data that we have. Unfortunately the licences under which we hold these data may be restricted to those in higher education

# Topics



Household Questionnaire
England
Office for National Statistics
2011 Census

Complete online
www.census.gov.uk
Your personal internet access code is:

OR fill in this paper questionnaire and post it back using the pre-paid envelope supplied.

Return to:
FREEPOST 2011 Census,
Processing Centre, UK

If your address is incorrect or missing, enter your correct address here:

Postcode

A message to everyone - act now

Everyone should be included in the census - all people, households and overnight visitors.

It is used to help plan and fund services for your community - services like transport, education and health.

Please complete your census questionnaire on 27 March 2011, or as soon as possible afterwards. You can fill it in online or on paper.

Taking part in the census is very important and it's also compulsory. You could face a fine if you don't participate or if you supply false information.

Your personal information is protected by law. Census information is kept confidential for 100 years.

So help tomorrow take shape and be part of the 2011 Census.

J. N. Matheson
Jil Matheson
National Statistician

Declaration

This questionnaire has been completed to the best of my knowledge and belief.

Signature

Date

Telephone number

We may contact you if we need to collect missing information.

If you have lost your envelope, please return to:
FREEPOST 2011 Census, Processing Centre, UK

Where can you get help?

www.census.gov.uk
Census helpline 0300 0201 101
Text Relay 18001 0300 0201 160

Help is available in large print and Braille

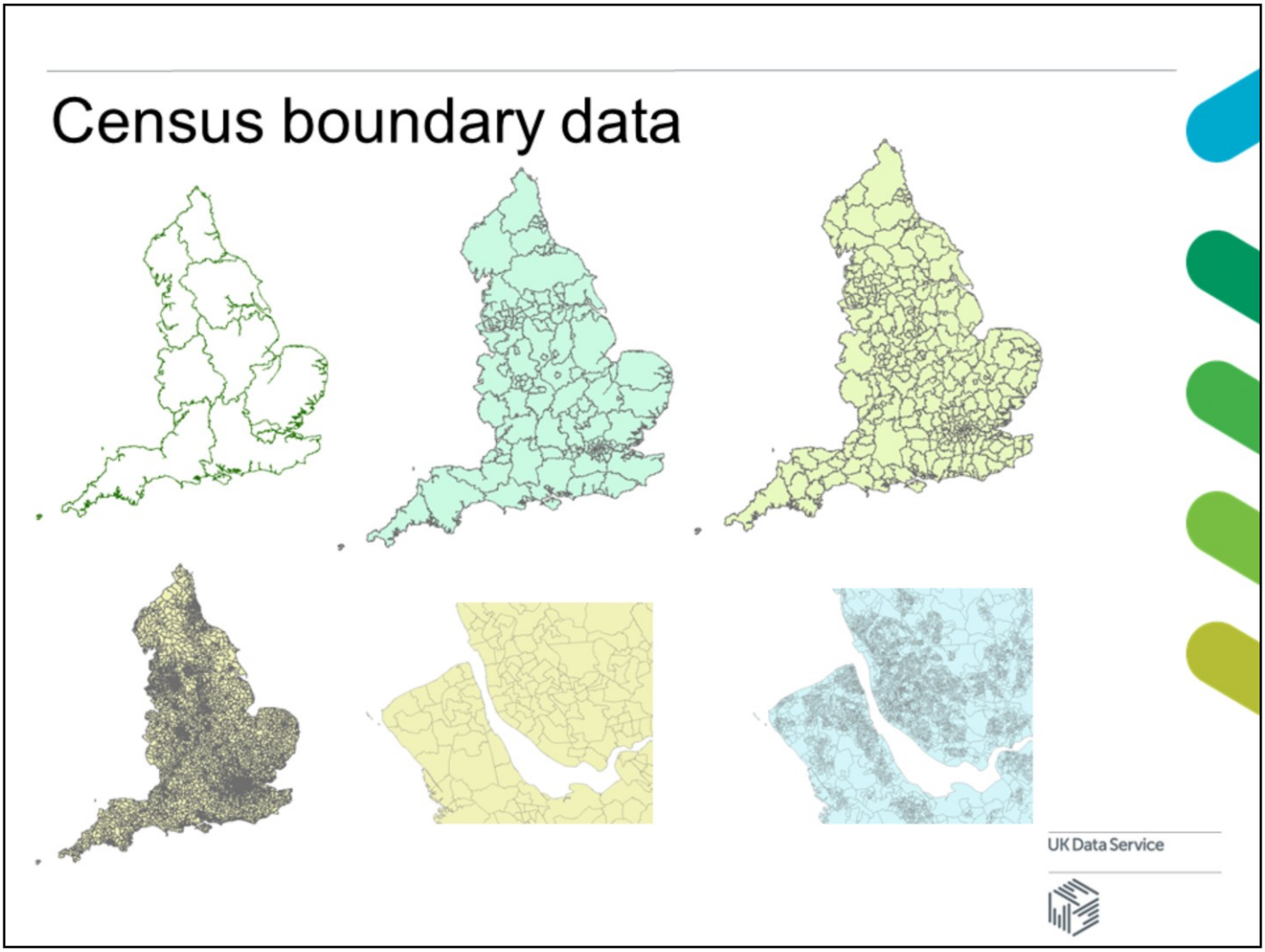
H1

Specimen

- age
- sex
- health
- religion
- ethnicity
- caring
- travel to work
- household composition
- housing
- car ownership
- employment
- education
- national identity
- second homes
- intention to stay

UK Data Service

Topics vary each census, and in the 2011 Census, new topics such as national identity and whether someone has a second home were included




We also provide access to various (1981 to 2011) census boundaries for a range of geographies.

The first map that you see on the top left – is the largest, regions and then you can zoom in to smaller and smaller geographical areas until you get to wards, and the smallest area is super output areas. Here you can see an example where we've zoomed into the Wirral.

The boundaries are available in  a range of GIS formats.!

# Census microdata

- individual-level data which has been anonymised to prevent disclosure
- Sample of Anonymised Records (SARs)

- looks like data you would get from doing a survey yourself
- analysed using an analysis package like SPSS or Stata
- more flexible as you can produce your own tables, populations and attributes
- 1991, 2001 and 2011 SARs data available

So moving on to our Census micro data…..
These are like surveys, but they've been produced from the census

Individual-level data which has been anonymised to prevent disclosure, and to protect confidentiality
Sample of 1-5% depending on dataset

These look like the data you would get from a social survey & you would analyse them in the same way using packages such as SPSS

# Qualitative data

Qualitative data can take a number of different formats: interview transcripts, visual data, focus groups, essays, diaries, online data, observation notes, documents, audio data, open-ended survey questions, case notes, newspaper clippings etc.

