



# UK Data Service

## Annual report

October 2012 – March 2014



Data enhanced research  
Data driven impact



The UK Data Service has been carefully designed so as to integrate previously disparate investments in order to maximise the potential of data reuse within the UK and beyond. This planning has paid off: the Service is now an efficient and innovative resource with more than 22,000 registered users. Its remit and reach have also expanded considerably, and we are working hard to increase access to a wider audience, and to open data where there are appropriate legal gateways in place.

This report highlights a number of the Service's achievements, but to me the greatest achievement is the huge success of maintaining the high-quality service, while undergoing considerable organisational change.

Gordon Marshall  
Chair, UK Data Service Governing Board

The ESRC is proud to fund the UK Data Service. The Director and his colleagues have been assiduous in constructing a value-for-money national infrastructure which is admired internationally. The UK Data Service is part of the foundation for many other ESRC investments including the Administrative Data Research Network and other elements of the ESRC's Big Data Network.

In the past 18 months, the Service has transformed itself seamlessly from a group of investments into a single entity which leads the world in access to data for social science research. There are so many strands of excellence across the Service, but its ability to provide input and advice for so many of the ESRC's activities is enormously helpful.

Paul Boyle  
Chief Executive  
Economic and Social Research Council

Establishing a new service on long-standing foundations has provided challenges, but I believe we have been hugely successful in every area of our combined endeavours.

Independently managed, reliable and carefully curated data for research are critical to inform policy and analyse its effectiveness; our service strives to ensure we meet this goal. We have a robust but flexible workplan to cover the whole of the five-year contract, and we are meeting the challenges of a continually changing data landscape.

'Big Data' was hardly in use when we started our contract; now we are developing plans to support the ESRC's Big Data Network. The Open Data agenda was looming on the horizon in 2011; now we are supporting the Cabinet Office in consulting stakeholders on new data sharing legislation.

We will continue to adapt and lead in the change which is ahead of us.

Matthew Woollard  
Director, UK Data Service

# Welcome to the first report of the activities of the UK Data Service.

# About us

Data are the lifeblood of research. The UK Data Service provides the social science community with a dedicated portal to data that can inform society.

We are a comprehensive national service established and funded by the Economic and Social Research Council (ESRC) to support researchers, teachers and policymakers who depend on high-quality data.

We were established in October 2012 to integrate several data services (Economic and Social Data Service, Secure Data Service and the Census Programme) into one seamless service that meets the needs of academic researchers and students as well as a widening range of data users and data owners in government, charities, foundations, think tanks and businesses. Today our website provides a portal to a broad range of secondary data including UK Census data, large-scale government surveys, longitudinal data, international macrodata, business microdata and qualitative studies.

We serve more than 22,000 registered users in addition to a growing number of Open Data users. Our helpdesk, guides, webinars and training are reaching students and researchers at all levels and sectors.

# Contents

## Research 04

*Enabling research that informs policy and improves practice*

## Skill-building 08

*Teaching core data skills*

## Reach 12

*Extending data access, support and expertise to new sectors and users*

## Infrastructure 16

*Providing for future data, technologies and research*

## Looking ahead 20

*Preview and priorities for the coming year*

# Getting users to the data

In an age of vast and growing resources, social science researchers want one trusted place to find the data they need. Discover is delivering just that.

Providing a single point of entry for data supporting the ESRC's programme of research was a founding principle for the UK Data Service. Finding the best way to integrate discovery tools across collections as diverse as the Economic and Social Data Service, Secure Data Service, the Census Programme and other research outputs was daunting. Doing it in a way that makes sense to as many of our users as possible required considerable consultation and testing.

Today all website visitors can use the simple Discover interface to enter a single search and then filter results by various criteria. Well-constructed ranking algorithms bring the most relevant data to the top of the list.

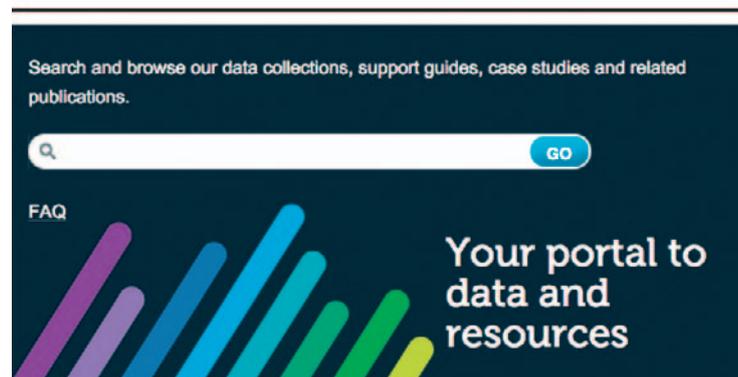
Discover was launched in March 2013 and, since then, has satisfied almost 265,000 searches, bringing our users to over 350,000 catalogue records.

We have also linked resources so that those looking for data can easily find related case studies, user guides and publications as well as similar data collections. Researchers can even search and browse for all these things at once.

We augmented Discover's core application with the Variable and Question Bank, allowing researchers to search for specific data variables and survey questions. We also added the ability to search for data by series – something of special interest to data owners and those who use our core survey collection. We have used controlled vocabularies wherever possible, in particular those developed by the DDI Controlled Vocabularies Group, to ensure the system is robust yet familiar to data users.

It's now a strong and reliable tool, and the front door to the Service – but our work is far from over. As technologies advance and user needs change, our journey of continuous improvement will continue with regular rounds of review, user experience testing and feedback.

