

UK Data Service



UK Data Service

Annual report

April 2015 – March 2016

UK Data Service



Data enriched research
Data enhanced impact

Welcome to the third report of the activities of the UK Data Service



This third annual report of the UK Data Service highlights some of its achievements during the past year. These build on impressive earlier foundations. The ESRC's Data Infrastructure Strategic Advisory Group (DISAC) continues to be impressed by the performance of the Service, its ability to meet the needs of the ESRC, and its engagement with other key stakeholders in the wider data landscape of the UK and beyond. This report shows how the Service meets its objectives across a wide range of activities, underpinned by the core requirement of making it easier for researchers to access and use data, so as to help generate high-quality and valuable research.

Gordon Marshall
Chair, ESRC Data Infrastructure Strategic Advisory Committee

The ESRC remains proud to fund the UK Data Service. For the ESRC, the UK Data Service represents a key part of the ESRC's data infrastructure, not only through the provision of high quality data from government and researchers, but through providing the means by which the ESRC can implement its Research Data Policy. The Service not only continues to maintain and develop a world-class service to researchers, but has been engaging strongly with developments in opening access to different data sources, including government data, as well as representing our interests in the RCUK's Concordat on Open Research Data. This report gives clear evidence of the value of data infrastructure in facilitating and enabling high quality research, and playing a strong role in the promotion of data skills, a significant achievement in the changing data landscape.

Jane Elliott
Chief Executive, Economic and Social Research Council

This report covers the middle of the current UK Data Service award (2012-2017) and at the time of writing we are thinking strongly about the future. However, I find it very pleasing to reflect on some of the success stories of the past year, and how some were planned, and some were more serendipitous. The ability of the Service to be both proactive and reactive in a constantly changing data landscape is a result of our flexibility and a consequence of the level of autonomy which our funders allow us. This review provides only some of the 'big stories' of the year. Lots of what we do is 'service as usual'; at the end of this year we had around 22,000 registered users of the Service; during the year 423 Secure Lab outputs had been checked and released and over 2,000 people had attended our webinars and workshops. I think it's a fantastic effort and it's down to the enthusiasm and skills of the UK Data Service staff wherever they're based.

Matthew Woollard
Director, UK Data Service

Governing Board

Chair

Gordon Marshall,
Leverhulme Trust

Vice-Chair

Simon Briscoe,
independent

Members

Fiona Armstrong,
ESRC

Roger Burrows,
Goldsmiths

Gerald Donnelly,
National Records of Scotland

Guy Goodwin,
Office for National Statistics

David Marshall,

Northern Ireland Statistics and Research Agency

Maurizio Pilu,

Connected Digital Economy Catapult

Gesche Schmid,
Local Government Association

Sharon Witherspoon,

Head of Policy at the Academy of Social Sciences

About us

The UK Data Service is funded by the Economic and Social Research Council (ESRC) to meet the data needs of students, teachers and researchers from all sectors, including academia, central and local government, charities and foundations, independent research centres, think tanks, business consultants and analysts, communities and the commercial sector.

The UK Data Service provides access to the UK's nationally and internationally significant social science data assets for research, teaching, skills development and policy-making. The collection includes most major UK government-sponsored surveys, cross-national surveys, longitudinal studies, UK census data, international aggregate data, business data and qualitative data.

We offer

- Access to high-quality 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
- Development of best practice data preservation and sharing standards
- Sharing of expertise with international data providers to remove barriers to accessing data.

Contents

Research 04

Skill-building 14

Reach 20

Infrastructure 28

Looking ahead 34

Research

Enabling research,
informing policy,
improving practice

Making access to high-quality, regional, national and international social and economic data and resources easier

The UK Data Service is increasing its collection of open data and continues to support data owners in making their data available as widely as possible

Opening more data: The National Food Survey

During 2016, the UK Data Service collaborated with the Department of Environment, Food and Rural Affairs (Defra) to open National Food Survey (NFS) data. The initiative was part of Defra's 'Year of British Food', a campaign to celebrate British food. The NFS, which ran from the 1940s to 2000, collected detailed information on individual and household food consumption, including food diaries.

Defra and the UK Data Service both hold safeguarded digital versions of NFS data for the years 1974-2000. The Service supported Defra by informing a disclosure review (Privacy Impact Assessment) of the data to ensure that an open version could be a worthwhile research source without respondents' confidentiality being compromised. The Service also provided advice on the use of an Open Government Licence to cover the new data resource. As well as the open NFS files, the Service

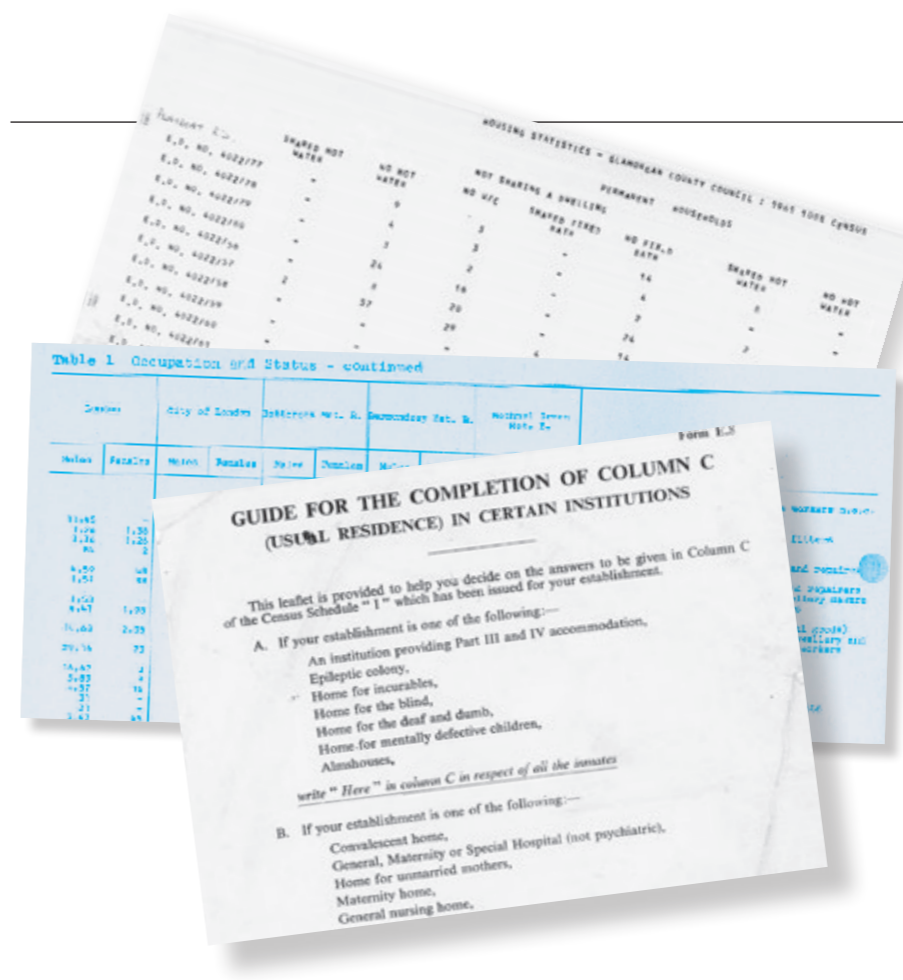
retains the official safeguarded versions to ensure that more detailed NFS data are available for registered users; the open files include household data with some cases removed and many variables banded or aggregated, whereas the safeguarded versions include detailed individual-level, household and diary data.

