Case study of a data archive: The UK Data Service

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

Workshop on Research Data Management
Conference ‘Changing Landscape of Science & Technology Libraries’
Indian Institute of Technology Gandhinagar
2-4 March 2017
What is the UK Data Archive

- Department of the University of Essex
- Established in 1967 by the Economic and Social Research Council (ESRC) as a ‘Data Bank’ for science
- 50 years of curating and providing access to data
- Data and support services for research, teaching and learning
- Speciality in social survey data, qualitative data, historical databases, some biomedical and coming…big data!
- Leads the ESRC UK Data Service
University of Essex

The Archive
What does the UK Data Service do?

- Put together a collection of the most valuable data
- Make data and documentation available for reuse
- Preserve data in the long term for future research
- Provide data management advice for data creators
- Provide support for users of the service
- Information about how data are used
- Easy access through website
Progressive data sharing in the UK

- Concordat with our National Statistical Institute
- Supporting the ESRC Data Policy since the 1960s
  - Role in policing the policy – compliance /data quality
  - Advise and support for applicants and award holders
  - Included qualitative and historical data from 1996
- First to publish RDM guide for researchers in 2005
- Our approach to making data shareable based on:
  - everyday challenges faced by researchers
  - experience in handling range of research data
Welcome to the UK Data Service

Your resource for quality social research data

A unified point of access to data from ESDS, Census Programme, Secure Data Service and others

Call for papers: Opinions and Lifestyle Survey user meeting

European Social Survey invites EU data users to visit and learn

Digital Futures: Your input needed on data digitisation

Standards: Data sharing and the role of interoperability

The UK Data Service is at the core of a network of trust that includes data owners, producers, funders and users.

Who can most benefit from the data we hold?

- Academic researchers and students
- Government analysts
- Charities and foundations
Our data portfolio

**UK Surveys**
- Large-scale government funded surveys

**Longitudinal**
- Major UK surveys following individuals over time

**International**
- Multi-nation aggregate databanks and survey data

**Census**
- Census data 1971 – 2011

**Business**
- Microdata and administrative data

**Qualitative**
- Range of multimedia qualitative data sources
Who is it for?

• academic researchers and students
• government analysts
• charities and foundations
• business consultants
• independent research centres
• think tanks
• citizen scientists, where skills enable analysis
Links with other data archives worldwide

Data archives
- Qualitative data
- Open data
- Secure datalabs
- Question banks

Explore online
- Data access policy
- Thesauri

ICSSR Data Service
Some statistics about our UK Service

Data for research and teaching purposes, used in all sectors and by many different disciplines

- **7,064** datasets in the collection
- **1008** qualitative and mixed methods collections
- **400** new datasets and new editions added within last 12 months
- **25,000** registered users
- **60,000** downloads worldwide per annum
- **4000+** user support queries per annum
Sharing data at the UKDS

Discipline specific repository

We use 2 ‘curation pathways’:

• Our ‘curated collection’ - high impact data, e.g. government departments, major surveys

• UK Data Service ReShare – self-deposit system for academic and smaller research data
UK Data Archive - digital data preservation

- Undertake long-term data curation and preservation
- Certified to ISO27001 for Information Security
- Data Seal of Approval (DSA) accredited
- Deeply involved in international preservation planning and accreditation activities

www.data-archive.ac.uk/curate
Data collection requisites

We ensure that data collections are:

- Long-term preserved and usable
- Self-explanatory for users
- Non disclosive where promised
- Rights are in place to redistribute
Collections Development work

Trawling

Line-catching
UK Data Service acquisition

- **Reactively** and **proactively** acquire data
- Data are deposited by:
  - UK government departments (contractual)
  - Intergovernmental organisations
  - Research institutes
  - Charities and companies
  - Individual researchers - ESRC Data Policy
- **Selection criteria:** Collections Development Policy
- Beware spontaneous gifts…
I have data! But I also need to empty my office......
Assessment for new deposits

Data Appraisal Group assesses data
Decision will usually be one of the following

- accept the data into our main collection
- deposit in ReShare repository
- deposit in an institutional repository
- use an alternative place of deposit
# Data Appraisal Grid

<table>
<thead>
<tr>
<th>STAGE 1: Strategic high-level value</th>
<th>Yes</th>
<th>No</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK Data Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKDS: Potential secondary use and analysis for research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKDS: Teaching and learning use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKDS: Replication and validation of research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data relevant to the UK Data Forum Data Strategy?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhanced longitudinal data on people and families</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addressing multidisciplinary challenges (e.g. biosocial data)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data about organisations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer databases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social media as research data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-sectional data on people and families</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Access conditions

Depositor selects, with guidance, the access category most appropriate for the data, under a license.