## Discover data



Discover, our search engine portal to a wide range of data and resources.

# Streamlining access conditions

Gone are the days of controlling access to every piece of data used for research. Yet at the same time, individuals and organisations are increasingly wary of how personal data are being shared and used.

In an era when pressures for open access are pushing against privacy concerns, those seeking access to research data can get stuck in a muddle of confusing access routes, with each data owner identifying separate and inconsistent criteria for who can access their data – and for what purpose.

With over 101,000 data downloads since the start of the service, and more people downloading data year on year, it's important for us to make the user's journey to the data as simple as possible.

In 2013 we took a proactive step to streamline and clarify our data access policy, identifying three clear and distinct routes.

**Open** is for data which are not personal and have relatively few restrictions for use. For these data, registration and authentication are unnecessary.

**Safeguarded** is for non-personal data where the owner identifies a risk of disclosure from linkage to other data, such as private databases. For these data, it's important to know who is using the data, and why; therefore registration and authentication are required.

**Controlled** is only for data which may be identifiable and therefore carries a risk of disclosing information about individuals or organisations. In these cases, we continue to promote research access via the Secure Lab by ensuring those who use it are accredited, and that their research purpose is approved by the relevant data access committee.

While the vast majority of our data collection is currently classed as *safeguarded* – most commonly under the standard End User Licence, sometimes with special conditions attached – it is our explicit long-term goal to work with data owners to identify and remove all unnecessary barriers to access.

“ We recognise that the UK Data Service plays a critical role in facilitating social science research through data access. Yet we also have a responsibility to data providers, ensuring they are used responsibly and to best advantage. By defining clear criteria and close collaboration, we can help define the optimum access standards for each particular data collection. ”

Matthew Woollard  
Director, UK Data Service

# Bridging language gaps across Europe

Enabling data access and use across national and cultural boundaries involves more than systems and laws.

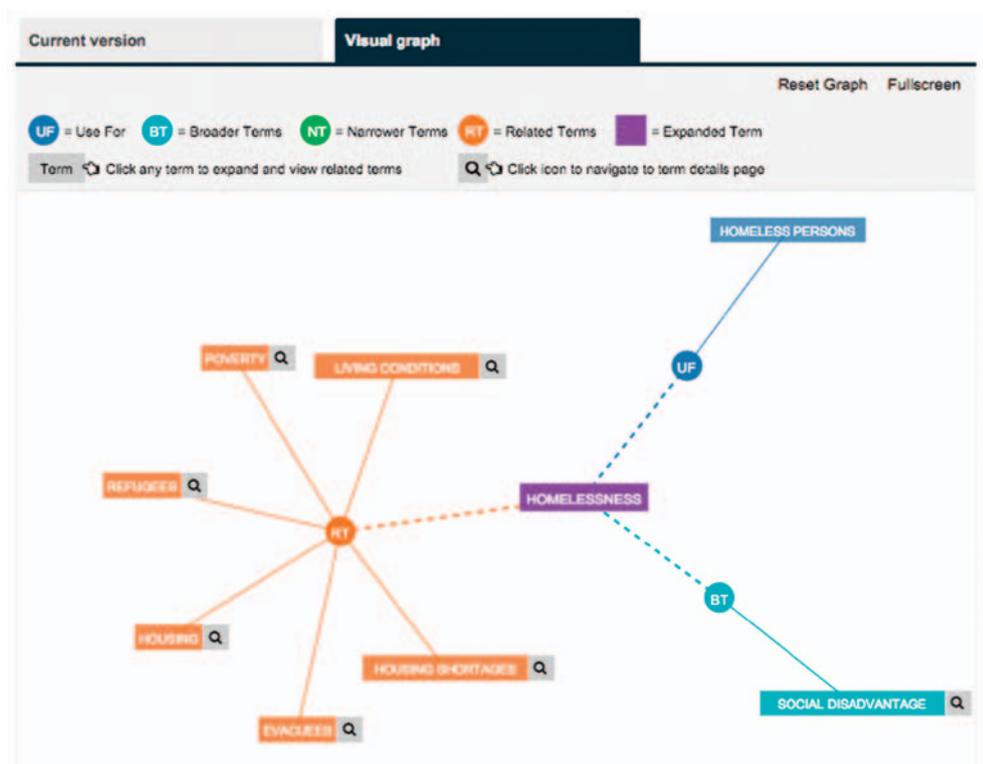
It also involves language differences, something that is critical for social scientists as it can clarify or obscure meaning.

To tackle this issue, our team at Essex are in the midst of a project to enhance access to – and align – the two thesauri that we manage. Humanities and Social Science Electronic Thesaurus (HASSET) is a resource in English. European Language Social Science Thesaurus (ELSST) is a multilingual resource developed over many years, often in association with members of the Consortium of European Social Science Data Archives (CESSDA). It is currently available in ten languages, with more on the way.

Our applications development team have been critical to this development, developing the Simple Knowledge Organization System (SKOS) instances of the thesauri and building a new online thesaurus management application that allows translators to upload translations, visualise the relevant mappings in a tree structure, and make interactive suggestions.

In November 2013, the first release of this new tool was shared with ELSST translators across Europe for the first wave of user testing.

Much work has also been taking place behind the scenes to align the two thesauri. Ideally, wherever the two share concepts they should have identical labels, metadata and relationships. There are times, however, when cultural differences in meaning make that difficult: for example, the concept of ‘right of way’ in the UK may mean something slightly different in Finland.



