

UK DATA SERVICE



ANNUAL REPORT

2016/17

The data collections held by the UK Data Service reflect the issues that affect all our lives and experiences, both close to home and internationally, covering every continent, from birth through education, employment, our social interaction and on to old age.



# CONTENTS

**07** Research

**19** Reach

**25** Infrastructure

**39** Skills building

**44** Future Plans

Welcome to the UK Data Service Annual Report which covers the period April 2016 to March 2017 and highlights the major achievements of the UK Data Service over the past twelve months. The UK Data Service is funded by the Economic and Social Research Council to meet the data needs of researchers, students and teachers from the academic, central and local government, charities and foundations, independent research centres, think tanks and the commercial sectors.

# UK Data Service Annual Report



We  
provide

- **Access** to high-quality local, regional, national and international social and economic data
- **Support** for policy-relevant research in the higher education, public and commercial sectors
- **Guidance** and training for the development of skills in data use
- The **development** of best practice data preservation and sharing standards
- The **sharing** of expertise with international data providers to remove barriers to accessing data

We curate  
data in order  
to ensure its  
long term use

## Supporting researchers as they create and share new data

The UK Data Service curates data to ensure its long-term use and encourages the reuse of data by researchers. We have been working with researchers to help them to consider how the data they create can be reused in future research and consider barriers to sharing their data.

### Challenging data deposits

To understand issues researchers face when working with challenging data we have undertaken interviews with key researchers who deposited data with us. Researchers shared their expertise on practical implications of sharing data from sensitive topics and undertaking anonymisation of their data. Involvement in research where time has lapsed between research and preparing data for deposit or where they needed to balance confidentiality with usefulness of data for other research. Although researchers we spoke to have worked with data that might seem difficult to share or reuse, all have overcome these challenges to create shareable and reusable data. We have collected their experiences in a series of 'depositor stories', available on our website as guidance for other researchers.

### Sharing research data

In September 2016 the UK Data Service, in collaboration with the London School of Hygiene and Tropical Medicine, was awarded a grant from the Wellcome Trust to investigate the current attitudes and actual practices of researchers to open science, barriers that inhibit or prevent sharing research data and code and measures research funders could take to encourage open science, data sharing and reuse. Towards Open Research gathered evidence from 583 Wellcome Trust-funded and 259 ESRC-funded researchers and found open research practices are increasing, with new collaborations, validation of research and increased citation rates being the most recognised benefits of sharing data. Barriers to sharing included lack of participant permission and concerns over sensitivity of data collected and concerns about financial and time cost to prepare data for reuse.

### Syntax upload facility

To help with reuse, the UK Data Service launched a Syntax Upload Facility, allowing researchers to share their syntax created for data in our collection. Available in all our catalogue records, files can be uploaded in various formats automatically creating a citation reference for the code and a file to download for reuse by others.



**I want to make sure that the data is as useful as it can possibly be because if it's not, then I'm doing a disservice to my respondents."**

Dr Karon Gush, Institute for Social and Economic Research, University of Essex

## Encounters with Big Data: South Africa Summer School

A team from the UK Data Service spent five days in Cape Town teaching and mentoring researchers and data scientists from across South Africa in a summer school which focused on applying big data methods to social science research. The summer school was part of the Smarter Household Energy Data Project, a collaborative research project between the UK Data Service and DataFirst in South Africa, which aims to create a shared research structure for large-scale household energy data.

The summer school covered aspects of extraction, exploration and statistical analysis of data using the UK Data Service's new Data Services as a Platform (DSaaP) initiative. The legal and ethical issues of using big data were considered and case studies on health and demographic survey data were developed. Delegates attending the summer school were highly experienced data users and creators

who worked with a wide range of data, from longitudinal surveys, public health and medical data to financial data and satellite data.

After three days of learning and discussion, summer school participants moved on to using their newly acquired skills to formulate their own group projects. Focusing on accessing structured open source data from the web, the groups imported, queried and linked data for exploratory analysis, modelling and mapping. Experimenting with big data tools gave delegates the practical experience that they can apply in their own research. The summer school was also an opportunity for data specialists to connect with other researchers from across South Africa and from the UK. The UK Data Service looks forward to continuing this collaborative work with the researchers who attended.

“

Our aim was to support researchers in understanding and analysing large and complex datasets, focusing on leveraging the power of popular statistical software within a big data environment.”

Nathan Cunningham, Director, Big Data Support Network

## Our Data Impact Fellows

In 2016 the UK Data Service welcomed our first Data Impact Fellows. Seven Fellows were selected following an open competition seeking early career or post-doctoral researchers, who use data from the UK Data Service in their work with a focus on impact. The Data Impact Fellows programme is designed to support the use of Service data and resources by new scholars through the research partnerships they develop and students they teach throughout their careers, whether they remain in academia or branch into the wider public or commercial sector.

The two-year programme aims to provide career development opportunities for researchers at a relatively early stage in their academic careers. The Impact Fellows meet to share expertise and to receive support and advice from Service staff. Fellows also receive financial support to cover impactful public engagement activities for their research.

Working in a diverse range of fields we have charted their progress on our Data Impact Blog.