Search Discover for Qualitative and mixed methods data

Examples of sociology data collections:

- Family Life and Work Experience before 1918, Middle and Upper Class Families in the Early 20th Century, 1870-1977 (SN 5404)
- Gender Difference, Anxiety and the Fear of Crime, 1995 (SN 4581)
- Mothers Alone: Poverty and the Fatherless Family, 1955-1966 (SN 5072)
- Affluent Worker in the Class Structure, 1961-1962 (SN 6512)

UK Data Service

So just mentioning very quickly qualitative data as well..

All of the data sources we've looked at so far have been statistical or numeric data sources,
Qualitative data contains many different kinds of data –
the most common is interview transcripts, although we also have a small amount of visual data as well as information that has been derived from focus groups, diaries, records of meetings etc.

You can search our Discover catalogue for qualitative data sources, but here's a list of some of our classic sociology data collections

# An example of qualitative data

Ray Pahl, SN 4867: *School Leavers Study*, 1978

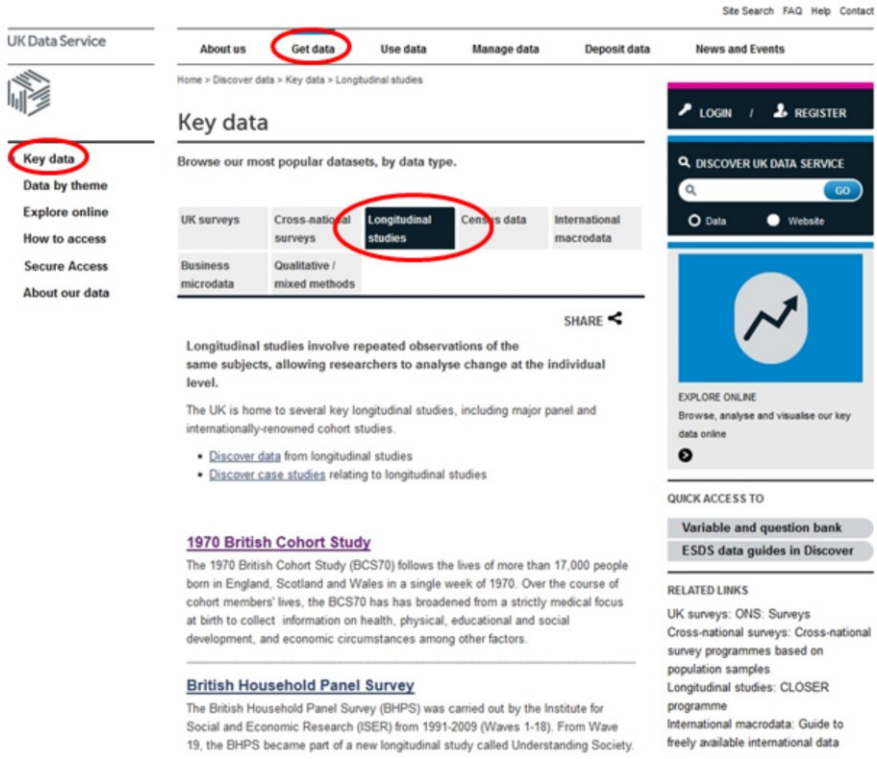
Teachers at a comprehensive school on the Isle of Sheppey were asked to set a particular essay to those pupils who were students in English lessons about ten days before they were due to leave school. The students were asked to imagine that they were nearing the end of their life, and that something had made them think back to the time when they left school. They were then asked to write an imaginary account of their life over the next 30 or 40 years.

**The resulting data:** 141 handwritten essays in 1978 by school leavers aged 15 and 16 years old.

UK Data Service

Here's one example of a qualitative study, taken from the School Leavers Study of 1978 that was deposited with us by Ray Pahl. This was a study conducted on the Isle of Sheppey in Kent where students aged 15-16 years were asked to imagine that they were at the end of their life and were thinking back over their lives. And these data are available through us in the form of essays.

# Useful tips: Discover – key data




If you want to search for any of our key data, you can do this through our Key Data pages.

The way to find your way to this section, is to use the Get data menu – click on Key Data, and then select the tab for whatever type of data you are looking for. And then a list of key data sources will be listed

# Accessing data

- Web access to data and metadata
- Data are freely available to students in higher education institutions
- Data supplied in a variety of formats
  - statistical package formats (e.g. SPSS, STATA)
  - databases and spreadsheets
  - word processed documents, PDF documents etc.

UK Data Service

So you may be wondering how you can access all these data sources….

Everything is done online, and you can access both the data and the metadata through our website – and this is freely available to look at so you can have a look at our data catalogue, download any user guides, questionnaires, technical reports, that are associated with a particular survey.

And the vast majority of data are freely available to anyone who registers with the UK Data Service. There may be some collections where the licence has been restricted so they may not be available to anyone, but most of our data are available under our End User Licence so you can download them in a variety of packages…

# Registration

- Click on the 'login' link on the UK Data Service homepage
- Click on 'login to the UK Data Service'
- Find your university in the list of institutions and proceed to the login page (or your UK Data Archive credentials if you are not at a UK University)
- Enter your login details
- Complete the registration form as a new user
- Accept the End User Licence
- Use the download/order link for the dataset you wish to access and create a project description
- Download the data in your chosen format

UK Data Service

DEMONSTRATION HERE

# Accessing data

- If you are at a UK university, you can use your own institutional username and password to log in and register with the UK Data Service
- If you are not at a UK University you will have to apply for a UK Data Archive username and password