<table>
<thead>
<tr>
<th>Access Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Open</strong></td>
<td>• available for download/online access under open licence without any registration</td>
</tr>
<tr>
<td><strong>Safeguarded</strong></td>
<td>• available for download/online access to logged-in users who have registered and agreed to an End User Licence</td>
</tr>
<tr>
<td><strong>Controlled</strong></td>
<td>• available for remote or safe room access registered users whose research proposal has been approved by an access committee and who have received specialist training</td>
</tr>
</tbody>
</table>
Safeguarded data – conditions of access

• Most common license choice
• Register with us - UK Federation
• Agree to an End User Licence (EUL)
  ✓ Appropriate data usage
  ✓ Full citation of data
  ✓ Informing us of reuse outputs
• Click ‘Download/Order’ from catalogue
• Specify a project for which the data are to be used
• Download data in choice of formats
Data discovery

Discover

Search and browse our data collections, support guides, case studies, and related publications.

Search: India

Results per page: 10

Displaying 1-10 of 198 results

SN 852597 New urbanisms in India: Urban living, sustainibility, everyday life
Sophie Hadfield-Hill, University of Birmingham

Full record

Download | DDI XML | Similar data collections

SN 7478 Young Lives: School Survey, India, 2010-2011
Solon,A., University of Oxford. Department of International Development

Full record

Download/Order | DDI XML | Similar data collections

SN 5380 Resisting Subjugation: Law and Power amongst the Santal of India and Bangladesh, 2002-2004
Shariff, F., University of Warwick. School of Law

Full record

Download/Order | DDI XML | Similar data collections
Quality assessment of data

• Assess data, metadata and documentation:
  • Data integrity
  • File, variable and value label metadata
  • Documentation coverage

• Set publishing standard - depends on likely use
  • Online data browsing tools requires high level of data enhancement

• Data anonymisation: disclosure review
UKDS strategies for enabling safe access

• **Informed consent** for long-term data sharing
• **Protection of identities** when promise
• **Regulated access** where needed (all or part of data) e.g. by group, use, time period

5 SAFES enables **safe access to data** that meet the needs of **data protection**. Yet fulfils the demands for open science and transparency
Protecting confidentiality: the ‘5 Safes’

- **Safe data** - treat data to protect respondent confidentiality
- **Safe people** - educate researchers to use data safely
- **Safe projects** - research projects for ‘public good’
- **Safe settings** - SecureLab system for sensitive data
- **Safe outputs** - SecureLab projects outputs screened

5 Safes Animation
Assessment complete

- Processing level allocated (A*, A, B, C) - 30 days
- Data enhancements: edits made in consultation with depositor
- Generate multiple data formats for dissemination and preservation
- Assemble documentation for users
- Prepare catalogue record (DDI)
- Archived to preservation system
- Released as zip bundles to catalogue
- DataCite DOI assigned - versioned
Health Survey for England, 2014

SN: 7919
Title: Health Survey for England, 2014
Alternative title: HSE
Persistent identifier: 10.5255/UKDA-SN-7919-2
Series: Health Survey for England [Health Survey for England, 1991-]
Depositor: NatCen Social Research
Principal investigator(s): NatCen Social Research
University College London. Department of Epidemiology and Public Health
Sponsor(s): Information Centre for Health and Social Care

CITATION

The citation for this study is:

SERIES ABSTRACT

The Health Survey for England (HSE), sponsored by the Information Centre for Health and Social Care and the Department of Health, began in 1991 and has been carried out annually since then. A number of core questions are included in every wave but each year's survey also has a particular focus on a disease or condition or population group, which are subsequently revisited at appropriate intervals in order to monitor change. The survey combines questionnaire-based answers with physical measurements and the analysis of blood samples. Blood pressure, height and weight, smoking, drinking and general health are covered every year. An interview with each eligible person in the household is followed by a nurse visit.