**Dharmi Kapadia:** is a Q-Step Lecturer in Sociology, University of Manchester with a PhD in Social Statistics and 10 years' experience of research to improve lives of people suffering from racist discrimination and mental and physical illnesses. Dharmi is currently looking at how racism influences mental illness in ethnic minority people using data from Understanding Society.

**Sarah Knight:** is an ESRC-funded PhD student, University of York. She has an interest in human well-being, nature connectedness, environmental quality and ecosystem services. Sarah is researching impact of natural environment on subjective well-being using Understanding Society and the British Household Panel Survey to explore the determinants of self-reported well-being on adults in the UK.

**Rachel Oldroyd:** is a quantitative human geographer, University of Leeds. Rachel started a part-time ESRC White Rose Studentship in October 2015 and her PhD is centred on spatial data analytics for food safety, with particular focus on foodborne illness.

**Marii Paskov:** is a sociologist and Research Officer, University of Oxford. Using the European Social Survey, the Wealth and Assets Survey, and the British Household Panel Survey Marii's research focuses on intergenerational social mobility.

**Bram Vanhoutte:** is a Research Fellow, University of Manchester. His research uses the English Longitudinal Study of Ageing, as well as the US Health and Retirement Study, and the Australian Life Histories and Health Study, and focuses on wellbeing in later life as a marker for successful ageing.

**William Shankley:** is a sociology PhD student, University of Manchester. His research uses census microdata and census flow data and focuses on internal migration patterns of Polish people in England and Wales and the factors that shape their migration decision-making.

**Claudia Zucca:** is a Marie Curie Early Stage Researcher working on the VOTEADVICE project, in collaboration with the University of Exeter, Kieskompas Amsterdam and Koç University Istanbul. Her research uses the British Social Attitudes Survey, the European Social Survey, the British Election Study and the European Values Study, and focuses on the impact of new technologies on political behaviour using advanced quantitative methods.

Encouraging  
new and  
emerging  
researchers





## #DataImpact2016

### #DataImpact2016

#DataImpact2016 took place on the evening of 13 October 2016 at the Lighthouse, Glasgow, with the theme Identifying data solvable challenges. The event showcased the impact of the Service, focusing on stakeholder groups from the higher education, government, commercial and civil society sectors in Scotland. A panel of leading data innovators explored data re-use in policy and research, sharing their experiences of demonstrating data enhanced impact.

- Roger Halliday (Chief Statistician at the Scottish Government)
- Rosemary Agnew, (Scottish Information Commissioner)
- Gillian Docherty, (Chief Executive Officer of the Data Lab)
- Des McNulty (Deputy Director of Policy Scotland)
- Bill Roberts (Founder of Swirrl) and
- Piyushimita Thakuriah (Director of the Urban Big Data Centre).

Our panel members considered data re-use in policy and research, sharing their experiences of demonstrating data enhanced impact.

### Open Data Dives

Two Open Data Dives brought together coders, data enthusiasts, graphic designers, project managers and pizza lovers to hunt down, crunch and beautify open social and economic data. Hosted at The Shed digital innovation hub at Manchester Metropolitan University, the first Data Dive led to the creation of 'London Spike' – a 3D visual representation of median local authority house prices in the UK. Based on boundary data from the Census, combined with land registry data, the model clearly demonstrated the inflated house prices affecting the capital.

Our second Open Data Dive took a challenge theme: On 5 May 2017 Greater Manchester will have a new Mayor: Your gift to her/him - a SWOT analysis of Greater Manchester and the Region. Attendance by local and national government, HE and commercial sector organisations, included a map and an app to look at the number of prescriptions per 100 patients, GP surgery registered patients by region and prescriptions for allergies, asthma, diabetes and an app showing how income affects life expectancy.

### Impact and Innovation Lab

We developed the Impact and Innovation Lab to compliment the Data Impact Blog. The Lab is designed to demonstrate impact and innovation through working more closely with tech entrepreneurs. It aims to develop inspiring digital solutions to social challenges with new and existing data users in a range of sectors, wherever they are found.



**Your data will never be perfect. Just start where you are."**

Gillian Docherty, Data Lab

# Enabling researchers to use data in a responsible and secure way

## Secure Lab: Promoting the responsible use of research data

Secure Lab is the ESRC funded service established to give qualified researchers safe, secure and free access to sensitive or confidential business, social and economic microdata. All researchers wishing to access data via Secure Lab have to take Safe User of Research data Environments (SURE) training which gives users a solid grounding in the processes for accessing sensitive data securely, the legal responsibilities for users of secure data and practical details of how the Secure Lab works. Use of the Secure Lab has increased steadily since it was first opened in 2011 and there are now over 550 users trained and registered to use the service. The SURE Consortium, comprising representatives of the UK Data Service, ADRN, HMRC and ONS, meets regularly to use their expert knowledge of legal and procedural obligations, statistical disclosure control, and training provision to continually monitor, and update where necessary, the content and delivery of the SURE training course.

Use of the Secure Lab is helping to support impactful research. Secure Lab data was used by the Institute of Fiscal Studies for their working paper '*The short run elasticity of National Health Service nurses' labour supply in Great Britain*' which used the Annual Survey of Hours and Earnings dataset to examine how differential labour market conditions across localities impact on the elasticity of supply of NHS nurses. Findings determined an increase in pay for NHS nurses in localised areas, as opposed to a national level, in combination with improved working conditions and area supplements, would benefit the employment and retention of staff.