A view inside the ELSST interface, where related terms are mapped in tree structures.

To bridge this gap, the team have adopted an innovative mapping technique based on (but extending) the ISO 25964 standard, where shared concepts are labelled with either ‘exact’ or ‘close’ equivalents. This is an important but mammoth task: the two thesauri share over 7,500 concepts and, in SKOS terms,

between them represent over 101,000 triples. The aim is to end up with two linked thesaurus tools which will allow researchers across Europe effectively – and accurately – to search for data across languages.

Our involvement with ELSST also encompasses the management of licences and user agreements.

**You can see all language versions of a term at one go, you can see both the source and your language version side by side when you translate, and you can see the translation status.**

Taina Jääskeläinen, Finnish Social Science Data Archive

# Fuelling research that benefits everyone

Data can only make a real difference when they're used to provide insights, influence debates or inform public policies.

Data from our collection have underpinned a number of research projects that are central to key public and policy debates. We've captured more than 150 of them in the form of case studies. These are just a few which use survey and longitudinal data.

## Retirement age for women

Retirement ages are on the rise – but do we know how the labour market will respond? Jonathan Cribb and colleagues from the Institute for Fiscal Studies used Labour Force Survey data to investigate how raising the UK state pension age for women might affect retirement decisions. The findings suggest that increasing the age by one year increased the employment rate of 60-year-old women by 7.3 percentage points; men married to wives aged 60 also saw a significant but smaller rise in employment (4.2 percentage points).

IMPACT: This research was introduced at the Nuffield Foundation in March 2013 to an audience including policymakers at the Department for Work and Pensions. The researchers were subsequently invited to present a seminar to policymakers at HM Treasury in April 2013 and their work was cited by the Office of Budget Responsibility in their Economic and Fiscal Outlook accompanying the 2013 Budget. This work has also been cited in the press and discussed on BBC Radio 4's Today programme.

## Buy-to-let landlords

As house prices rise to record highs, many are forced to rent instead of buy. It's a flourishing market for landlords, but for renters it often means rising rents, additional fees and unsafe conditions. Chris Lord and colleagues from NatCen and the Strategic Society Centre used data from two surveys – Wealth and Assets and the British Household Panel Survey – to investigate the size, financial assets and plans of UK private landlords. Their findings indicate that compared to tenants,

the majority are privileged in terms of education, employment and wealth, with a mean asset base of roughly £75,000. In fact, 62 per cent report that they're able to withstand the economic risk of a quarter drop in income for a year or more.

IMPACT: The evidence has informed the debate on landlord and tenant rights, specifically that policy changes that favour tenants over landlords may have little impact on landlords. This work was widely featured in the media including *The Economist*, *The Guardian*, *The Times* and Sky News.

## Public support for nuclear energy

Energy policies are ripe for public and political debates, as they're entwined with challenging issues including carbon emissions, climate change, environmental impact, financial profits and international relations. Wouter Poortinga and colleagues at Cardiff University joined Japan's National Institute for Environmental Studies to examine public perceptions on both sides of the world before and after the Fukushima nuclear accident, specifically to understand how they may impact current and future energy strategies.

IMPACT: Using nationally representative surveys – two in Japan, two in the UK – they found that the Fukushima disaster did not impact attitudes toward nuclear power alone; it affected public views on Japan's energy future more generally. The Japanese are now less likely to think that any specific energy source will contribute to a reliable and secure supply of energy. This work was published in the journal *Energy Policy* as well as by the UK Energy Research Centre.

*We work regularly with researchers to uncover the innovative ways our data are being used around the world. You can search our library of case studies through Discover.*

# Raising the bar for research data management practices

## Good data management practices lead to quality data.

Quality data can promote sharing and reuse, which can greatly accelerate social science research and teaching at all levels.

With research funders and institutions increasingly pressing for data management plans that promote data sharing, those who collect data are looking for an authoritative source for answers and support. To fill this gap, many UK universities are setting up research data management support services for their researchers. According to website analytics, the source they most often cite is the UK Data Service, its predecessor Economic and Social Data Service (ESDS), and its sister service, UK Data Archive.

In 2013-2014, there have been more than 115,000 unique pageviews for the detailed online support provided at both the UK Data Service and UK Data Archive websites. These web pages have become a 'must-see' for researchers, a place where they can find step-by-step guidance from writing data management plans and gaining consent for data sharing through to documenting, formatting and storing data for long-term use.

The companion print guide, *Managing and sharing data: Best practice for researchers*, now in its third edition, remains highly popular with more than 1,000 copies sent on request – and free of charge – to universities across the UK. More recently, the trainer's resource – a binder packed with case studies and exercises – is supporting institutions to develop their own in-house expertise, creating a multiplier effect in terms of setting high and consistent standards across institutions.

This online and print guidance is augmented with a programme of workshops where researchers get hands-on support with their own data. In 2013-2014 we provided training on request to universities from northwest Scotland to southeast England in addition to 12 in-house events, reaching an audience of 330 researchers, postgraduate students and support staff.

IMPACT: This work is now reaching around the globe. The core guide for researchers has been translated into Japanese by the Social Science Japan Data Archive. Parts of the trainer's pack have been repurposed to guide trainers at the University of Queensland, Australia. The World Agroforestry Centre based in Nairobi, Kenya has reused significant parts of the best-practice guide to train their researchers and publish their own data management guide.



In 2013 the Social Science Japan Data Archive published a Japanese translation of *Managing and sharing data: Best practice for researchers*.