DATA ACCESS

<table>
<thead>
<tr>
<th>SN</th>
<th>Study Description</th>
<th>Access Online</th>
<th>Download / Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>7919</td>
<td>Health Survey for England, 2014</td>
<td>🔄</td>
<td>📦</td>
</tr>
<tr>
<td>7649</td>
<td>Health Survey for England, 2013</td>
<td>🔄</td>
<td>📦</td>
</tr>
<tr>
<td>7480</td>
<td>Health Survey for England, 2012</td>
<td>🔄</td>
<td>📦</td>
</tr>
<tr>
<td>7260</td>
<td>Health Survey for England, 2011</td>
<td>🔄</td>
<td>📦</td>
</tr>
<tr>
<td>6986</td>
<td>Health Survey for England, 2010</td>
<td>🔄</td>
<td>📦</td>
</tr>
</tbody>
</table>
Sets out any what we did & any key data issues for users
Online browsing for surveys: Nesstar

Dataset: British Social Attitudes Survey, 2009

Variable IncomGap: R say that the gap between those

<table>
<thead>
<tr>
<th>Values</th>
<th>Categories</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>too large</td>
<td>1797</td>
<td>79.3%</td>
</tr>
<tr>
<td>2</td>
<td>about right</td>
<td>360</td>
<td>15.9%</td>
</tr>
<tr>
<td>3</td>
<td>too small</td>
<td>51</td>
<td>2.2%</td>
</tr>
<tr>
<td>8</td>
<td>Don't know</td>
<td>58</td>
<td>2.6%</td>
</tr>
<tr>
<td>9</td>
<td>Refusal</td>
<td>1</td>
<td>0.0%</td>
</tr>
<tr>
<td>-7</td>
<td>off route due to corrupt sample file</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>Skip, version C</td>
<td>1148</td>
<td></td>
</tr>
</tbody>
</table>

SUMMARY STATISTICS
Valid cases 2267
Missing cases 1154
This variable is numeric

UNIVERSE
VERSIONS A AND B: ASK ALL
Online browsing for qualitative - QualiBank

Quali Bank

Search and browse qualitative surveys, interviews and open-ended questions.

Search: typhoid

SEARCH RESULTS

Displaying 1-10 of 17 results

1. Interview with Mrs. Omison
   SN2000 The Edwardians, 1870-1973
   Sex: Female, Age group: 75-84, Socio-economic status: Routine, Region: North West
   ... very little? No, I don't think so. I think we'd be getting on, about 7 or 8 before we went on our own. Did you share the bedroom with anybody? With the maid. And then I had typhoid fever, once, and then the maid had to go and sleep at home. She had her family at Plank Lane, but she didn't use to...

2. Interview with Mrs. Clark
   SN2000 The Edwardians, 1870-1973
   Sex: Female, Socio-economic status: Semi-routine, Region: Wales
   ... lost her you see, she - she got typhoid fever and was - called home. Yes. We were only in England about seven years and then she was - she left us. Mm, it was sad because we - I suppose if - father and mother had gone back - to their work - at the other side of the world, missionary work, I suppose she...
Online browsing for variables & questions
Self upload – supporting ESRC award holders

- Upload data to our ReShare EPrints repository
- Harvests project information from Funder Gateway
- DataCite DOI assigned
- Records feed into central data portal
- Supports Nature’s Scientific Data journal as a host repository
Easy to publish and upload data

Edit collection: Data Collection #CS 5.5

To deposit a data collection, you must accept the ReShare Terms and Conditions:

- I confirm that I am the owner of the copyright and associated intellectual property rights in the whole Data Collection or am otherwise lawfully permitted to grant this licence on behalf of such and every owner;
- I grant a non-exclusive, royalty-free license to the UK Data Archive (a department of the University of Essex and not a separate legal entity) at Minford Park, Colchester, CO4 3SQ (the "University") to hold, make copies of, and disseminate copies of the Data Collection in accordance with the access conditions I will specify when uploading data files: open data accessible to users without registration or embargoed data accessible to users registered with the data services provided by the UK Data Archive.