Secure Lab was also used for analysing data for a report produced highlighting hotspots of creative activity in all nations and regions of the UK. Using data from the Business Structure Database the research found that creative industries are a high growth sector making economic contributions with a Gross Value Added of £84.1 billion and accounting for 5.2 % of the whole UK economy.

### Developing best practice

Staff from the UK Data Service continue to be part of the working group for Safe Data Access Professionals, alongside staff from ADRN, The Health Foundation, HMRC, CRUK and The Jill Dando Institute. The group is designed to share and promote best practice in relation to safe data access, including developing, agreeing and maintaining relevant standards, and aims to bring further professionalisation to the sphere.

We have also worked with the Office for National Statistics and staff at the University of the West of England to start developing a new training course, designed to meet the needs of the Digital Economy Act. Members of the Access team at the UK Data Service are part of the Expert Advisory Group that feeds their expertise into the development and content of this course.



The report used a number of official data sources available from the UK Data Service, combined with geographical analysis of data on the supply of university graduates and research, and analysis of social networking activity."

## Creative clusters in the UK

Creative industries are a high growth sector and make important economic contributions with a Gross Value Added (GVA) of £84.1 billion, accounting for 5.2 % of the whole UK economy. Although it was well known that creative industries are engines of growth in urban areas, importance in other parts of the UK was less widely documented.

Nesta and Creative England produced a report highlighting hotspots of creative activity for all regions of the UK, providing data that economic development agencies and policymakers in charge of promoting business growth could use when setting strategic priorities.

### Key findings

The research demonstrated that the creative industries are growing more rapidly than other sectors in most parts of the country particularly in services activities like design, digital media and advertising. Findings show creative industries are gaining in economic importance across the UK, but more could be done to strengthen the UK's creative clusters. At the local level, governments could take action to boost creative industries growth and incorporate it into their strategy whilst on a national level could play a more active role to scale up creative clusters outside London and the South East. Relative scarcity of talent, relevant research and

knowledge exchange, as well as low levels of local networking were issues raised that could be developed to support creative clusters.

### How the UK Data Service was used

The report used a number of official data sources available from the UK Data Service, combined with geographical analysis of data on the supply of university graduates and research, and analysis of social networking activity. The study measured business activity using Business Structure Database, which contains location, employment and turnover information about most registered businesses in the UK, complemented with employment figures

from the Annual Population Survey. GVA figures were estimated using detailed location and financial data from the Annual Business Survey. Analysis was conducted on geographies at the level of the Travel-to-Work-Areas; ONS geographies bounding local labour markets based on commuting patterns from the UK 2011 Census.

### Use for policy

Used by the Arts Council England as evidence in the The arts and culture sector and exit from the European Union report focusing on how the UK's exit from the EU can deliver the best outcome for arts and culture.

# Setting the standards internationally

## International engagement

### Indian Council for Social Science Research

In June 2016 the Indian Council for Social Science Research launched their new national data service, an endeavour which has been realised with the support of the UK Data Service. Staff from the Service helped to develop a road map and implementation plan for ICSSR's data service, outlining the proposed information architecture for the data repository, setting out key components of data acquisition, data pre-processing and analytics, metadata, software requirements and technology. Data experts from the Service were also involved in in-depth training of ICSSR staff. The new ICSSR data service is based on the model of the UK Data Archive, lead organisation of the UK Data Service.

There is an abundance of data of interest to the social scientist generated across India including the Census, as well as sector specific data such as the Annual Survey of Industries (ASI), the principal source of industrial statistics in India. As in many countries, academics also generate their own data through funded research projects. Making data available

for research is an ongoing challenge in India, with open access an exception; government departments cautious to grant access to data; and private sector data, wherever accessible, expensive to make available for research. The launch of the new data service has made social science data more available and progress has been impressive, with 33 surveys being published to the survey data catalogue in the first month of operation.

### OECD Global Science Forum Expert Group

The OECD Global Science Forum's Expert Group on Research Ethics and New Forms of Data for Social and Economic Research's report on the same topic was soft-launched in April in Chicago. A response to the growing challenges of using new and emerging forms of data, these guidelines are the culmination of a two-year project to provide a set of high-level recommendations that can be used to underpin the ethical governance of research. Professor Matthew Woollard, Director of the UK Data Service, represented the international social science data archive community on this Expert Group.



**New forms of data have the potential to revolutionise research in the social sciences, but come with significant responsibilities for data producers, researchers and archives."**

Matthew Woollard, Director, UK Data Service



## Administrative Data Research Network

The Administrative Data Research Network (ADRN) is a UK-wide partnership between academia, government departments and agencies, national statistical authorities, the voluntary, community and social enterprise sector, funders and the wider research community to facilitate new economic and social research based on routinely collected government administrative data.

In the three years since its foundation, the ADRN has established new legal, secure and efficient pathways for the research community to access functionally anonymised, linked administrative datasets. These datasets can provide detailed information about large sections of society, which researchers can analyse to provide a greater evidence base to inform policy. In 2016 the ADRN's work on infrastructure, data linkage and privacy protection processes came to fruition, with many exciting research projects underway, some publishing results, and many more in the pipeline. The ADRN also helped researchers and institutions to learn and harness the skills that are needed in the new digital economy.