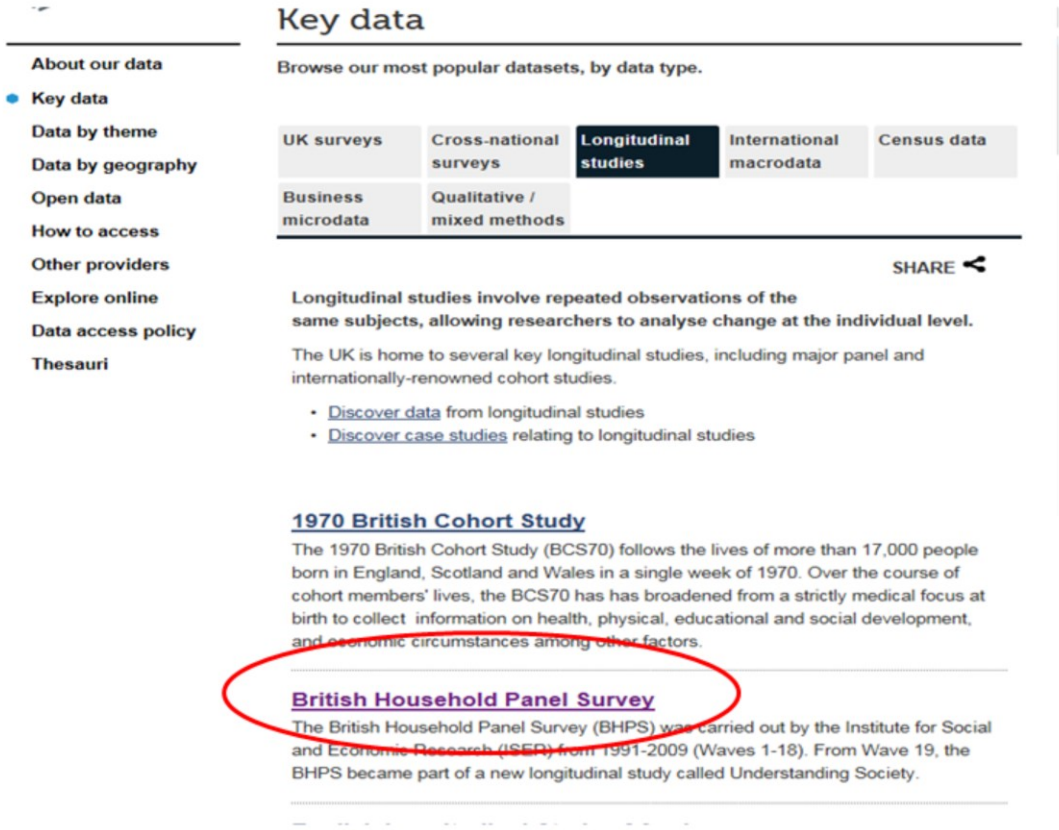
UK Data Service

The way you register with us can vary depending on which particular sector you come from…

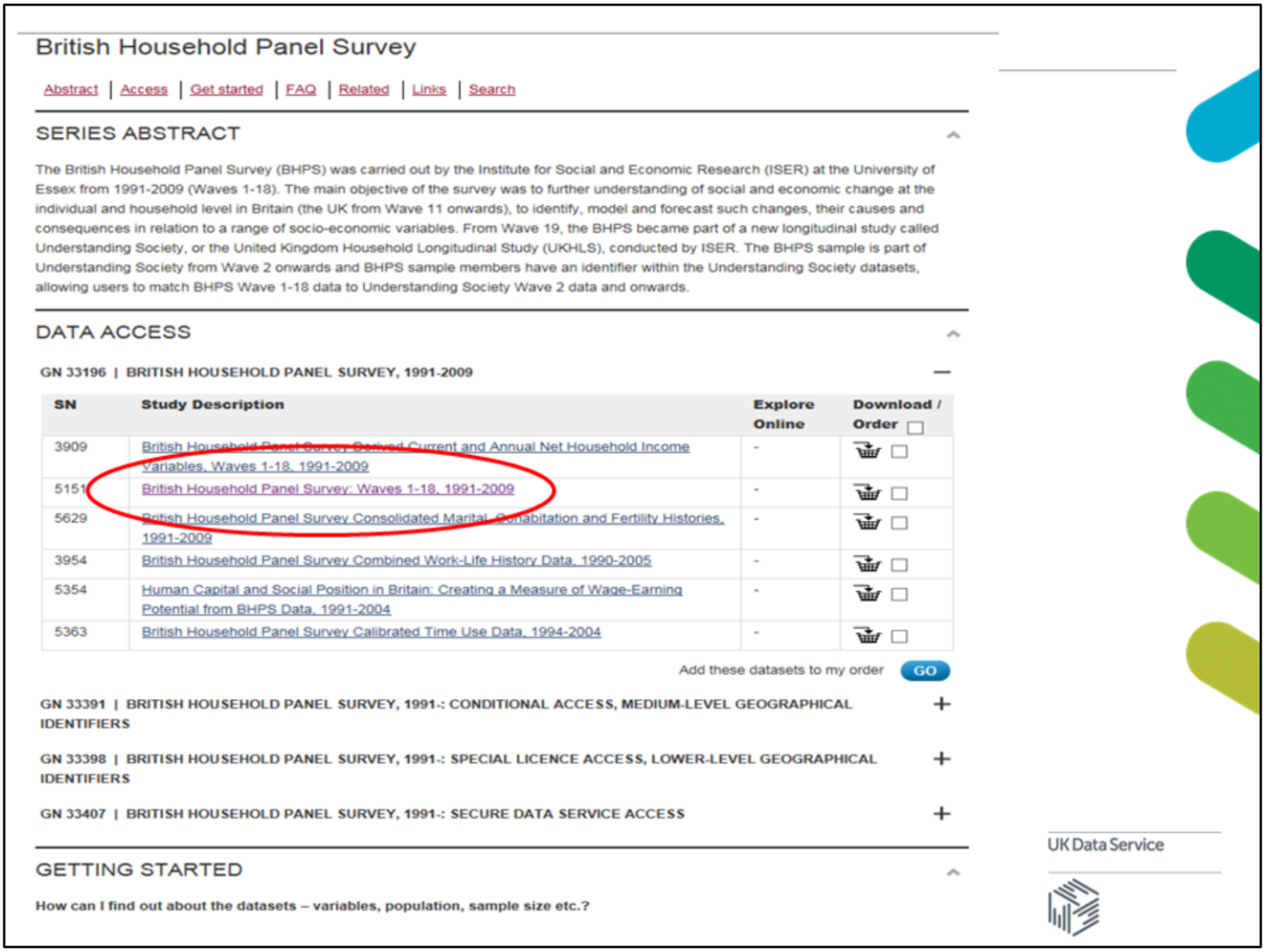
So if you're at a UK university, then you can use your own institutional username and password to log in and register with the UK Data Service

If you are outside a UK university, then you can still register with us, but you will first have to apply for a UK Data Archive username & password – so you just have a small extra step to go through.