In 2001 the NFS became the Family Food component of the ONS Expenditure and Food Survey (now the Living Costs and Food Survey) also held by the UK Data Service. Disclosure control can be more difficult as the data are more recent, which means disclosure risks are higher. The open files will also be available from the Service.

Many of the reports have been digitised by the Hathi Trust at the University of Michigan. The UK Data Service helped in negotiating copyright transfer for the University of Michigan and Defra, and liaising with the academic team researching digitisation and data extraction from the reports.

It has been an exciting opportunity for the UK Data Service to become part of the open government data landscape. Defra's aim to become a more open, collaborative and data-driven organisation has meant that the Service has been able to support Defra and use its data expertise in new ways.

Matthew Woollard,
Director, UK Data Service



The 1961 Census: First steps towards a new open data collection

Much of the UK census aggregate data from 1970 to 2011 are available online from the UK Data Service. Aggregate and small area data and reports for the 1961 census are not available online. The UK Data Service and the ONS funded the PRIMA Research Lab at the University of Salford to carry out a proof of concept and feasibility project for the extraction of aggregate census statistics from a range of digital formats. The project aims to offer the potential for an unprecedented low-cost automated process to rescue and preserve these data, currently inaccessible in the form of micro-fiche and scanned images. PRIMA carry out research in document digitisation, optical

character recognition and other image analysis applications. A three-month proof of concept and feasibility project focused on a structured sample of the census. Initial results were encouraging, offering a 98% character recognition rate.

The results of the pilot are so promising that the Service is extending the capability of the methodology across a fuller range of output types from the 1961 census outputs to provide a methodology for digitisation of the complete set of 1961 outputs. The next stage will extend the existing framework to establish the new challenges and trial and evaluate the solutions.

#CiteTheData

Users can now enhance their search and download data citations in a variety of standard citation formats

Our data search function has been enhanced with the addition of an advanced search function so users can refine their search using fields such as title and depositor. Multiple search terms can be entered and combined by using the Any, All or Exact buttons.

Once researchers have found the data they need they can now easily get the citation in a variety of formats. There is also an option to export the citation into XML for CSL or Endnote. Making citation easier is at the core of what we do and improving mechanisms for data citation is so important to demonstrate the value of data sharing.

Our #CiteTheData campaign highlights a key objective of the Service – increasing the impact of the re-use of data in the collection, not least through the promotion of data citation.

Victoria Moody, Director of Impact and Communications, UK Data Service

InFuse: Our platform for UK Census aggregate data

InFuse, the UK Data Service's online search facility for open census data, is enabling tailored search and investigation of UK census statistics; opening new opportunities for impact

Using the Census to inform policy

InFuse is one of a range of advanced search and exploration applications developed by the UK Data Service Census Support team for providing access to census statistics.

UK census data are fundamental to improving understanding of life in the UK, supporting informed policy decisions by providing relevant and independent statistics. In the past, analysis of UK census data has been constrained by the separation of data and metadata, limiting social science investigations and research.

Data held in InFuse was made open in 2013, widening access to aggregate statistics from the UK 2011 and 2001 Censuses. Combining data supplied by each of the three national statistical agencies into InFuse widens the scope for research projects which require access to UK-level census aggregate data and those which compare geographical areas across the UK. The opportunity for users to access these data from a single integrated system is a unique feature of InFuse, and access to data via InFuse underpins a wide range of research across the higher education, public policy and commercial sectors.

1.

Sustainable travel

InFuse data were used for the *Smarter Travel* research project which studied how 'smart choices' for sustainable travel could be implemented and supported in transport planning.

2.

Integration in the UK

A study by the Marie Curie charity into the differing needs of Black, Asian and minority ethnic groups in Scotland for end of life care used InFuse to determine that the minority ethnic population in Scotland has doubled since 2001, highlighting the need for greater and more appropriate provision.

3.

Education and homelessness

A Knowledge Transfer Partnership between homelessness charity Llamau and Cardiff University used InFuse data to show that Welsh young homeless people participating in the study were over twice as likely to have left school with no qualifications compared to UK-wide figures for their age group and gender.



Engaging with the open data agenda

The Service has been working with the Open Data Institute to create open datasets which comply with the ODI Open Data Certificate

ODI Open Data Certificate

To reach the highest levels of certification the ODI Open Data Certificate needed really detailed stand-alone machine-readable information. The UK Data Service uses rich ODI metadata to describe the creation, availability of and rights associated with datasets in the collection, however the existing metadata fell a little short of expectations for supporting machine harvesting, so a range of additional information to the web content, metadata records and user documentation were added. A link was also added to the appropriate Creative Commons licence and database rights statement for every file available as well for as the whole collection.

After in-house checks the final decision about disclosure risk is made by data owners following their own in-house practices. Transparency about such decision-making on disclosure risk is often lacking, so the Service's Privacy Impact Assessment statement was enhanced to include information about higher-level protocols followed by social survey data producers.