# Publishing our expertise for global impact

The academic publisher SAGE recognised the numbers of researchers turning to our guidance, and decided to give it an even wider audience.

In 2014 they published an expanded and updated version of our *Best practice for researchers* guide in handbook form.

*Managing and sharing research data: A guide to good practice* is a 240-page compendium of advice, examples, case studies and exercises covering the essentials of good research data management practices (formatting, organising, storing, transferring, publishing), plus specific guidance on privacy in data sharing, intellectual property rights, and data citation.

The authors – Louise Corti, Veerle Van den Eynden, Libby Bishop and Matthew Woollard – have compiled and improved the guidance over decades of working directly with UK researchers, funded by the Economic and Social Research Council (ESRC), Jisc and other UK research councils.

“Researchers need to be armed with the knowledge and skills to ensure that the data they create and manage can be exploited to the maximum potential for further research,” says Paul Boyle, outgoing Chief Executive of the ESRC. “This book offers these skills in an approachable way.”

Furthermore, having a publisher like SAGE acknowledge this vital and under-taught topic should help it enter the mainstream research methods literature that makes up standard academic reading lists.

“[The book’s] great strength is the way it combines thought-provoking practical exercises, exemplars and checklists for action, and case studies with the rationale for action on each topic, addressing potential concerns and pitfalls,” says Jude England, Head of Social Sciences for The British Library.

“This is a much-needed resource that will serve the field well,” adds Nigel Fielding, Professor in Social Research Methodology at the University of Surrey.

The book is available in hardback and paperback and can be ordered online.

“ This is a timely, comprehensive and really useful guide to good practice in the management and sharing of research data. ”

Jude England, Head of Social Sciences,  
The British Library

# Bringing real-world data into the curriculum

## Data aren't just for researchers.

Students are increasingly being introduced to research data to help develop the skills needed to analyse and evaluate evidence. Teachers are also discovering that real-world data can make their courses come alive with relevance, both in and out of class.

"The employability agenda is very much pushing research methods to the fore," writes Jackie Carter, Mark Brown and Bethany Morgan Brett in a recent issue of the journal *Health and Social Care Education*.<sup>1</sup> "By engaging with research, students learn a variety of transferable skills such as group work, critical appraisal, and report writing... [And] the UK Data Service provides a highly interactive learning environment in which students can discover and interrogate the wealth of real-world data hands-on."

### A source for data-driven teaching

There are a growing number of datasets in our collection designed specifically for teaching and learning. These files are designed so they're easier for learners to engage with, containing a manageable number of variables. They include cut-down datasets from major surveys as well as qualitative teaching files.

Our website is also becoming a hub for data-driven teaching resources. A suite of web pages points teachers to teaching datasets as well as online teaching modules, hands-on workbooks and dataset guides.

Those working with qualitative data, for example, can find detailed guidance on interviewing methods as well as thematic guides ranging from gender studies to youth culture to crime and social control. Teachers using quantitative data can find videos on how to analyse international macrodata using UKDS.Stat and Nesstar.

They can also find supporting materials for software and other online tools such as SPSS, Stata and R. Collectively these resources assist in both methods and substantive teaching across the social sciences.

### Supporting the network of teachers

Increasingly, the UK Data Service is helping teachers support other teachers. We have 32 case studies on the website that demonstrate how data from the collection are making a difference across the curriculum – a valuable resource that can inspire teachers and encourage reuse of both data and teaching materials. Some examples:

- Getting to grips with statistics: Jane Fielding, University of Surrey
- Understanding the global economy using real-world data: Danny Quah, London School of Economics and Political Science
- Connecting data to subjects that inspire students: Mark Brown, University of Manchester
- Getting sociology students into archived qualitative data: Jackie Turton, University of Essex
- Helping students design their own day analysis projects: Melanie Beres, Otago University, New Zealand

"The UK Data Service makes it easy to access data on almost every imaginable debate in contemporary social science", says John MacInnes, a sociology professor at the University of Edinburgh.

"'Real data' are not only more attractive to students, they also give them a much better understanding of how research is done, and the difference between evidence and armchair philosophy."

<sup>1</sup> Carter, J., Brown, M. and Brett, B. M. (2014) 'Creative teaching in health and social care using the UK Data Service', *Health and Social Care Education*, 3(1), pp. 16-22

# Opening and improving access to census data

Our Census Support team have taken a major step in removing unnecessary barriers to data access.

In April 2014, they provided free and open access to aggregate statistics from the 2011 Census via InFuse, an online tool which allows users to build queries and extract the data they need. InFuse provides academic and non-academic researchers with access to 2001 (England and Wales) and 2011 (UK) Census aggregate statistics under the Open Government Licence. This means there's no need to register or login to extract the data. InFuse also makes it quick and easy for users to find and download information for places of interest across the UK.

## Easier to access and use

Rather than making the aggregate statistics data available to users in the raw tabular formats supplied by the national statistical agencies, the Census Support team have made the data more accessible and easier to locate and extract than through other interfaces.

Consequently, InFuse operates from a fundamentally restructured and integrated version of the 2011 aggregate statistics. It integrates and standardises data and geographies to enable unique features including:

- searches for data across the whole collection by selecting combinations of variables rather than searching through hundreds of traditional census tables
- selection of data for equivalent geographical areas across the UK in a single operation rather than separate operations for different countries

The three UK census agencies are producing thousands of constituent datasets and it is estimated that the InFuse collection will eventually contain around ten billion counts relating to two million areas arranged in multiple geographical hierarchies.