# Accessing data: Step by step




UK Data Service



British Household Panel Survey

Abstract | Access | Get started | FAQ | Related | Links | Search

SERIES ABSTRACT

The British Household Panel Survey (BHPS) was carried out by the Institute for Social and Economic Research (ISER) at the University of Essex from 1991-2009 (Waves 1-18). The main objective of the survey was to further understanding of social and economic change at the individual and household level in Britain (the UK from Wave 11 onwards), to identify, model and forecast such changes, their causes and consequences in relation to a range of socio-economic variables. From Wave 19, the BHPS became part of a new longitudinal study called Understanding Society, or the United Kingdom Household Longitudinal Study (UKHLS), conducted by ISER. The BHPS sample is part of Understanding Society from Wave 2 onwards and BHPS sample members have an identifier within the Understanding Society datasets, allowing users to match BHPS Wave 1-18 data to Understanding Society Wave 2 data and onwards.

DATA ACCESS

GN 33196 | BRITISH HOUSEHOLD PANEL SURVEY, 1991-2009

| SN | Study Description | Explore Online | Download / Order |
|---|---|---|---|
| 3909 | British Household Panel Survey: Derived Current and Annual Net Household Income Variables, Waves 1-18, 1991-2009 | - | |
| 5151 | British Household Panel Survey: Waves 1-18, 1991-2009 | - | |
| 5629 | British Household Panel Survey Consolidated Marital, Cohabitation and Fertility Histories, 1991-2009 | - | |
| 3954 | British Household Panel Survey Combined Work-Life History Data, 1990-2005 | - | |
| 5354 | Human Capital and Social Position in Britain: Creating a Measure of Wage-Earning Potential from BHPS Data, 1991-2004 | - | |
| 5363 | British Household Panel Survey Calibrated Time Use Data, 1994-2004 | - | |

Add these datasets to my order GO

GN 33391 | BRITISH HOUSEHOLD PANEL SURVEY, 1991-: CONDITIONAL ACCESS, MEDIUM-LEVEL GEOGRAPHICAL IDENTIFIERS

GN 33398 | BRITISH HOUSEHOLD PANEL SURVEY, 1991-: SPECIAL LICENCE ACCESS, LOWER-LEVEL GEOGRAPHICAL IDENTIFIERS

GN 33407 | BRITISH HOUSEHOLD PANEL SURVEY, 1991-: SECURE DATA SERVICE ACCESS

GETTING STARTED

How can I find out about the datasets – variables, population, sample size etc.?

UK Data Service

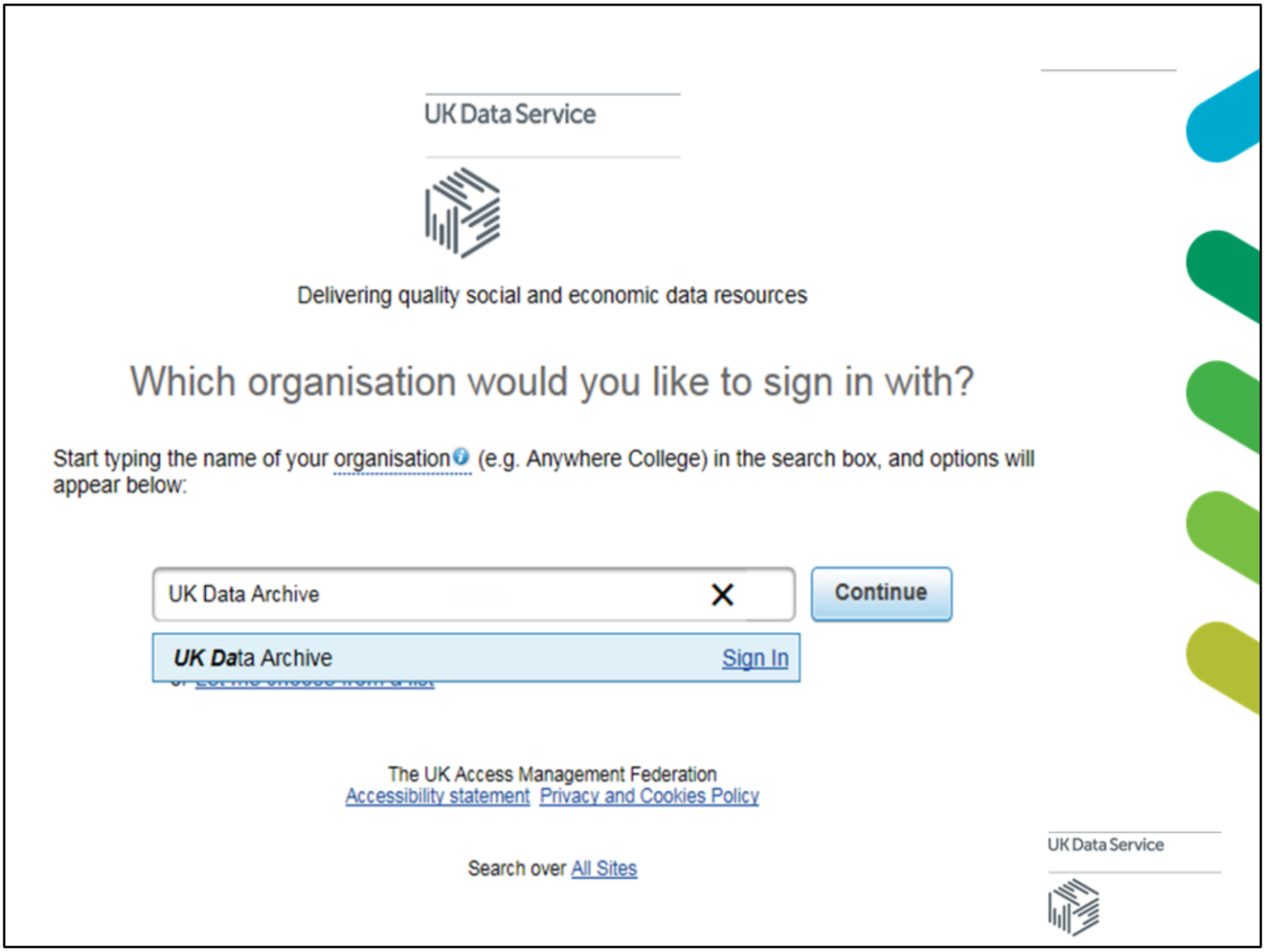
http://discover.ukdataservice.ac.uk/catalogue/?sn=5151&type=Data%20catalogue