After completing this work, the European Quality of Life Time Series, 2007-2011 and the Family Life and Work Experience Before 1918, 1870-1973 achieved Platinum level certification, which means these open data collections are an exceptional example of information infrastructure.

Concordat on Open Research Data

Matthew Woollard, Director of the UK Data Service, was part of the *Concordat on Open Research Data* Working Group. Designed to help ensure that research data created by the UK research community is openly available for use by others, *Universities UK*, *Research Councils UK* (RCUK), the *Higher Education Funding Council for England* (HEFCE) and the *Wellcome Trust* worked together to capture the best practice in the research community. Launched in July 2016, with an introduction by the then Minister of State for Universities and Science, the Concordat is intended to establish sound principles for working with research data, promoting the opening up of data for scrutiny and reuse, whilst protecting privacy, confidentiality and respecting consent terms. It proposes a set of principles for working with research data:

- open access to research data as an enabler of high quality research
- the right of creators of research data to reasonable first use
- good data management as fundamental to the research process and should be established at the outset
- the use of others' data should always conform to legal, ethical and regulatory frameworks, including appropriate acknowledgement

// This Concordat represents a more unified view across UK research data related stakeholders about access conditions and has the possibility of leading to a step change in the understanding and application of data access. //

Matthew Woollard,
Director, UK Data Service

Supporting excellence in social science research

Data in the UK Data Service collection has been used in research that supports policy, stimulates debate and informs the public on key social science issues

Data use in research, teaching and policy

The UK Data Service works with researchers to understand the innovative ways that data in the collection are being used to develop case studies to communicate and promote the use of the data in research, teaching and as a basis for informing policy. The case studies can be searched through the Service's Discover interface.

Does broadband access make UK firms more successful?

Popular wisdom might assume that access to broadband enables businesses to have greater connectivity and therefore productivity, however there was no evidence for this until Timothy DeStefano, Richard Kneller and Jonathan Timmis from the University of Nottingham investigated whether the adoption of Asymmetric Digital Subscriber Line (ADSL) broadband impacts firm performance in the UK.

Their research focused on a five-year period during which there was disparity in the regional availability of ADSL broadband, where firms in one location had access to broadband services while those in the other location did not. The results showed that whilst greater access to broadband infrastructure significantly increased the likelihood of firms adopting it, there was no evidence suggesting that using broadband, as a result of availability of ADSL, had an effect on employment, sales, labour productivity or firm survival. This project suggests that ADSL broadband access does not have economic impacts for UK firms.

Nearly a third of Welsh adults struggling to cope with the pain of chronic conditions

Dr Ivy Shiue, Senior Research Associate in Northumbria University's Department of Public Health and Wellbeing and formerly of Heriot-Watt University, conducted a study exploring long term conditions among Welsh adults and found that thousands of people have not learnt to live with the symptoms of long term health conditions.

Findings showed that nearly 30% of the respondents reported that they hadn't learnt to cope with the mental or physical pain associated with their condition. Around 10% reported untreated depressive or anxious symptoms. Around 20% however, said they had learnt to live with their chronic conditions.

// Future public health intervention and nursing rehabilitation programs that could help ease pains for people with historical health conditions, in particular mental and neurological disorders, to assist them with learning to live with their symptoms should be prioritised along with health prevention nationally. //

Dr Shiue

Creating percentile charts of body mass index by age and gender

Crystallise, an analytics and modelling company that services the finance, healthcare and pharmaceuticals sector, explored the distribution of body mass index as a health risk factor. More specifically, the company tried to identify the distribution of body mass index by age and gender among the population in England. This work, using the Health

Survey for England 2012, contributed to Crystallise's development of morbidity and mortality models to service the finance, healthcare and pharmaceuticals industries.

Products developed as a result of the research include percentile charts and a modelling tool, Sonata, which are freely accessible on the company's website. Sonata includes some of the features behind the Crystallise framework. It is complemented by visualisations useful for clinical decision making and shared decision making in clinical practice allowing users to explore the impact of differing assumptions of the relationship between risk factor values and mortality.

The Data Impact Blog

Bringing in a new audience stream for the Service and reaching almost 8,000 new users in its first 18 months. The Data Impact Blog showcases the impact of the Service and the use of data in the collection in the development of impact in research and teaching. The blog is a resource for all stakeholders to engage in debate, share best practice and keep up to date with news, events and the latest data-driven research and policy-making. Posts come from a range of data creators, researchers and teachers as well as demonstrating the expertise of UK Data Service staff.

Read the Data Impact Blog at:

<http://blog.ukdataservice.ac.uk/>

A PhD journey: Social and spatial inequalities among ethnic minorities in England and Wales

As part of a new strand of work supporting early career researchers, the Service is focusing on the PhD journey. Carolina Zuccotti was awarded her PhD by the European University Institute and her research explored the production and reproduction of social and spatial inequalities among Black, Asian and minority ethnic group in England and Wales.

The study was based on data from the ONS Longitudinal Study and on UK Census aggregate data from 1971 to 2011. In particular, Carolina chose to analyse data from the ONS Longitudinal Study because it enabled her to combine neighbourhood information with information on individuals, including their ethnic and social backgrounds.

Carolina's career journey: Carolina is now a Research Fellow at the University of Brighton and works with Professor Jacqueline O'Reilly on the STYLE (Strategic Transitions for Youth Labour in Europe) project. The aim of this project is to provide a comprehensive understanding of the causes of very high unemployment among young people and to assess the effectiveness of labour market policies designed to mitigate this phenomenon. Carolina is investigating whether and to what extent the family of origin affects the employment probabilities of young individuals in the UK with different ethnic backgrounds.