### Sharing expertise

The ADRN received a delegation with experts from the South Korean Government in September 2016. They travelled to the UK specifically to visit the ADRN and find out how the UK uses government data

for research. They were eager to understand the processes and procedures behind the ADRN and very interested in the challenges the ADRN has overcome. Statistics Korea, the country's equivalent of the UK's Office for National Statistics, conducted their 2015 population and housing census using government data, rather than a census. The visit allowed them to hear about practical aspects of handling such data, and about negotiating with data owners, working with researchers and engaging the public to encourage understanding and support for this kind of research.

### Digital Economy Act, 2017

ADS Co-Investigator, Matthew Woollard was one of the experts who provided guidance and expertise during the Cabinet Office consultation process on the Digital Economy Bill. The Cabinet Office brought together government departments, researchers, privacy groups and others for this consultation. Throughout the consultation, it became clear that the policies, procedures and infrastructure the ADRN has put in place could be a model for what can be done nationally through the legislation. When the Bill was published in July 2016, it had adopted the ADRN model as best practice for accessing and linking administrative data for research. The Bill received Royal Assent at the end of April 2017, at which time it passed into legislation as the Digital Economy Act.



# UK service provider for CESSDA

## CESSDA

The UK Data Service has worked with the Consortium of European Social Science Data Archives (CESSDA) on a number of new initiatives to strengthen and widen the social science data research infrastructure. The UK Data Service has teamed up with other European social science data archives to develop several cutting-edge tools: Product and Services Catalogue (PaSC).

A one-stop-shop to find datasets for social science research, the PaSC's automated collection of metadata keeps content up-to-date and directs users to the right source/service to retrieve any data of interest.

### Euro Question Bank

This centralised search facility across all of CESSDA's survey holdings is a step towards a single point of access for survey questions across Europe. Researchers can find survey questions on different datasets in different languages and information on associated studies, datasets and variables. Tools include the ability to search for a specific concept and switching languages to view translations of the same question.

### Metadata Management

This project aims to develop, promote and implement a standardised metadata design, content and practice for all CESSDA data assets in the PaSC, which is fully compliant with the Data Documentation Initiative (DDI) international standard for describing statistical and social science data.

### Controlled Vocabulary Manager

CV Manager enables controlled vocabularies (CVs) to work more quickly, be less labour-intensive and less error-prone. It allows for the creation, versioning and maintenance of controlled vocabularies, their translation into all member languages, and access to all CVs, helping to retrieve information in the CESSDA Product and Services Catalogue and Euro Question Bank.

### Data Discovery Training

A pan-European training programme offering workshops for researchers from across Europe, the data discovery training sessions focus on finding data from CESSDA Service Providers, and promote the understanding of complicated datasets and their potential.

In common with other CESSDA members, we have also worked on a technical framework which will guide all members when developing software and services that are part of the CESSDA Research Infrastructure, making collaboration and cooperation easier.

# Supporting data infrastructure development

## Big Data Network Support

Established by the ESRC in 2014, the Big Data Network Support (BDNS) aims to optimise the use of new and emerging forms of data for research. The enormous volume and complexity of the data has the potential to hugely benefit academic research, organisations and society, but the technical, practical and ethical challenges of utilising big data require specialist understanding and support. The BDNS provides support for research centres using new and emerging forms of data, alongside technical innovation for the UK Data Service to enable researchers to access and analyse big data.

In the last twelve months the BDNS has delivered a successful pilot of the Data Services as a Platform (DSaaP) technical infrastructure, preparing the UK Data Service to support new and emerging forms of data, alongside supporting the Big Data Research Centres (DRCs) around the UK. Together the BDNS and the DRCs have provided training and webinars on big data for researchers and data users, collaborated on work plans for the ingest, curation and access of data and considered issues around data acquisition and licensing.

The BDNS team at the UK Data Service has been strengthened through the inclusion of commercial partnerships with Hortonworks, Cloudwick and Amazon Web Services who provided specialist expertise at key points of the project. Significant staff training and development has also taken place this year to fully utilise the specialisms within the team and to share knowledge and best practice with the DRCs.

### Data Services as a Platform (DSaaP)

This project has been developed as a partnership with Hortonworks, a leader in Hadoop-based data technologies. The DSaaP uses solutions that deliver data at scale, with speed and security, including Hive, Spark, and Zeppelin, which integrate seamlessly with popular data analysis environments like R and Python. DSaaP is a flexible and cost-effective infrastructure that links data on physical servers and cloud-based data storage. This hybrid model enables tight security but also provides users with a seamless way of accessing the data they need.



The BDNS helps researchers make the most of big data for knowledge exchange and impact. In order to present data in a way that is useful to researchers, we needed the right framework.”

Nathan Cunningham, Director Big Data, UK Data Service



## Building Research Capacity

To prove the robustness and usability of the Data Services as a Platform infrastructure (DSaaP), Big Data Network Support (BDNS) developed a set of case studies which use new and novel forms of data in DSaaP, demonstrating to partners and data users the benefits of using the platform for research. The case studies focus on key big data technologies, such as Hadoop and data lakes.