Characteristics will be defined in terms of over 100 topics divided into several thousand categories. Counts will be produced for several hundred thousand different combinations of categories. Data for many of these combinations of topics and categories will only be available for some of the areas.

## No more registration for other Census tools

Changes have also been made to our data access tool for historical censuses, Casweb, allowing any user to have open access to the UK Data Service back catalogue of aggregate data from the 1971, 1981, 1991 and 2001 Censuses. In addition, users requiring origin-destination statistics or flow data from 2001, 1991 or 1981 censuses can now extract data from WICID by logging in using a guest account without the need to register for access to census statistics.

Work is underway to extend open access to as many of the census geographical boundary datasets as possible in the near future.

**// InFuse is a fantastic tool that saves plenty of time and effort!**

Webinar participant Max Berks,  
Centre for Workforce Intelligence, 13 February 2014



# Providing more options for international data users

Researchers who depend on international macrodata can access some data directly from intergovernmental organisation websites.

However access is often limited to certain data, with no way to search for data across providers.

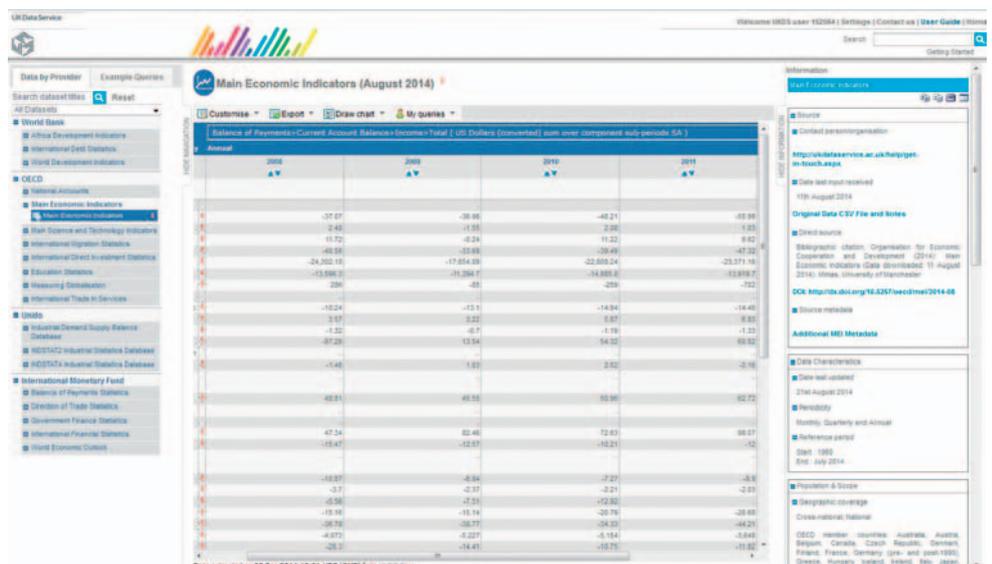
Now there's UKDS.Stat, our new cutting-edge data platform that allows academic users to search across the full content of datasets from a range of different sources, including World Bank, UNIDO, International Monetary Fund (IMF) and Organisation for Economic Co-operation and Development (OECD). Users can also:

- combine data, to view and download similar data from different datasets
- save data searches as queries, which can be re-run to search new data releases
- share queries with collaborators

In addition, it offers an impressive selection of visualisation options including animated time-series charts and maps, enabling users to easily see how the data change through time and across geographies.

Early feedback indicates this is a major step forward. "I prefer it because it allows users to customise the layout – particularly, to download data in 'long' format," says an economics lecturer from University of Bath.

The full international macrodata portfolio of more than 50 datasets is currently being ingested into the UKDS.Stat platform, set to replace the Beyond 20/20 platform by December 2014.



# Widening access to an eLibrary of financial resources

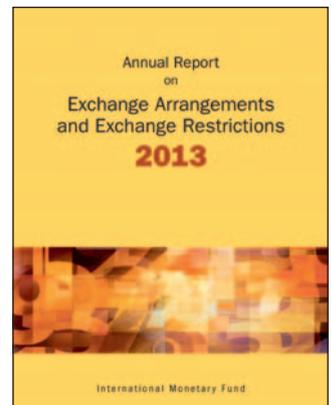
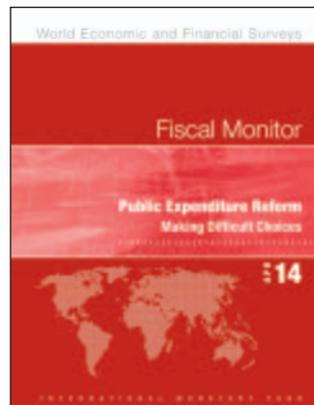
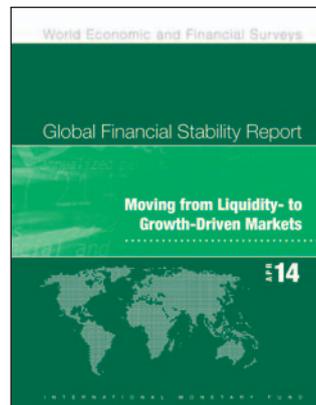
In a move to support wider access to research-related publications, we have opened a portal for academic users to find, access and download documents published by IMF.

Thanks to special funding from the Department of Business, Innovation and Skills, users in higher and further education can explore and use the IMF's vast eLibrary, a treasure trove of more than 10,000 digital documents dating from 1951 to the present.