// Although I knew that I could link neighbourhood data to individuals, it was only when I started working in on it that I realised the potential of this link - and this led me to new research questions and a new chapter in my thesis. //

Carolina Zuccotti, University of Brighton



Creating better knowledge for a better society

In summer 2015 the UK Data Service and AppChallenge collaborated to launch a developer contest using open data about the Quality of Life of European citizens

#EULife AppChallenge

The UK Data Service #EULife AppChallenge crowdsourced app ideas from developers across the world, with entries from countries including India, Sweden, Serbia, Germany, Finland, Estonia, Poland, the United Kingdom and even New Zealand. Nine prizes were available with a prize fund of £15,000 judged by Google, the Open Data Institute, the UK Data Service, Royal Statistical Society, Hack & Craft, Digital Catapult, DSRPTN and the data owner, Eurofound. The nine AppChallenge winners were:

- £2,000 – Young Person’s prize (Aged 16-24 Years) Tomek Kwiatkowski for EU Connected Quizzes
- Five Judges “commended” prizes of £500 each
- Nebojsa Kamber for EU Fun Stats
- Jessie Johns & Milady Capitan for Euro Hero
- Philip Hardwick for 30 Day Impact
- Tomasz Mendak for Life Meter
- Leon Kukuk for EUConnect

Two students from Barking & Dagenham College in East London, Milady Capitan and Jessie Johns were commended in the #EULife AppChallenge for their entry, Euro Hero, which sets out missions to educate users about EU member countries.

£7,500

First Place

Tomek Kwiatkowski

for EU Connected Quizzes which connects a geography quiz with questions that develop the user’s understanding of European society. Tomek’s design is being developed into the Service’s first app aimed at the 14-16 age group to develop their engagement with data.

£2,000

Second Place

Lyall Davies for Citizens Europe

£1,000

Third Place

Deepak Bhatia for TopicCast

“While I was studying IT, I got interested and learned how to code. I think coding particularly and programming is a very male dominated domain...and I think it’s really important to showcase women entering into these competitions. It’s...a really good step to showcase that they can do these things and hopefully we’ll see more women using coding in their general day-to-day work.”

Jessie Johns,
AppChallenge prize winner

ONS Approved Researcher Scheme

The UK Data Service provided expert guidance to the Office for National Statistics on the new Approved Researcher Scheme

The Approved Researcher scheme is used by ONS to grant access to data that cannot be published openly, for statistical research purposes, as permitted by the Statistics and Registration Service Act 2007 (SRSA). It became the gateway through which ONS provided access to microdata which has a risk of disclosure. In 2015 the scheme was reviewed and revised, with the aim of the new scheme to make research using sensitive data more transparent and to widen access to data to researchers and other user communities who currently have difficulties accessing these type of data.

Matthew Woollard, Director of the UK Data Service, was part of the expert steering group that advised on the new scheme.

The main changes to the Approved Researcher scheme are:

- separate approval for researchers and projects, allowing researchers to take their ‘approved’ status to different projects
- the option for researchers to request ongoing access to a dataset
- changes to the level of experience required before accessing data, with explicit qualifications and research experience to be set in order to obtain Approved Researcher status
- new Provisional Approved Researcher status for new researchers who are working under the direction of a full Approved Researcher
- allowing researchers from overseas to apply for Approved Researcher status, if their proposed research benefits the UK

These changes should allow more postgraduate students to analyse these types of data and help early-career researchers develop their research portfolio. Researchers from beyond the Higher Education and Government sectors will also be able to apply for access to these data, but only if the research has benefits to the public and all research is published.

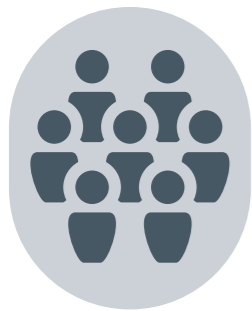
“We very much welcome these new guidelines, which will improve the situation for researchers who want to use these datasets in their work. Having separate approval for researchers and projects, for example, means a researcher can be approved for up to five years, and make multiple applications for different sets of data in that time. Also, if they have ongoing access to a dataset, they don’t have to re-apply if a new wave is released. The new provisional status will allow researchers without previous experience to work on a project under the direction of a ‘full’ Approved Researcher – helping those new to the data community to learn and develop, increasing our capacity in the UK to carry out more of this vital work.”

Matthew Woollard, Director,
UK Data Service

Skills-building

Teaching core
data skills

Supporting data users through training and capacity building in data skills



New developments in our webinar programme

The UK Data Service has been running a series of webinars introducing big data analytics and solutions, such as *What is Hadoop?*, and showcasing some of the research projects from the ESRC Big Data Network. We have also produced a series of short online guides related to the big data webinars, giving users more detailed information on big data analytics.

Themed webinars have been introduced to support research and teaching on particular topics including ethnicity, education, ageing and ICT.

// Good course, has definitely improved my knowledge and made me more confident in Hive //



Helping students with their dissertations

The Service has developed a suite of student resources webpages for students using secondary data in their research. The webpages contain information about UK survey data, census data, international macrodata and qualitative data held by the UK Data Service. We also developed a guide to using survey data to help students conducting project work and dissertations get started with projects using quantitative collections.

We continue to offer new ways of ensuring students, researchers and teachers have the right skills to use data in the collection.



Student Forum

We have developed a Student Forum on Facebook. The Forum hosts discussions on how to source and use data from the UK Data Service. Open to all students; undergraduate, taught postgraduates and doctoral students, the forum helps facilitate peer to peer support.

// It made me aware that there are data sharing issues within my own PhD that I had not fully considered. //



Teaching Resources

Following consultation with teachers the Service has developed a Teaching Resources Rubric – a one stop shop which provides information on the suite of Methods Guides available from the UK Data Service.

We have also updated our Teaching with Data web pages which now include guidance on making your own teaching data, as well as the quantitative resources and qualitative/mixed methods resources available for teaching. The UK Data Service also has datasets specifically designed for teaching. Based on actual datasets in our collection, but reduced to a manageable size and simplified to make them more accessible and engaging for students. Newly released open teaching datasets include the 2013/14 Crime Survey for England and Wales, the Jan-March 2015 Quarterly Labour Force Survey, the April-May 2015 Opinions and Lifestyle Survey and the 2012/13 English Housing Survey.



Working collaboratively with other ESRC training investments

The Service has worked collaboratively with other ESRC data-related training investments including the National Centre for Research Methods (NCRM), Administrative Data Research Network (ADRN), the CALLS-Hub, Cohort & Longitudinal Studies Enhancement Resources (CLOSER), Understanding Society, the Centre for Longitudinal Studies (CLS) and the Big Data Network Phase 2, to provide more joined up communications for users and joint events, such as a webinar on data resources for research students and a face-to-face workshop on supervising quantitative research degrees.