The RCUK Centre for Energy Epidemiology (CEE) at UCL Energy Institute is the UK's first dedicated energy epidemiology research centre. They used combined data from various sources, including energy, geospatial and weather data, to develop statistical models to illustrate underlying drivers

of energy demand in UK dwellings. Accurate and cost-effective characterisation of energy demand in dwellings is required for effective energy policy. At present, Energy Performance Certificates (EPCs) are used to assess energy demand in UK dwellings but require on-site visits. This project aimed to use big data methods and smart meter data to generate empirical models of dwellings based on their energy consumption profiles and associated weather data.

### Research challenges

The introduction of smart meters has provided high quality, high-resolution electricity and gas consumption data across thousands of dwellings in

the UK, offering enormous potential but cannot be analysed using traditional analytical methods. CEE's legacy data processing frameworks were struggling to cope with the volume and computationally intensive analytics. This made exploratory analysis significantly slow and time consuming for researchers detracting from addressing important research questions. The project was selected as a pilot project using the UK Data Service's new Hadoop-based initiative Data Services as a Platform (DSaaP). Domestic energy consumption data is considered personal data under the Data Protection Act; data confidentiality and data security capabilities of DSaaP were critical elements for the feasibility of the project.

### How the BDNS helped

Sensitive datasets were safely stored by the UK Data Service and DSaaP's dataflow management platform.

The BDNS scaled out the CEE's data processing procedures to over 8,000 dwellings equipped with smart meters in the UK. DSaaP provided an integrated, safe environment for the interactive exploration and analysis of smart meter and weather data at scale, which enabled researchers at the UCL Energy Institute to build new and innovative models for the accurate thermal characterisation of UK dwellings.

“

The use of the UKDS Data Services as a Platform made a real difference to this project. The increased performance and scalability enabled UCL researchers to refine and test computationally intensive models at scale, while the suite of tools facilitated interactive exploratory analysis in much shortened timeframes. In effect, near real-time analysis on billions of data points is now possible while operating in a trusted, secure environment”.

*Professor Tadj Oreszczyn, the Director of CEE and UCL's Bartlett School of Environment Energy & Resources*



# Making ground-breaking data openly available via UKDS.Stat

UKDS.Stat provides many data usage tools, including integrated metadata, searching across datasets from different providers (IMF, OECD, International Energy Agency, World Bank, UNIDO and UNESCO), data visualisation options and the ability to save and share queries. To help users UKDS.Stat has been updated, alongside the API guides.

### Human Rights Atlas 2012

The *Human Rights Atlas 2012* dataset was originally created by the Institute for Democracy and Conflict Resolution at the University of Essex and Mackman, for the ESRC-funded Human Rights Atlas Project; the main objective of this work was to produce a visually pleasing, interactive interface for the dissemination of global human rights data. The project team used published data from the World Bank, the United Nations, academics, non-governmental organisations and the other bodies to generate a dataset which includes more than 240 different measures of economic, social, political, and legal life for over 200 different countries across the world and gives a picture of the lives and rights of human beings over a 30-year period from 1981 onwards.

The Human Rights Atlas data was archived and made available at the UK Data Service back in March 2015 via the Service's Discover Catalogue. Making the data available via UKDS.Stat allowed users to explore, visualise and subset the Human Rights Atlas data prior to download. The data can be extracted in Excel, PC-Axis, SDMX(XML), CSV formats and accessed via a Developer RESTful API (application programming interface).

### UNESCO datasets

These datasets are a valuable addition to the UK Data Service international open data portfolio and have come about as a result of membership of the Statistical Information Systems Collaboration Community (SIS-CC). Set up by the Organisation for Economic Co-operation and Development, the SIS-CC is a group of international institutions working together to share experiences and expertise and to enable cost effective innovation in large-scale data dissemination. The UK Data Service has been working collaboratively with SIS-CC to develop UKDS.Stat since 2012.



The web interface is very user-friendly, and the resource has been used for educational and research programmes, including a MOOC on combating contemporary slavery delivered by the University of Nottingham to over 6000 participants around the world. It is great that the underlying dataset is available in this new format."

Todd Landman, Professor of Political Science, University of Nottingham

# Collecting, processing and releasing data for research

## Responding to changes in the national data landscape

Introduction of new data legislation in the Digital Economy Bill changed data sharing across government departments, improving the possibilities of linking government administrative data for research purposes, whilst protecting confidentiality. Prior to the Bill, the Cabinet Office undertook a consultation on data sharing, in which Professor Matthew Woollard, Director of the UK Data Service, took a key role.

### Concordat on Open Research Data

To help ensure that research data created by the UK research community is openly available for use by others and to capture best practice in the research community Matthew Woollard was part of the Concordat Working Group. The Concordat's key principles include open access to research data as an enabler of high quality research, right of creators of research data to reasonable first use, good data management as fundamental to the research process and the use of others' data should always conform to legal, ethical and regulatory frameworks.