The catalogue includes periodicals, books and working papers that provide information and perspective on macroeconomics, globalisation, development, trade and aid, technical assistance, demographics, emerging markets, policy advice and poverty reduction. Examples include the *Global Financial Stability Report*, *Fiscal Monitor*, *IMF Survey* and *Research Bulletin*.

And these documents are all available in a range of digital formats for mobile access, including PDFs, iPhone and Android (for smartphones), ePub (for iPad, Sony, Nook and other devices), and Mobi (for Kindle readers).

You can find the IMF eLibrary at [imf-text.mimas.ac.uk](http://imf-text.mimas.ac.uk)



# Reaching new users through training videos

Learning is no longer a strictly classroom affair.

As new technologies open the door to effective online training, we are adapting many of our core materials to electronic formats – webinars and video tutorials – that reduce travel time, lower costs, and improve convenience.

Webinars and video tutorials are fast becoming a popular choice, helping data users get the information and support they need, right from their desktops.

Perhaps more importantly, these formats are already reaching new and more users and showing promise in breaking down barriers between academic and non-academic communities.

The most popular webinars for non-academic audiences are those that introduce the UK Data Service and its core data – most particularly from the 2011 Census other major surveys – which suggests these are new users who are getting to grips with the resources available through the UK Data Service. One introductory webinar on international databanks prompted an immediate spark in online activity: following its live debut on 15 January 2014, the Service saw a noticeable jump in both pageviews and unique pageviews.

Building on this success, in late 2013 we made past webinars easier to find via a website library and YouTube channel.

We also introduced a series of openly licensed video tutorials that talk users through specific concepts and tasks, often with online demonstrations. These range from commonly requested information (*How to download UK survey data*) to specialist training for advanced data users (*Creating an intensity map using Google Fusion tables*). They're also easily found alongside the data catalogue in Discover.

The response has been swift and positive, with compliments coming in via Twitter and focus group discussions. They're not just popular in the UK: Open Education Europa, an online portal for educational resources across the EU, quickly indexed the tutorials and promoted them to its members.

Of the 2,392 who attended our training events in this reporting period, the greatest percentage attended via webinar, outpacing our face-to-face workshops.

User meetings 21%

Workshops

37%

Webinars

42%

# Shaping and influencing UK data strategies

Providing strategic advice to inform data strategies across the UK is core to our vision, especially for the social sciences. In the last 18 months, we have taken some long strides in this area.

Our guidance is grounded in the philosophy that barriers to access should be proportionate to the risk of disclosure, or real commercial sensitivity. We also believe that data access strategies are not just about access, they must also be about data quality, integrity, 'findability', long-term curation, and the support mechanisms around data. We persistently and consistently approach these topics while considering the best methods of access.

We believe, first and foremost, in making data available to researchers who need them, and making them available in a timely fashion. However we have to follow the law, and we have to meet the needs of the legal owners of data which we hold, and provide access to, on their behalf.

In the last 18 months, discussions with the Office for National Statistics have led to a single, shared concordat. We expect this framework to extend to other large data suppliers. A uniform approach based on our new Access Policy makes the Service more efficient, and makes decisions about access criteria easier.

We advise the ESRC – through numerous channels – on the implementation of their Research Data Policy, and the wider Research Councils UK (RCUK) activities on research data. We have provided detailed input into the National Data Strategy, responded to several relevant government strategies on data (notably the Open Data consultations from both the Cabinet Office and the Public Accounts Select Committee). We've provided separate input into the ESRC's initiative on the new EU Data Protection Legislation.

In addition, our staff sit on cross-council bodies like the Expert Advisory Group on Data Access and various Data Access Committees. In each case the message is the same: barriers to access must be proportionate to the risk of disclosure. Data where there is no possibility of the disclosure of personal information should be open. Users of data must be traceable where there is a residual risk of disclosure through linkage with other public sources of information. Finally, researchers who are allowed to use personal data must be appropriately trained and vetted.

**// We are in almost constant dialogues with our data owners to widen access, or reduce barriers to access.**

Matthew Woollard  
Director, UK Data Service



# Accelerating sustainable data services in Europe

What do you do when your nation has rich reserves of social science data – but they’re not adequately inventoried, archived or shared?

This is the situation that some European nations are finding themselves in, with data assets held across university departments and institutes in varying formats and conditions. Three countries – Bosnia and Herzegovina, Serbia and Croatia – have made a commitment to establish archives to address this important need, but time is pressing: they need infrastructure, standards and procedures in order to accelerate research access and use.

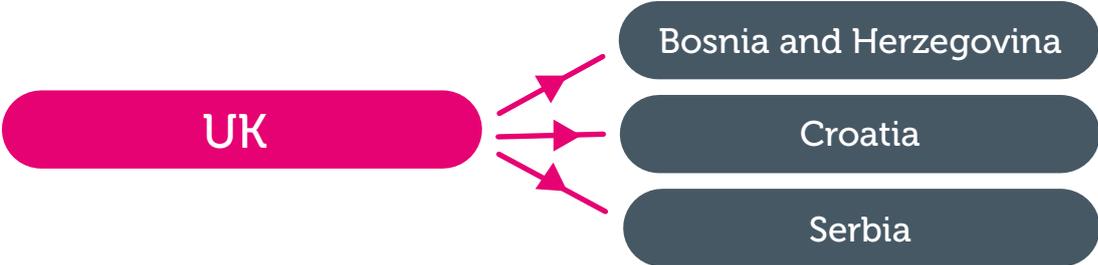
The SERSCIDA project was set up with funding from the European Commission to bridge this gap by harnessing existing expertise in support of these countries. With roadmaps written by each emerging archive, we worked in partnership with other members of the CESSDA to help them quickly establish formal archives with strong foundations.