// Very good event. The information will be very useful in my job and I have made contact with the right people in several areas. //



Access to sensitive data for research: The Five Safes

Under the Secure Data Service we have provided secure remote access to data deemed too confidential for download

Safe People, Safe Projects, Safe Settings, Safe Outputs, Safe Data

The UK Data Service's remote access system, known as the Secure Lab, enables researchers who have completed the necessary steps to log in to a secure server based at the UK Data Archive, University of Essex, and access data for analysis. Once the researcher's analyses are completed, the statistical results are screened by Service staff to ensure no link between the results and the original data can be made, preserving the confidentiality of the data subjects. When declared 'safe', the results are returned to the researcher.

We operate the facility on five simple protocols:

Safe People: Only approved researchers may access data through the Secure Lab. These are researchers whose interest in accessing the data is for public benefit. Researchers must register with the UK Data Service, submit a detailed project application justifying their need to access the data and suitability for accessing and handing the data. Researchers must also have approval from their own institution and undertake a mandatory SURE training course.

Safe Projects: Only projects which serve the public good and are approved by the data controller can be undertaken in the Secure Lab. Researchers must explain how their research will benefit the public good when they apply. Data controllers make the decision what is or is not in the public good.

Safe Settings: The UK Data Archive Secure Lab uses Citrix secure remote access technology, frequently used by the banking and military sector and renowned for its robustness. The UK Data Archive is accredited to the ISO 27001 Information Security Management standard and the outcomes of the 6-monthly ISO surveillance audits are made available to the dozen or so Government Departments and other agencies that supply the Service with confidential data for research.

Safe Outputs: Only statistical outputs which have been screened by staff to ensure they could not be used to identify the data subjects can be released to the researcher. Our staff carry out disclosure control checking to ensure that confidentiality is maintained.

Safe Data: Because of the high security standards applied, the Service is able to provide access to 'unsafe' data. These are data from which it is relatively easy to identify and data about individuals with sensitive variables. Direct identifiers such as names and addresses have been removed, but the data are still confidential or sensitive and are considered 'personal' under the Data Protection Act or the Statistics and Registration Services Act.

A new integrated approach: training researchers to use sensitive microdata

Implementing the Five Safes: the new Safe Use of Research data Environments training (SURE) for new users of Secure Lab and other Research Data Centres



It is vital that researchers using our remote access Secure Lab share our views on data security and have a full understanding of the processes in place to ensure that the sensitive microdata held in the Lab are used in an appropriate, and therefore safe, way.

SURE training

SURE is the latest evolution of the UK Data Service's training programme for users of controlled data. A driving force behind the redevelopment of this course was the opportunity to collaborate with other secure access providers. Several Research Data Centres (RDCs) operate across the UK, all of them offering access to sensitive microdata. Each RDC was providing broadly similar training and it was clear that the current process of requesting that researchers train through one RDC and then, if they want to access data from further RDCs, undertake a second, or even third, training course, was not the best use of resources.

The new course was developed through collaboration between the UK Data Service, HM Revenue and Customs, the Office for National Statistics and the Administrative Data Research Network with the aim of making accreditation transferable between these services. Not only does collaboration reduce the training burden on researchers who were previously required to travel to different

locations to train for each service that they wished to use, it also makes for better use of resources for each RDC, which is no longer compelled to train researchers that have already been trained elsewhere.

Working responsibly with secure data

The course covers many facets of working with these data, including trust in researchers, the Five Safes security model, relevant legislation, sanctions for security breaches, and statistical disclosure control. Researchers are encouraged to think about how all of this relates to their own proposed research.

The approach helps to cement the core messages and gives the researcher a good idea of what those checking for disclosure issues are looking for and why. By having a full understanding of the various ways in which potentially disclosive information may be presented, researchers can avoid submitting outputs for release from Secure Lab that will be returned to them for changes before they can be passed as 'safe'.



Any researcher attending this course has a greater understanding of the 'principles-based' approach to statistical disclosure control and an increased awareness of how to use these sensitive data as safely as possible, helping them to carry out their analyses in a safe and secure way.

Reach

Extending access,
support and expertise

UK Data Service
ukdataservice.ac.uk



 @UKDataService
 blog.ukdataservice.ac.uk

#DataImpact2015

Data innovators from the public, commercial and higher education sectors discussed the challenges and opportunities for data infrastructure in relation to the big issues in public policy debate at #DataImpact2015

#DataImpact2015 took place on the evening of 2 December 2015 at the Digital Catapult Centre, London, with the theme *the impact of social and economic data re-use*.

A panel of leading data innovators explored data re-use in policy and research, sharing their experiences of demonstrating data enhanced impact. Introduced by Matthew Woollard, Director of the UK Data Service and Chaired by Jane Elliott, Chief Executive of the ESRC, speakers included Sharon Witherspoon, former Director of the Nuffield Foundation, Will Moy, director of Full Fact, an independent fact-checking charity, Dan Sheldon, digital strategy lead at the Department of Health and Andrew Collinge, Greater London Authority lead for the Smart City Agenda.

Over 60 delegates attended from the higher education sector, including Universities UK and LSE Impact Blog; central government, including Defra and ONS, and local government including Westminster City Council; the voluntary and community sector, including Wellcome Trust, Open Knowledge Foundation and DataKind, and commercial sector including ScraperWiki. #DataImpact2015 briefly trended on Twitter on the evening of 2 December, a first for the Service.

#DataImpact2016

takes place in Glasgow on 13 October 2016.

“There is huge potential to share data much more effectively.”

Andrew Collinge, Assistant Director
Greater London Authority

“We’re the organisation that listens to what people are saying in public debate, whether that’s politicians or journalists or pressure groups and we go back to the data and find out what it says and we make that data available to the public, the media and decision makers to make up their own minds about the big issues.”

Will Moy,
Director Full Fact

“This sort of event matters because you’re bringing together people from public sector, private sector and the more we can bring together those disparate audiences to talk about it, the better a public policy discussion we’ll have.”