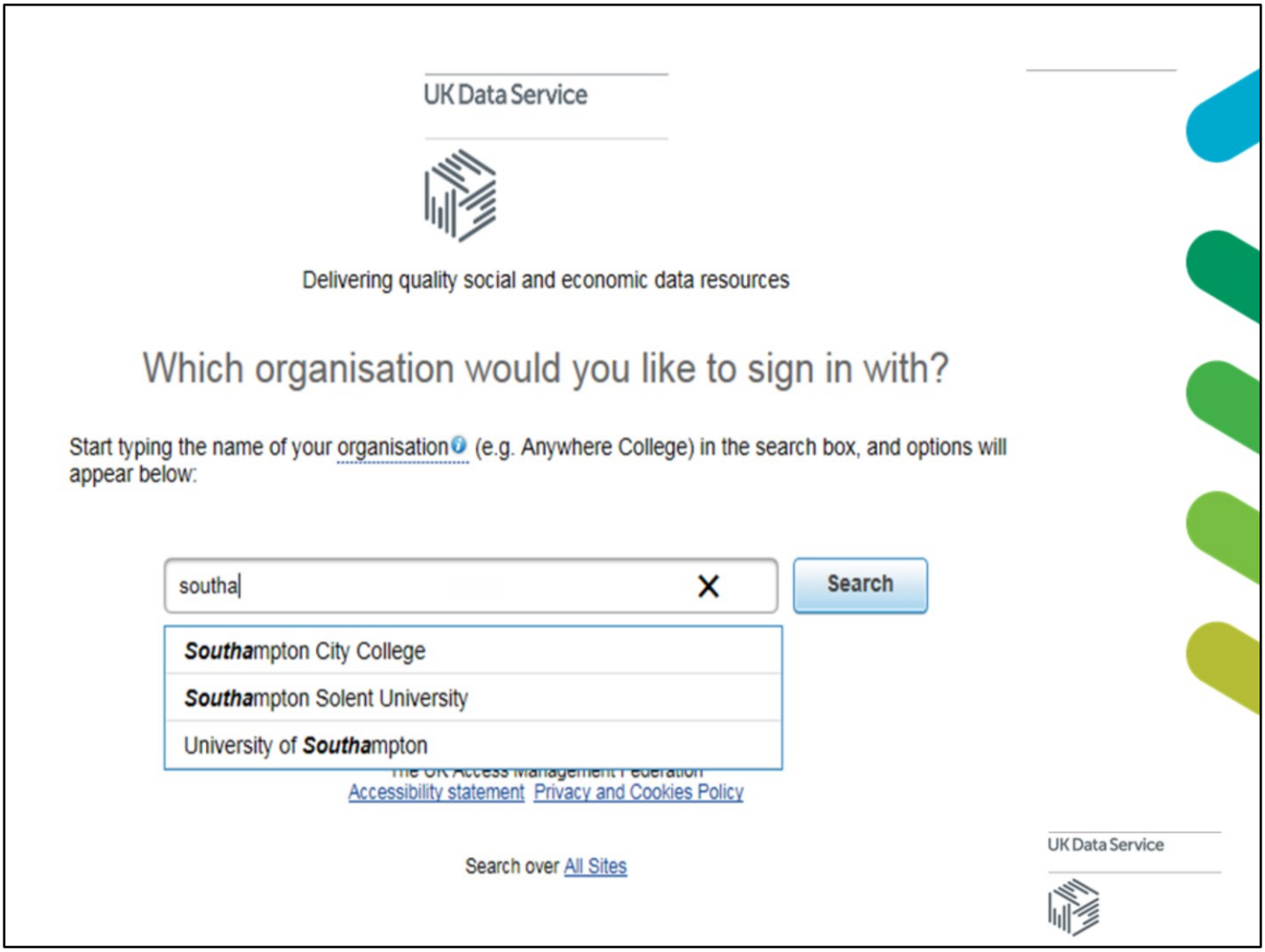
**For approved researchers using controlled data**

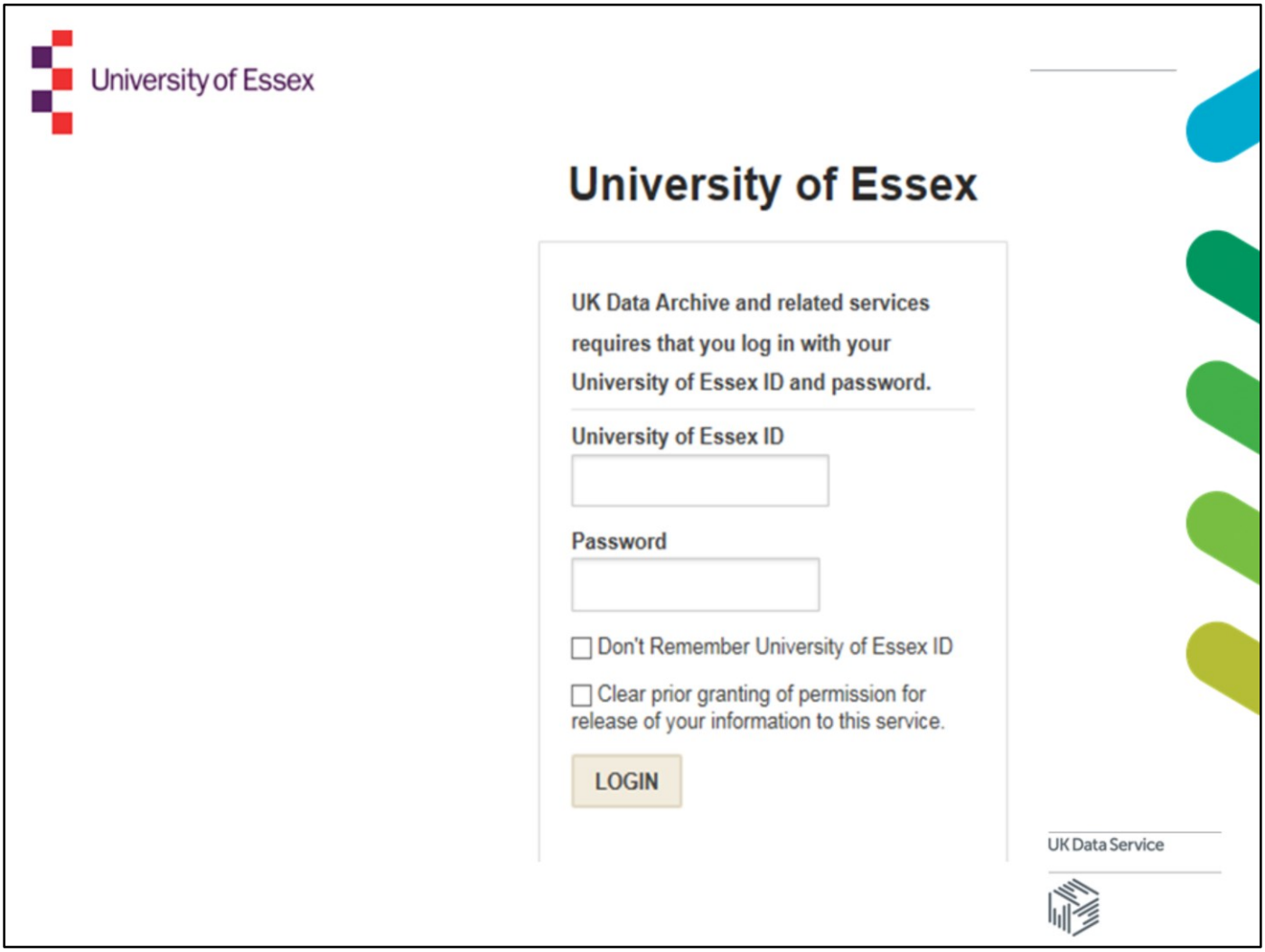
Use the Secure Lab login which is only available to researchers who have successfully applied for access to restricted data which cannot be downloaded.