### Special Licence Review

Teams across the UK Data Service worked with the Office for National Statistics to produce a detailed review of data collections hosted by the Service available via the ONS Special Licence scheme. This allowed Approved Researchers access to potentially disclosive datasets with more detailed variables than Safeguarded data, made available under End User Licence, but less detailed than Secure Access data. ONS wished to ensure their detailed data were only used in a secure environment, such as the UK Data Service Secure Lab or Virtual Microdata Laboratory.

The UK Data Service and ONS carried out a statistical disclosure review of the data, assessing each dataset for risk and determining release under End Licence or Secure Access. From the end of November 2016 the Service stopped providing access to ONS Special Licence data via previous access mechanisms and researchers were given the option to access this data via Secure Lab or the Virtual Microdata Laboratory. Researchers actively working with ONS Special Licence data were contacted by the Service prior to the switch to the new system, allowing a smooth transfer to the new system and minimal disruption to research projects.



**This Concordat represents a unified view across the UK research data related stakeholders about access conditions."**

Matthew Woollard, Director, UK Data Service



The study brought together a variety of datasets to analyse the impact of relative energy prices on trade. The coverage and detailed disaggregation of the data went well beyond previous work, allowing the first global ex-post analysis of the relationship between trade and energy prices.”

## Energy prices and international trade

### Do higher energy prices affect international trade?

Dr Misato Sato and Dr Antoine Dechezleprêtre, London School of Economics and Political Science, studied climate change policy and its effects on trade. Emissions trading policies are regulations implemented by many countries and cities to cost effectively reduce industrial greenhouse gas emissions. However, according to the European Commission on Climate Action they can result in carbon leakage and may affect the competitiveness of businesses. This research explored whether and to what degree changes in relative energy prices might influence trade and competitiveness.

### Key findings

Findings showed changes in relative energy prices have a statistically significant, but very small impact on imports. On average, a 10% increase in the energy price difference between two country-sectors increases imports by 0.2%. The impact is larger for energy-intensive sectors, but even within these the effect is minor — changes in energy price differences across time explain less than 0.01% of the variation in trade flows.

The authors calculated that a 30% increase in energy prices across Europe would cause exports to fall by only 0.5% and would increase imports by 0.07%.

### How the UK Data Service was used

The study brought together a variety of datasets to analyse the impact of relative energy prices on trade. The coverage and detailed disaggregation of the data went well beyond previous work, allowing the first global ex-post analysis of the relationship between trade and energy prices. The researchers created a unique and comprehensive dataset of industrial energy price indices at the country and sector levels covering 48 countries and 12 industry sectors for 1996 to 2011. The dataset used data from International Energy Agency World Energy Balances, International Energy Agency Energy Prices and Taxes and the World Bank. Additionally, the researchers

used data on GDP and population obtained from the International Monetary Fund’s World Economic Outlook (2012) and data on wages from the United Nations Industrial Development Organisation (2011).

### Use for policy

This research found unique evidence suggesting concerns about risks of carbon leakage may have been overplayed. The findings have been included as research evidence in the ‘Evaluation of the EU ETS Directive’ report carried out by the European Commission and subsequently informed policy measures implemented by the Commission regarding the revision of the EU ETS Directive.



## Census flow data: New functionality

Census flow data relate to the movement of people between places and differ from other census data as they link two locations — an origin and a destination.

Over the last year we have developed the functionality of our WICID platform for accessing census flow data and have added additional data to complement the census flow data, this time focused on local authority cross-border flows of state funded pupils in secondary, primary and special schools in England. We have also added census based flow data – published as commissioned outputs.

To aid users' discovery of flow data tables we have worked in collaboration with Office for National Statistics on the development of a table finder facility, so users can filter tables by keyword and geographical level, as well as searching table titles.

We've developed a new integrated mapping component with flow data in mind, so users can select origins and destinations of interest. The tool allows both origin and destination zones to be selected in a number of ways, whether one area at a time, through automatic selection of areas falling into a defined region, or falling into a specified buffer region.

The new mapping system developed for WICID has also been used in a collaborative project with The National Archives on mapping intellectual networks during the Enlightenment.

**William Shankley**, UK Data Service Data Impact Fellow and Sociology PhD student at the Cathie Marsh Institute of Social Research, The University of Manchester uses census microdata and census flow data and focuses on internal migration patterns of Polish people in England and Wales. Focusing on the factors that shape their migration decision-making, William's research combines 2011 UK Census data for England and Wales with interview data collected from forty Polish people living across Greater Manchester.

Mapping the Enlightenment Project:  
[mapping-the-enlightenment.org/](http://mapping-the-enlightenment.org/)

WICID Census Flow Data:  
[wicid.ukdataservice.ac.uk/cider/wicid/query.php](http://wicid.ukdataservice.ac.uk/cider/wicid/query.php)



Research involved face-to-face visits with young carers and their families and combined the data collected with demographic data of young carers from the 2011 UK Census."

## The lives of young carers in England

According to the 2011 Census, there are almost 166,000 young carers in England who provide regular care and emotional support to a family member. With the implementation of the Care Act 2014 and the Children and Families Act 2014 local authorities needed to assess the support needs of young carers. To evaluate the effectiveness of this new legislation, the Department for Education commissioned a qualitative report 'The lives of young carers in England'.

The research was conducted by TNS BMRB, in partnership with Professor Jo Aldridge, Director of the Young Carers Research Group (YCRG) at Loughborough University.