Our support included an intensive three-day workshop modelled on the UK Data Archive’s highly regarded ‘How to Set Up and Run a Data Service’ workshops. There, the West Balkan archivists learned the essentials, starting with the structure of a data archive and its functional components using the OAIS model, and then identifying the necessary documentation and procedures for each component.

To augment this introduction, we then formalised our training materials and documentation into a practical online guide that’s comprehensive, cohesive, and easy to navigate. All resources are supplemented with links to similar documents at CESSDA partner sites to demonstrate flexibility in approach. The guide also highlights how each data service is unique – especially true in this case as each is at a different stage of development, with different governance structures and different ways of doing things.

This resource is already proving fruitful. “I finally have a complete picture of the need and importance of establishing institutions which would allow open access to and dissemination of primary data collected through different surveys,” says Sunčica Stefanović Šestić, Head of the Statistical Department for Education, Scientific and Culture in Serbia.

Colleague Dalibor Pančić in Bosnia and Herzegovina has already used what he’s learned: “The knowledge and skills I have gained in this project and with the current technical infrastructure of [my institution], we have created some of the most significant pre-conditions for the establishment of a digital data archive at the National and University Library of the Republic of Srpska.”



## Improving EU archiving standards

Effective research depends on quality data that are easy to find and use – no matter where in the world they are.



Those who seek data from European sources, however, often encounter a patchwork quilt of nations and repositories, each in a different language and with varied organising principles and archiving standards.

We're rising to this considerable challenge by working with partners in the CESSDA to create the standards and processes that will allow seamless access.

In 2013 staff from the UK Data Service led a group of experts in a series of workshops with CESSDA partners to look specifically at how best to integrate standards and processes to establish a set of 'best practices' across European archives.

Using the Data Seal of Approval (DSA) to structure the discussion, the group self-assessed and peer-reviewed each other against the 16 trusted digital repository criteria. In this way they were able to build consensus around what works best along with the cooperation necessary to align systems across organisations.

The result: a range of recommendations ranging from aligning approaches to creating preservation policies to developing a common description of archival storage practices. These recommendations are now under consideration by the CESSDA board, with an aim to establish a common standard for trusted digital repositories and a formal approach to managing CESSDA partners' alignment with the CESSDA statutes.

## Shining a light on the costs of curating digital data

No one would argue that it's not important to invest in curating and preserving digital data for long-term use.

But given the size, complexity and variety of most collections, it's often difficult to identify where to direct limited funds to ensure the best return on investment.

We're working with the 4C Project to shine a light on the investment issues facing European organisations. 4C stands for Collaborating to Clarify the Costs of Curation, which essentially summarises the group's forward-looking mission.

"We're leading a task that's identifying and defining core cost concepts around resources and activities and their interactions with issues critical to stakeholders such as 'risk', 'value', 'quality' and 'sustainability'," explains Hervé L'Hours, Preservation Planning Manager.

Standardisation of underlying concepts and approaches will both allow organisations to select cost models appropriate to their needs and support the wider digital curation community by making the outputs of these curation cost methods more directly comparable.

"There is further work to do to adequately join up existing models and to define new ones that will help to make sense," says task member Ulla Bøgvad Kejser from the Royal Library, Denmark, "but this work promises to provide a more coherent perspective on the economics of digital curation."

# Collaborating for the good of global statistics

We may have 'UK' at the start of our name, but our data and influence reach far beyond national borders.

International macrodata have long formed a core part of the data collection, serving an active community of academic users. Now the UK Data Service is building on that reputation by becoming the first and only academic institution to join the worldwide Statistical Information System Collaboration Community led by the OECD.

Under a memorandum of understanding, our team at Mimas, University of Manchester, are collaborating with a number of inter-governmental and national statistics organisations – including the International Monetary Fund, European Commission and UNESCO – to develop statistical systems for aggregate data and to promote established standards for data exchange to improve data infrastructures. Members meet quarterly, attend developer workshops and specialist task forces, and contribute to a portal for reporting issues and sharing code.

This work is already having an impact: for example, all members of the Collaboration Community have implemented DotStat to disseminate data to their respective user groups and are working together to co-develop the DotStat software. This should lead to greater efficiency in using and sharing data between statistical organisations, which in turn should lead to innovative technical developments for the DotStat data loading tools and browser.

"Our team adds an important voice to the group," says Keith Cole, Deputy Director for the UK Data Service, "as we're able to highlight academic researchers' needs in using international data."

**// This collaborative approach allows member organisations to use established standards and modern technologies to significantly improve their data infrastructures and also to benefit from each other's extensive experience. //**

Keith Cole, Deputy Director, UK Data Service

# Looking ahead

As this report shows, we've come together and put in some intensive work to establish and grow the UK Data Service.

The results may be impressive, but we're far from finished. Many major new initiatives are just around the bend – too new to report outputs or impact, but not too small to mention. Here's a glimpse of some of the exciting work underway.

## Supporting ESRC Big Data initiatives

In the coming years we look forward to supporting the new Research Data Centres by increasing capacity within the UK Data Service to work on aligning data purchase and access approaches for new types of data. We will collaborate with the Digital Preservation Coalition on two linked projects to investigate best practices for curating new forms of data, especially transactional data and social media data.