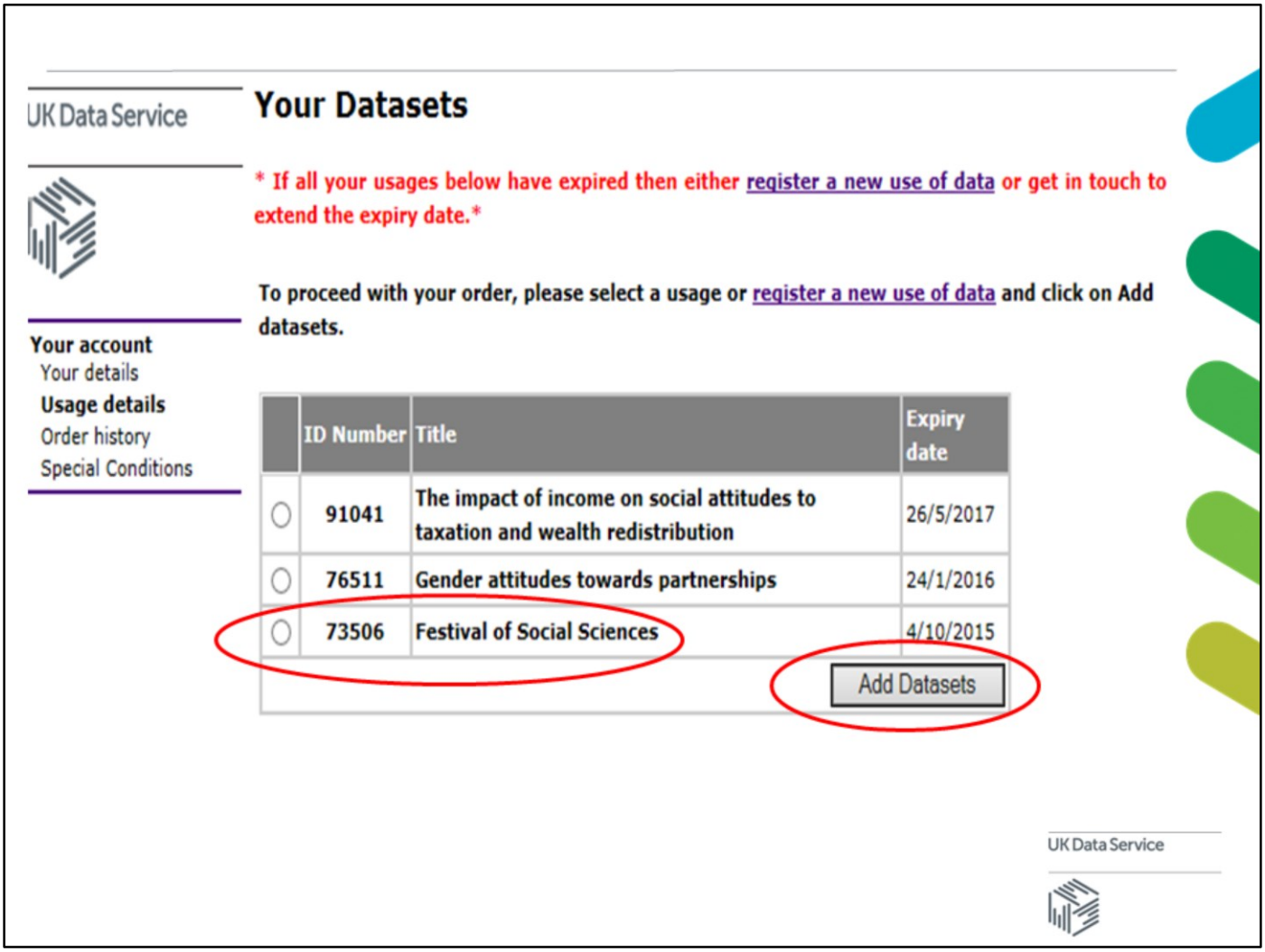
PLEASE NOTE: The Secure Lab will be unavailable every Tuesday between 10.00 and 12.00 for maintenance.

UK Data Service


UK Data Service
Delivering quality social and economic data resources
Which organisation would you like to sign in with?
Start typing the name of your organisation (e.g. Anywhere College) in the search box, and options will appear below:
UK Data Archive
Continue
UK Data Archive
Sign In
The UK Access Management Federation
Accessibility statement
Privacy and Cookies Policy
Search over All Sites
UK Data Service


UK Data Service
Delivering quality social and economic data resources
Which organisation would you like to sign in with?
Start typing the name of your organisation (e.g. Anywhere College) in the search box, and options will appear below:
southa
Search
Southampton City College
Southampton Solent University
University of Southampton
Accessibility statement
Privacy and Cookies Policy
Search over All Sites
UK Data Service


University of Essex
University of Essex
UK Data Archive and related services requires that you log in with your University of Essex ID and password.
University of Essex ID
Password
Don't Remember University of Essex ID
Clear prior granting of permission for release of your information to this service.
LOGIN
UK Data Service


UK Data Service
Your Datasets
* If all your usages below have expired then either register a new use of data or get in touch to extend the expiry date.*
To proceed with your order, please select a usage or register a new use of data and click on Add datasets.
Your account
Your details
Usage details
Order history
Special Conditions
ID Number
Title
Expiry date
91041
The impact of income on social attitudes to taxation and wealth redistribution
26/5/2017
76511
Gender attitudes towards partnerships
24/1/2016
73506
Festival of Social Sciences
4/10/2015
Add Datasets
UK Data Service