Sharon Witherspoon,
Head of Policy at the Academy
of Social Sciences

GeoConvert: An easy-to-use web interface for converting between geographies

Enabling users to match UK postcodes with census geographies and convert between them with GeoConvert

Matching different geographies

The boundaries of geographic areas (such as districts, wards, output areas, postcode sectors) are a form of geospatial data stored as coordinates that can be used for mapping distributions of attribute data, for example census characteristics. Some geographies form hierarchies in which smaller areas from geographies at lower levels nest neatly within larger areas from geographies at higher levels, but many do not. Most geographies change over time as areas change their names or extents, or are terminated, split or amalgamated to form new areas. GeoConvert automates the matching and conversion processes between different geographies using the best information and methods available to identify and record relationships between the individual areas and within and between geographies.

Postcode building blocks

In the current version of GeoConvert, area relationships have been derived using information from the three UK census agencies (ONS, NRS, NISRA) about the populations of the sets of special enumeration postcodes that were the fundamental building blocks of areas used for outputs from the UK 2011 Census, together with their relationships with larger geographical areas. The platform enables the populations of postcodes to be aggregated into different census areas and the area relationship factors are then defined between pairs of areas based on the spatial distribution of population recorded in the 2011 Census. GeoConvert enables users to supply sets of identifiers for areas of one geography and have them matched to identifiers for areas of another and data supplied for one area to be reapportioned to different areas based on the area relationships between them.

Open data

All the data held in GeoConvert falls under Open Government Licence and users don’t need to register or log in to use the platform. In addition to matching geographies and converting between 1991, 2001 and 2011 Census geographies, GeoConvert also offers a range of commonly used data, for example indices of deprivation for some areas, primarily postcodes.



Smarter Household Energy Data

Strengthening data expertise and research partnerships between South Africa and the UK

Developing an internationally networked cohort of experts

Smarter Household Energy Data is a collaborative project between the UK Data Service and DataFirst, South Africa, to investigate access to, and the use of household energy data.

The project will strengthen data expertise and research partnerships between South Africa and the UK through the formation of a new network aimed at investigating access to and use of household energy data for policy-making and planning. The work planned focuses on data infrastructure and brings together data professionals, energy researchers and policy makers in South Africa and the UK.

“The availability of household energy is one of the key challenges for developing countries, like South Africa. This project has the potential to help shape the way decisions are made by providing robust data infrastructure that can enable better insight into impacts of energy supply and usage on household behaviour, and the situation of individuals within households. I’m very excited about the potential impact this project could have and am delighted to be heading up our multi-national team.”

Louise Corti, Director Collections Development and Producer Relations, UK Data Service

The challenge

Substantive cases from South Africa highlight some of the challenges faced both by the shortage of data needed to provide evidence for evaluation and impact monitoring of for example, energy consumption, fuel poverty and demand reduction, but also the lack of systems and tools to extract and present intelligence from data.

Teams will work across the network to address key policy challenges through shared scoping, planning and analysis of UK and SA household energy data. This proposal connects centres of expertise in areas where there is less capacity than is needed for progressing the next generation of data infrastructure, and in turn, foster new research agendas. Because the centres involved are recognised as key players in their industries, the fusion of their activities will enhance international pathways to impact.

A big focus of the project is anticipating and finding solutions for the challenges of ‘big’ and complex data in three main areas. The first is confidentiality, legal and ethical issues, particularly important in relation to administrative data and local level census and Demographic Surveillance Site data. The second relates to size and complexity, for example municipal datasets are of a size that social scientists in South Africa are not used to analysing, and the added challenges that come with combining data from different data sources raises analysis issues of a different type. The third is surrounding data quality – to identify and address missing data, data errors and uncertain or unknown provenance.

For the UK Data Service the main aim of the project is to develop and test infrastructure for big data with energy data as a realistic use case.

Supporting human rights organisations to develop insight from data

Improving the way civil society organisations can translate data into knowledge

The UK Data Service held a workshop “Supporting human rights organisations to develop insight from data” on 29-30 October 2015 at the University of Essex with support from the ESRC as part of its work to engage with civil society.

The workshop provided a forum for participants to discuss the strategies, tools and skills required for civil society organisations to help shape ways of managing and sharing data ethically, analysing the data, and gaining insight to develop impact from these rich data resources.

Human Rights organisations attending the workshop included those involved in supporting victims of human trafficking, torture, unfair trials, conflict and war and other vulnerable situations. Data about the people represented by these charities offers particular challenges around privacy and data protection, and hence, wider sharing; challenges that the Service is uniquely positioned to address.

Speakers included Neil Serougi (Trustee, Freedom from Torture), David Walker (the Guardian), Emma Prest (DataKind UK), Nigel Fielding (Surrey University), Matt Williams (COSMOS, Cardiff University), Tracy Gyateng (New Philanthropy Capital), Christina Rowley (ESRC’s Civil Society Engagement), and Louise Corti and Libby Bishop (UK Data Service, University of Essex).

“Really fantastic, useful and interesting workshop, speakers managed to bring the subject to life, make it practical and understandable.”

(Workshop attendee)



Smarter
Household Energy Data

Preserving Social Media and Transactional Data

We commissioned the Digital Preservation Coalition to develop two Technology Watch Reports on preserving social media and preserving transactional data

Social media plays an increasingly important role as we embrace networked platforms and applications in our everyday lives. Transactional data are generated through interactions with database systems and can be data created, for example by automatic sensors, or by interactions between government database systems and citizens. Transactional and social media data have potential for research into consumer analytics, but their use will only lead to new discoveries and insights if they are curated and preserved to ensure they can be reproduced.