### Key findings

The research found caring was seen to be very rewarding by the majority of the young carers but also had adverse effects such as anxiety, tiredness, strained family relationships, restrictions in social activities and under-engagement in education. Parents often struggled to discuss their health conditions with their children or professionals for fear of potential repercussions for their family. Formal or informal support helped reduce young carers' responsibilities. The findings highlighted the need for whole family approaches when working with young carers and their families, as well as professionals to consider young carers' needs, as carers and as children. Early

intervention should ensure access to appropriate information about the parent's health condition and support available. The report recommended children should not have to become carers and the need for good universal services to prevent this. When young people become carers, health, social care and education professionals should work to build resilience in young people and reduce anxiety.

### How the UK Data Service was used

Research involved face-to-face visits with young carers and their families and combined the data collected with demographic data of young carers from the 2011 UK Census. They used InFuse, a UK

Data Service tool providing easy access to aggregate data from the UK 2011 and 2001 censuses, to find the number of young carers in England and hours of unpaid care they provide per week.

### Use for policy

This qualitative study helped provide current context of caring among children and young people at a key point in time; the introduction of the new legislative framework (the Care Act 2014 and the Children and Families Act 2014). Shortly after publication, the Department of Health ran a consultation seeking views on a new strategy for carers and how to support them. The Department of Health is analysing feedback.

# Promoting good data management practices

## Collaborative training

We held two user conferences in collaboration with Office for National Statistics (ONS) - the Labour Force Survey and Annual Populations survey user Conference and the Crime Surveys User Conference. As well as showcasing research and developments with these data, the events provided ONS and users opportunity to discuss and gather feedback on the ONS Special Licence review.

Also, in collaboration with the ONS Annual Population Survey well-being data team, we organised our first personal well-being user event in London to showcase well-being data from ONS surveys and to highlight research that uses these data.

International collaborative events have also taken place. A joint webinar was held with the Gateway to Global Ageing Data Finding, which introduced cross-country analysis in Stata using harmonised health and retirement datasets. Another joint webinar was run with Eurostat on *Understanding Eurostat Data*. Both webinars received very positive feedback. In addition, one invited workshop was held at Academia Sinica at their Survey Data Archive on preparing qualitative data, two invited

workshops were held in India - one on Research Data Management at the 2017 International Conference on Changing Landscapes of Science and Technology Libraries and another on *Preparing Qualitative Data for Sharing and Reuse* for the new Indian Social Science Data Archive, and another on transparency in qualitative research at the Nottingham Conference on The Gold Standard of Reproducible Research: How to make it work in qualitative and quantitative research. Finally, a workshop was given to staff from the Digital Repository of Ireland (DRI) to learn more about Nesstar and survey archiving practices.

### Webinars

Our knowledge and expertise in delivering training webinars has become increasingly recognised, with the Service being approached to provide guidance and support to organisations starting their own webinar programmes. For example, in the last year we have supported CESSDA, the National Centre for Research Methods and other organisations/services.





## Using data in the classroom

The User Support team has been working with teachers across the UK to help social sciences educators develop their student's data skills. The first workshop aimed specifically at teachers who lead courses teaching undergraduate and postgraduate quantitative analysis skills was held in September 2016 and looked at ways of encouraging students to choose their own data, the skills that social science researchers need and where to find resources for teaching with data. The workshop was led by Service staff with contributions from the directors of two Q-Step centres. Group discussion benefitted from the mix of participants, which included both established lecturers and early-career researchers.

### Teaching resource

In response to a consultation with teachers, we have created a resource that provides a list of our teaching datasets with useful summary information for each dataset such as the type and number of variables, sample size, substantive topics and potential uses. We have also provided a way for teachers to administer the UK Data Service Access Agreement for teaching in Virtual Learning Environments, such as Blackboard or Moodle. This includes a short Data Access Test that can be imported to a course site and when successfully completed allows students to then access the data.

### Undergraduate Dissertation Prize

The Dissertation Prize is an opportunity to recognise and reward undergraduate students who use data from the Service in their dissertations. Over 30 dissertations were entered into the competition with first prize being awarded to Victoria Smith from the University of Manchester for her dissertation 'An Exploration in to the Factors Shaping Victim Reporting of Partner Abuse to the Police'. Using Scottish Crime and Justice Survey data, Victoria explored victim reporting of partner abuse to the police presenting bivariate and logistic regression analyses of the factors involved.

### Quiz App

The UKDS Quiz App, launched last summer has been populated with additional country quizzes and new quizzes have been developed based on social facts in context, to coincide with various public awareness days like mental health data awareness day and vegetarian week. Future plans for the App include creating quizzes linked to the citizenship curriculum for 11-15 year olds.



The material was very well-pitched for students developing their undergraduate dissertation proposals. The session really opened the students' eyes to the wealth of 'real life' data they can explore in their research projects."

Dr Rachel Bennett, University of Gloucester



This House expresses concern at the findings of the recent report by Independent Age, entitled *The overlooked over-75s*, that an estimated 20 per cent of older pensioners are living in poverty; and urges the Government to take immediate remedial action to reduce this figure so that those who have contributed to this country their entire lives can enjoy their retirement in dignity.”