# Datasets for usage 93579 - UKDS Webinars

Use the download column to download now or request for download or the other media column to e.g. request data on CD. Add more datasets to this usage below.

| SN | Study Description | Status | Download | Explore Online | Other media |
|---|---|---|---|---|---|
| 5151 | British Household Panel Survey: Waves 1-18, 1991-2009 | | Download | - | ☐ |
| | | | | | GO |

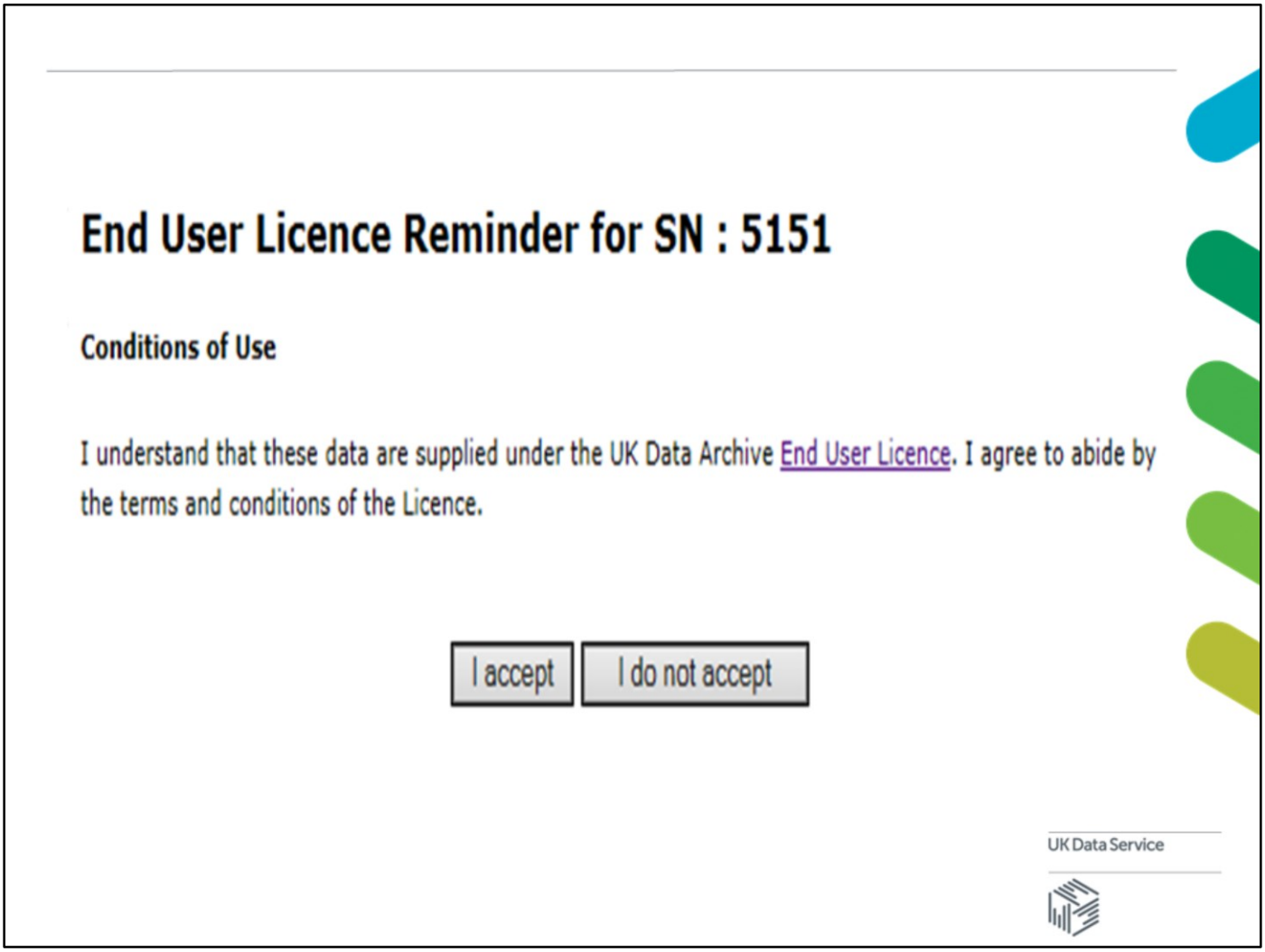
All requests are usually made available within 15 working days. Download requests are usually supplied in SPSS, STATA, TAB and RTF (qualitative only) formats and include documentation (in Adobe Acrobat PDF format). If data are required on CD or in another format, please use the 'Other media' option.

**Add a dataset to this usage - enter the study number (e.g. 1234)** GO

If you do not know the study number(s), find datasets using Search catalogue, Major studies and use the 'Download/Order' links to add the study to this usage.

**Order or view previous orders for datasets from a non-UK archive that do not have a study number in our online catalogue** GO

Details for this usage


End User Licence Reminder for SN : 5151
Conditions of Use
I understand that these data are supplied under the UK Data Archive End User Licence. I agree to abide by the terms and conditions of the Licence.
I accept
I do not accept
UK Data Service

You have selected to download the following dataset -
**SN: 5151 British Household Panel Survey: Waves 1-18, 1991-2009**

For Usage -
**Usage number: 93579**
**Usage title: UKDS Webinars**
**Intended use: NON-COMMERCIAL**

Under 'Format download' click on your chosen format button. **You must choose to 'Save this file to disk'.**

See **Download and format help** if you have problems downloading and for advice on choice of formats, file contents, and unzipping the compressed zip file.

| Format download | Filename | File size |
|---|---|---|
| SPSS | 5151SPSS_F864887C4E6472809B315AD4CBDA22FD.ZIP | 191.99 MB |
| STATA | 5151STATA8_51ECE7BDA43B49735C6EAAEEBC99B03F.ZIP | 174.31 MB |
| TAB | 5151TAB_F601B93B6EFBDA4B9D31A571C65970ED.ZIP | 213.34 MB |