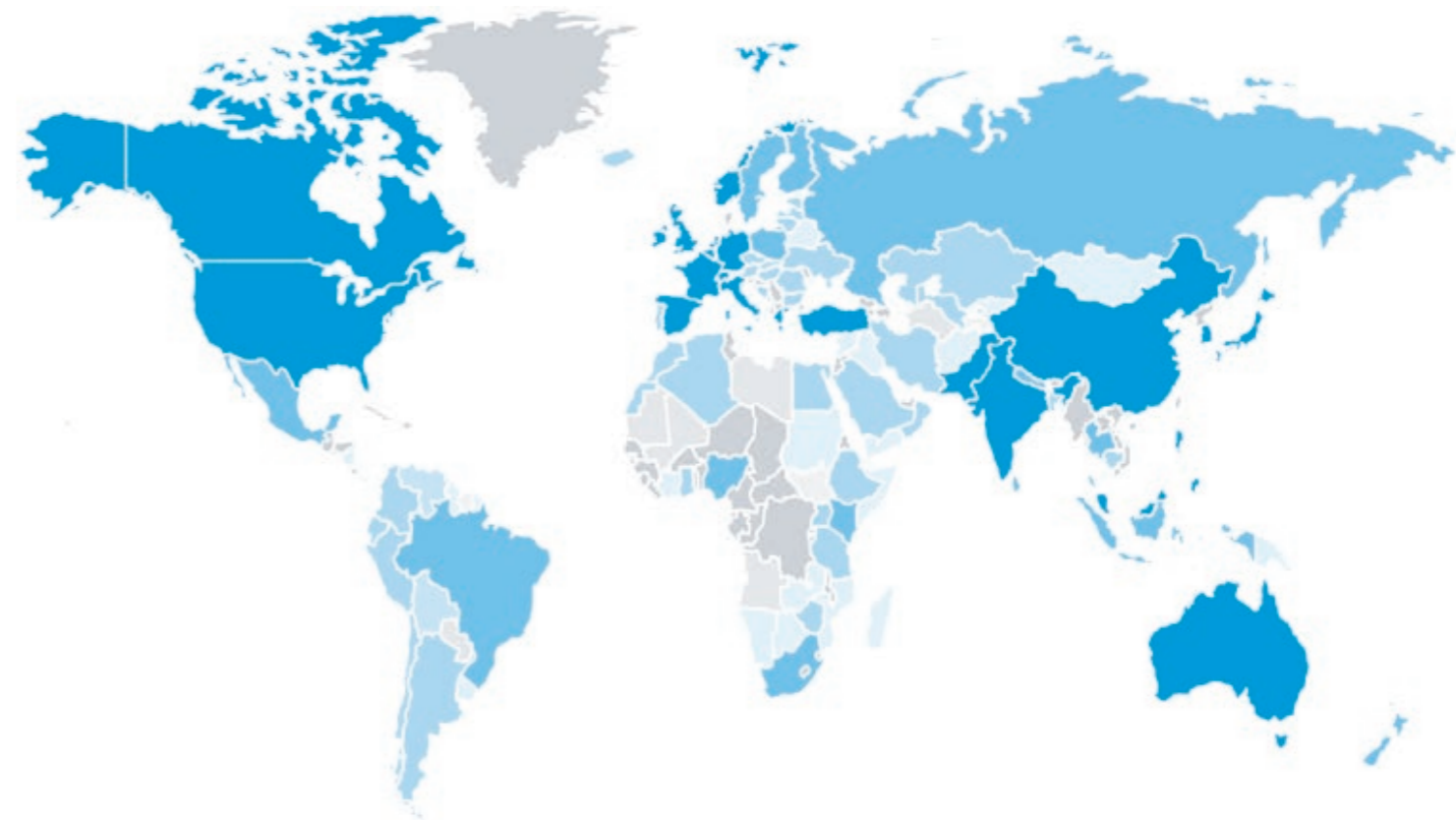
Preserving Social Media Data provides guidance for researchers wanting to access social media for research purposes, the institutions who support them, and all organisations with a need to preserve social media data. It describes the landscape of archiving social media data, including the attendant legal and ethical obstacles to long-term access.

The report lays out a number of proposed approaches to the preservation of social media data - a valuable resource currently at relatively high risk of disappearance if not actively addressed. Though the report addresses a number of significant challenges, it focuses on new developments and growing motivation across disciplines to ensure that future generations have access to social media created today.

“ The preservation of social media has a wide appeal and this report is likely to be of interest not only to DPC members, but many organisations throughout the digital preservation community who face the challenge of keeping user generated content through social media accessible in the future. ”

Neil Beagrie, editor,
Technology Watch Reports

Preserving Transactional Data considers strategies for preserving these data and the challenges for companies and institutions that re-use them. The preservation methodologies developed by government and other organisations for transactional data may differ from the preservation methods needed for analysis or academic research, which will involve developing strategies which take into account the fragmented nature of many database systems.



> 270,000

Global use of the UK Data Service continued to grow in 2015/16. More than 270,000 UK-based and 25,000 US-based new users accessed our website in a year when the UK Data Service was used in almost every country in the world.

Infrastructure

Providing for future
data, technologies
and research

Data Service as a Platform (DSaaP): providing a connected and accessible framework for ‘big data’

The public and commercial sectors generate data at volume, speed and a level of complexity, which present challenges for their curation and reuse. These ‘big data’ offer a significant resource for research, but how do we share and curate them’?

New infrastructure for ‘big data’ analytics

The UK Data Service is developing new infrastructure to provide a connected and accessible framework for ‘big data’; an ‘on-premises’ data centre based at the UK Data Archive, University of Essex and a cloud hybrid platform.

Our Data Service as a Platform (DSaaP) is based on the Hortonworks Data Platform on the Apache Hadoop software. By developing this hybrid model we can understand how to deliver a complete big data service which takes into account data access requirements for both open and controlled (secure) data. The DSaaP is developed from a set of standard open source technologies and uses a stable reference architecture (HDP 2.4) to deliver big data solutions for social science research. The main purpose of the DSaaP is to allow researchers to access and analyse data in an environment which meets their needs, but it will also allow us to provide a single data access point for all these new and novel forms of data. And, it will allow us to experiment on curation routines to maintain the data.

To develop the DSaaP we are running a number of targeted case studies with real data, including Smart Energy Meter Data, Pay Thru Data and local authority data from the ESRC’s BLG Data Centre, labour force data, cohort studies and open data useful for linkage such as climate data, geographies and historical census data.

New resources for users

Helping users to find the information and data they need



Providing user access to the 2011 Census flow data

The 2011 Census Origin-Destination Statistics collated and released by ONS are more extensive than those generated from previous censuses; covering student mobility and second home ownership as well as a greater number of tables detailing migration and journey to work flows.