## Promoting our Open Data portfolio

A key element of the Service's mission is to widen access to data, and one of the key ways to do this is to persuade data owners that existing access conditions may not be appropriate.

In the next year we'll be undertaking a programme of examining previous licences for our holdings to advise, where applicable, which data might be made more openly available. We'll also be working with the Open Data Institute to certify our Open Data collections to offer greater visibility to researchers beyond academic communities.

## Administrative Data Research Network

In October 2013 the ESRC invested in this major new initiative as part of its £64 million Big Data Network, and asked us to play a leading role in coordinating its four national research centres. The Administrative Data Research Network is now a fledgling service that offers researchers in all four UK countries an unprecedented gateway to find, access and link to a rich and growing source of data: de-identified administrative data routinely collected by government departments and other public organisations.

Our work to build this network will draw on in-house expertise in producer relations, data quality standards, secure access and disclosure control. In addition, as concerns over data privacy grow, our work to inform and engage the public will be critical to establish the trust necessary to negotiate more widespread use (and reuse) of administrative data.

## Innovation Fund

Thanks to this incubator fund from the ESRC, seven external partners are now working side-by-side with staff to enable new ways to find, access, share, link or otherwise exploit the value of research data.

Nine projects were pitched at a Dragon's Den-style event in February 2014; seven were 'green-lighted' for a three-month pilot phase. Bids came from businesses as well as academics; a few were joint applications. In July the Innovation Fund panel will evaluate progress and opportunities and will fund up to three projects for full development. Exciting times ahead; we look forward to sharing news in the next annual report.

## QualiBank

Searching for qualitative data has complications and nuances not common to quantitative data. When it comes to data in audio or image files, it becomes even trickier.

A team of staff specialists have been developing QualiBank, a search-and-browse interface that offers unprecedented capacities for qualitative and mixed methods collections. It searches the content of text files such as interviews, essays, open-ended questions and reports. It also searches related metadata, such as descriptions of images and audio recordings, and enables hyperlinking to related objects. Perhaps most impressively, it allows users to cite an entire object (such as a study) or just extracts (such as a paragraph).

At the end of the reporting period, QualiBank was in beta format for testing before a major public release.

## ReShare

Researchers who collect, manage and archive their own digital data now have a new self-deposit repository that's state of the art.

ReShare uses an extensively customised version of EPrints that's more suitable for complex data collections. The upgrades and benefits include an easy-to-use depositor interface; an intuitive workflow which simplifies deposits of multiple files, types and related documentation; an expanded metadata profile and vocabularies based on standard schemas; and free Digital Object Identifiers (DOIs) attached to each deposit, allowing researchers to cite and track their own data collections.

The ReShare project is funded by the ESRC to replace the current ESRC Data Store; all collections currently archived there are being upgraded and validated to move to the new system. We expect to further open up this self-deposit system to researchers beyond the ESRC who currently have nowhere to share their social science data.

# Who's who

The UK Data Service is delivered by staff with special expertise in research data based at universities across the UK.

These are the organisations that provide the foundation for our data, resources and services.

## UK Data Archive, University of Essex

Leader 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 is also the host of the Administrative Data Service.

## Mimas, University of Manchester

Hosts a significant number of the UK's research information assets, and builds applications to help people make the most of these rich resources. Provides access to, and specialist support for, the databanks of the International Monetary Fund, and also aggregate statistics from the 1971 to 2011 UK censuses.

## Cathie Marsh Centre for Census and Survey Research, University of Manchester

A centre of research excellence specialising in the application of advanced quantitative methods in an interdisciplinary social science context. Since 1993, CCSR had a role in the support and development of UK secondary microdata including the Sample of Anonymised Records (SARs) from the UK Census.

## School of Geography, University of Leeds

The School has a longstanding 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 Environment, University of Southampton

The department's work comprises: spatial population analysis and modelling; cultures, spaces and practices of care; and population health. Members work closely with the Office for National Statistics and co-direct the ESRC National Centre for Research Methods.

## EDINA, University of Edinburgh

Acts as the ESRC specialist geography unit for the UK Census, supporting UK Data Service users in accessing and using the geography outputs of the 2011 and previous censuses.

## 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.

# Governing Board

---

The Governing Board has the responsibility and authority to ensure that the UK Data Service is developed, managed and maintained in a manner that maximises its benefit as a long-term, world-class data resource.

## Chair

**Gordon Marshall**  
Leverhulme Trust

## Vice-Chair

**Simon Briscoe**

## Members

**George Alter**  
ICPSR, University of Michigan

**Richard Alldritt**  
UK Statistics Authority (2012-2013)

**Fiona Armstrong**  
Economic and Social Research Council

**Micheline Beaulieu**  
University of Sheffield (2012-2013)

**Keith Brown**  
University of Manchester

**Rachel Bruce**  
Jisc

**John Curtice**  
University of Strathclyde

**Keith Dugmore**  
Demographic Decisions

**Peter Elias**  
University of Warwick

**Nigel Fielding**  
University of Surrey

**Todd Landman**  
University of Essex (2014)

**Paul Longley**  
University College London

**David Marshall**  
Northern Ireland Statistics and Research Agency (2014)

**Andrew Morris**  
University of Dundee

**David Sanders**  
University of Essex (2012-2013)

**Glen Watson**  
Office for National Statistics (2012-2013)

**Sharon Witherspoon**  
Nuffield Foundation

## Ex-officio

**Christopher Dibben**  
University of St Andrews



ukdataservice.ac.uk



---

UK Data Service

---



Funding provided by