UK Data Service

# Online analysis with Nesstar

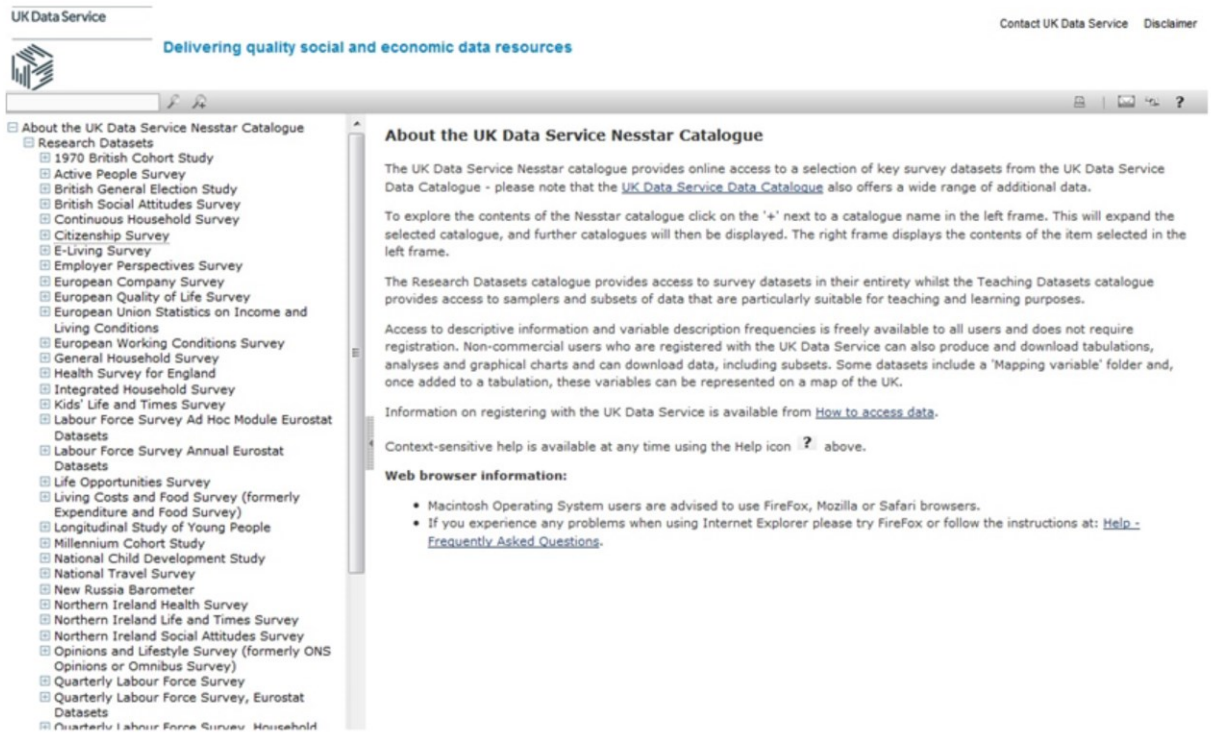
- Online data browsing and analysis system
- Allows users to search for, locate, browse and analyse and download a wide variety of statistical data within a web browser
- UK Data Archive, as service provider for the UK Data Service, hosts a Nesstar server populated by its most popular data series
- Registration is required for analysis such as cross-tabulations and regressions
- http://nesstar.esds.ac.uk/webview/index.jsp

You may have noticed that some of our datasets are available for online analysis using a system called Nesstar

In the previous example, I was downloading data in SPSS format, but I could have browsed the data using Nesstar – the good thing about Nesstar is that you use it in your web browser and so you don't need any special software

So at the bottom of the page, you can see a link which will take you to our nesstar system

# Using Nesstar




This is what the Nesstar interface looks like

On the left hand side you can see a whole series of data collections, which is increasing all the time.

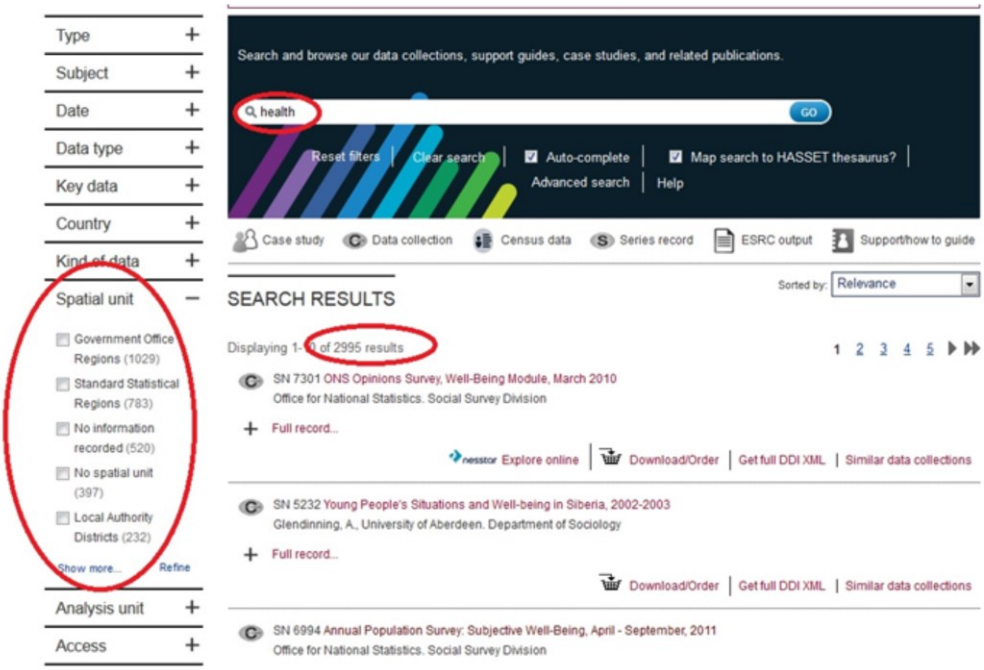
# Using Nesstar

UK Data Service

And by navigating through that list, you can find a particular survey and start to produce your own visualisations in your own web browser

Such as this….. This is an example taken from the Citizenship Survey 2010-11 & it's a graph showing how strongly people feel they belong to a neighbourhood or community

# Discover – conducting a search




http://discover.ukdataservice.ac.uk/

UK Data Service

Discover is our search engine.

Here's an example of a search I've run using health as my key word. You can see that it's returned a large number of results, so you can use the options on the left side to refine your search a little – this will make the search more targeted, and so will return fewer results. So you can see here that if I refined my search by selecting Government Office regions, this would reduce the number of results to 1029.

Discover is a really good tool for searching for data sources that you may not be familiar with already, so I would highly recommend that you take some time to explore this when conducting any research.

# Questions

Contact Details:

Veerle Van den Eynden

veerle@essex.ac.uk