*The House of Commons, Early day motion 133*

## The overlooked over-75s

In April 2016 the UK government introduced a new State Pension for people who reach State Pension age on or after 6 April 2016. Following the introduction, Independent Age reviewed financial circumstances of current pensioners who will not benefit from the new system.

In collaboration with Dr Matt Barnes of City University London, Independent Age carried out new analysis to better understand the differences in the financial circumstances of older people, particularly those aged 75+. They analysed income data and looked at differences in wealth between groups of older people

### Key findings

The research found that older pensioners' incomes are on average £59 a week lower than younger pensioners and £112 a week lower than working age adults. One fifth of those aged 75 and over are living below the poverty line, with women, single people and those in rented accommodation particularly likely to be financially overstretched. Over-75s are also twice as likely as under-75s to have been in poverty persistently for the last four years and are unlikely to see significant increases in their income during their lifetime. The Findings also show older pensioners are less likely to claim Pension Credit, designed to boost income of the poorest pensioners – an estimated three quarters of

a million over-75s are entitled but are failing to claim, often unaware they qualify. The report recommends the government should re-energise efforts to promote the take up of Pension Credit and other benefits to the groups of older people most at risk, introduce a 'triple lock' on Pension Credit to guarantee recipients of the 'old' State Pension do not suffer a relative decline in their state income and ensure lower income pensioners continue to receive vital universal benefits like the Winter Fuel Payment and the free bus pass.

### How the UK Data Service was used

Data from the Family Resources Survey (2013/2014) and Understanding Society (2009-2014) were used

to analyse incomes and circumstances of private households in the UK. The official poverty estimates used in the report were published by the Department for Work and Pensions and are presented in the Households Below Average Income (HBAI) series.

### Use for policy

The House of Commons submitted an Early day motion 133 in June 2016 to debate the findings of the report: The report has also been used by the Finance Foundation as evidence in their report on the issues faced by elderly pensioners in managing routine financial matters.

## Data Infrastructure Strategic Advisory Committee

- ◊ Gordon Marshall, Chair, Leverhulme Trust
- ◊ Simon Briscoe, Vice-Chair
- ◊ Fiona Armstrong, Economic and Social Research Council
- ◊ Roger Burrows, Goldsmiths, University of London
- ◊ Roma Chappell, Office for National Statistics
- ◊ Chirdeep Chhabra, Digital Catapult
- ◊ Gerald Donnelly, National Records of Scotland
- ◊ David Marshall, Northern Ireland Statistics and Research Agency
- ◊ Suzy Moat, University of Warwick
- ◊ David Pye, Local Government Association
- ◊ Sharon Witherspoon

## Future Plans

### Data Skills Modules

Data specialists from the User Support and Training team are working with an award winning e-learning technologist to produce a set of self-learning 'data skills modules', using Articulate storyline software. Three modules introduce key data types (surveys, aggregate data and longitudinal). Each module describes key features of each data type and how the data are used, set tasks for learners and provide feedback to learners through the use of online quizzes.

### Improving the user experience

The Access and User Support teams at the Service will be working with a user experience (UX) developer to improve the processes of accessing data and support.

### Impact Ambassadors

The UK Data Service has an objective to increase collaboration between the research community and non-academic organisations and communities; to become a more active 'data facilitator' identifying, responding to and leading emerging trends in data driven research impact. We plan to continue to support public, commercial, voluntary and community sector representatives and data users from among the academic community jointly to understand and support their data needs, and identify opportunities and challenges to coordinate a programme of activity focused on deriving impact through identifying and addressing 'data solvable' societal challenges defined by partners.

## About the UK Data Service

The UK Data Service is delivered by staff with specialist expertise in research data based at higher education institutions across the UK:

### The UK Data Archive, University of Essex

Lead organisation of the UK Data Service and an internationally acknowledged centre of expertise in acquiring, curating and providing access to data. The Archive has been providing coordinated data service infrastructure to the ESRC for 50 years.

### Jisc

A registered charity which champions the use of digital technologies in UK education and research. Jisc hosts a significant number of the UK's research information assets and builds applications to help people make the most of these resources.

### Cathie Marsh Institute for Social Research (CMIST), University of Manchester

A centre of research excellence specialising in the application of advanced quantitative methods in an interdisciplinary social science context. CMIST plays a key role in the support of microdata from UK Censuses and Government Surveys.

### School of Geography, University of Leeds

A long-standing association with both the spatial analysis of census data and the development of web-based systems to ensure their use in the wider academic community

### Geography and the Environment, University of Southampton

An outstanding reputation for leadership in population and health research, based on the combination of innovation in methodology covering geographical information systems (GIS), spatial analysis and quantitative and qualitative methods with substantive expertise in census, population and health issues.

### EDINA, University of Edinburgh

The ESRC's specialist geography unit for the UK census, supporting UK Data Service users in accessing and using the geography outputs for the census.

### Department of Information Studies, University College London

An international centre for knowledge creation and transfer in the fields of librarianship, archives and records management, publishing, information science and digital humanities.

### Centre for Advanced Spatial Analysis, University College London

One of the leading forces in the science of cities, generating new knowledge and insights for use in city planning, policy and design and drawing on the latest geospatial methods and ideas in computer-based visualisation and modelling.