The UK Data Service’s Census Support team has been working with ONS to provide user access to the safeguarded flow data from the 2011 Census through the WICID interface available from the Census Support website of the UK Data Service. Since there may be a risk of disclosure, users of these datasets are required to register with the UK Data Service and log in prior to building queries and extracting data, and to abide by usage and publication conditions.

New resources on our homepage

We launched our refreshed homepage in October 2015 with several new features and a focus on changing content, with the aim of increasing user activity on the website. An upward trend in visits is demonstrated, with a 25% increase in sessions up from 363,354 to 453,866 and an increase in returning visits at 51% for this period as opposed to 34% in the same period last year. Google Analytics shows that the homepage bounce rate has fallen by 87%, meaning significantly more users are continuing on to view other pages in the website than before the homepage refresh, getting to the resources they need faster. We have added a ‘latest data’ feed, and a rolling programme of videos from data creators, researchers, teachers, students and platform developers. We’ve also included a data impact feed showcasing data use in relation to the big issues in public policy debate.

The animation *About the UK Data Service* has been viewed nearly 1,000 times in just a few months.

CESSDA, the Consortium of European Social Science Data Archives

Providing a common set of tools, tests and code repositories for CESSDA members to use, helping to develop practices and align standards for research data infrastructure across Europe

Developing a technical framework for CESSDA

The UK Data Service has been working with our European partners to develop a technical framework for the Consortium of European Social Science Data Archives (CESSDA) members to use as part of their research infrastructure. CESSDA supports national and international research and aims to provide a comprehensive, distributed and integrated social science research data infrastructure in the European Research Area. The technical framework provides a common set of tools, tests and code repositories for CESSDA members to use, helping to develop best practice and align standards for research data infrastructure across Europe.

The project will enable systems to be built and tested to comply with common standards, enabling CESSDA to provide a full scale sustainable research data infrastructure that supports the research community across Europe.

Metadata management

The CESSDA Metadata Management project started in November 2015 and will develop a standardised metadata design, content and practice for all CESSDA data assets. The CESSDA Metadata Standards Portfolio Version 1 will become the standard for all Service Providers within CESSDA.

The Portfolio will be built mainly upon the Data Documentation Initiative-Lifecycle standard, but will include elements from other relevant standards where appropriate. The model will be platform independent and designed to enable the adoption of the Portfolio for past, present, and future CESSDA data assets.

The Portfolio will also contain CESSDA controlled vocabularies for relevant metadata fields, taking into account and supporting the Data Documentation Initiative Controlled Vocabularies Group (DDI CVG) work.

There are eight CESSDA service providers involved in the project with the UK Data Service leading the work on the controlled vocabularies.



SERISS project

The UK Data Service is also participating in an EU-funded project that complements the work that is being carried out within the CESSDA-ELSST (European Language Social Science Thesaurus) project. The Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS) project aims to address some of the key challenges for cross-national data collection, including facilitating greater harmonisation of data collection, analysis and curation across social science infrastructures.

The UK Data Service Thesaurus Team will be working on 'Maximising equivalence through translation', investigating different approaches to the translation of questionnaires, and conducting a comparative empirical assessment of thesaurus keywords. This project will evaluate ELSST and compare the translation quality of ELSST concepts with that used in questions in surveys conducted by the European Social Survey (ESS) or Survey of Health, Ageing and Retirement in Europe (SHARE).

Data Access Policy for CESSDA

During this year, the UK Data Service was funded to develop a unified Data Access Policy for CESSDA, based on the requirements of the OECD's Principles and Guidelines for Access to Research Data from Public Funding. The policy takes the view, that all data holdings which are not subject to privacy (or other legal) requirements should be available to anyone regardless of status, whether academic, government sector or member of the public, regardless of nation or type of use. CESSDA believes in open science while promoting the protection of the privacy of data subjects, as well as the protection of the intellectual property rights of the data owners.



© CESSDA

Looking ahead

This report demonstrates the wide range of projects we have developed to improve the experience of UK Data Service for users and depositors and to attract new users from a range of sectors. There is more to come – the UK Data Service's team of experts has already begun work on some new and exciting initiatives, on which we can report outputs and impact in next year's Annual Report. Take a look at some of the work we have underway:

Data Skills Modules

Data specialists from the User Support and Training team are working with an award winning learning technologist to produce a set of self-learning data skills modules. Three modules introducing key data types (surveys, aggregate data and longitudinal) will be produced over the next few months with the expectation that a qualitative module will follow. Each module will describe key features of each data type and how the data are used, set tasks for learners to try out and provide feedback on understanding through the use of online quizzes.

Pop up workshops

We have developed a series of flexible 'pop-up' workshops in locations where appropriate computer clusters are difficult to find. The workshops involve practical computing sessions for users to get hands-on experience with the data. We have purchased 16 laptops for delivery of the pop-up workshops and we have a series of workshops planned for the coming months.

Data Impact Fellows: Development opportunities for early career researchers

The UK Data Service aims to establish additional ways to support the long-range use of its data and resources by new generations of scholars, extending this usage through the research partnerships they develop and by the students they teach - from the earliest stages of, and throughout their career - by establishing a prestigious Impact Fellows programme. Funded by the ESRC the programme aims to provide career development opportunities for scholars at a relatively early stage of their academic careers with a proven record of research. We will report on the Data Impact Fellows' journeys so far in 2017.

Who's who

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

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 almost 50 years. It also hosts the Administrative Data Service, and carries out research and development in other areas of data service infrastructure.

Jisc

Jisc is 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 rich resources. Through the UK Data Service Jisc provides access to, and specialist support for, the databanks of such prestigious organisations as the International Monetary Fund, the World Bank and OECD as well as aggregate statistics from the 1971 to 2011 UK censuses.

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 UK secondary microdata including Census and UK Government Survey microdata.

School of Geography, University of Leeds

The School has 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

The Population, Health and Wellbeing research group has 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. Their work comprises spatial population analysis and modelling; cultures, spaces and practices of care and population health.

EDINA, University of Edinburgh

EDINA acts as 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.

ukdataservice.ac.uk

UK Data Service