I agree to the ReShare data deposit terms and conditions

Save for later  Cancel  Next >
**Collection period:**
1 October 2008

**Date from:**
30 September 2012

**Date to:**

**Country:**
United Kingdom

**Data collection method:**
Collection of information from official documentation across several countries

**Observation unit:**
Organisations, Text units

**Kind of data:**
Alpha-numeric

**Type of data:**
Qualitative and mixed methods data

**Resource language:**
English

---

**Access and Administration**

**Data sourcing, processing and preparation:**
Information from various sources, including the International Bureau of Fiscal Documentation

**Copyright holders:**
Name: Devereux, Michael
Email: Unspecified
Affiliation: University of Oxford
ORCID: Unspecified

**Contact:**
Name: Devereux, Michael
Email: michael.devereux@sbs.ox.ac.uk
Affiliation: University of Oxford
ORCID: Unspecified

**Notes on access:**
None

**Publisher:**
Economic and Social Research Council

**Last modified:**
28 Apr 2014 22:42

---

**AVAILABLE FILES**

**Data**

* CBT_Tax_database.xlsx

**Documentation**

* CBT_Tax_database_description.pdf
Self deposit: handling queries on deposit

- Some hand-holding for depositors prior to upload
- Full-time repository administrator (junior researcher with MA in political science)
- 1 in 10 questions relayed up to more senior staff
  - Ethics and disclosure review; more technical issues

- Query tracking system to manage/log responses
  - Can see past queries and responses
  - SLA – mostly answer within 3 working days
  - Easy to add common issues to your FAQ
User support and resources

- Help desk, individual user support
- Promotional events/ workshops/webinars
- User guides/ thematic guides (link from catalogue records)
- Data Impact blog
- Case studies of use
- Teaching & learning resources
- Undergraduate student dissertation prize
Prepare and manage data

“Good data habits from the moment you start planning your research”

Data created from research are valuable resources that can be used and re-used for future scientific and educational purposes.

Good data management practices are essential in research, to make sure that research data are of high quality, are well organised, documented, preserved and accessible and their validity monitored at all times. This results in efficient and credible research. Well-managed data are shared and can be used by other research or to duplicate and validate existing research.

Data management needs to be planned early in a research project, so that practices can be implemented throughout the research cycle.

Roles and responsibilities of the various players in the research process need to be explicitly established.

UK Data Service

Deposit data

“Valuable resources that can be used and reused”

Data created or generated during research or administrative processes are valuable resources that can be used and reused for future scientific and educational purposes. Sharing data can:

- facilitate research beyond the scope of the original research
- encourage scientific enquiry
- avoid duplication of data collection
- provide rich resources for education and training

We proactively acquire data that are suitable for use in research and teaching and that fall within the thematic scope of our Collections Development Policy. We offer different pathways for regular depositors of major survey or other data studies, ESRC award holders and other types of data collections. Please use the appropriate sections to help guide you.

How to deposit data

Different steps for different depositors

“Depositing data is straightforward and rewarding”

At the UK Data Service we offer different pathways for regular depositors of major survey or other data studies, ESRC award holders and other types of data collections.

New depositors

New depositors, new data and ESRC-funded can offer their data by sending us a short description of the data collected. The approval data according to our Collections Development Policy.

Regular depositors

Regular depositors for depositors of large-scale surveys or government data service we offer guidance on how to prepare data, completing our deposit form and licence agreement, information on the survey data is carried out on received data, and the ways in which we're rating and data.
Student resources

"Welcome to our student pages"

The UK Data Service provides access to the UK’s largest collection of social, economic and population data.

On these pages you will can find resources to help you explore and make use of this collection such as:

- information about types of data
- advice on getting started
- answers to FAQs
- our guide 'Using Survey Data' – especially written for dissertation students
- details of our Dissertation Prize

About the UK Data Service

The UK Data Service collection reflects the issues that affect all our lives and experiences, from birth through education, employment, our social interaction and on to old age. Find out more about our service and data by watching our short video.
Getting creative with open data

• Used a crowdsourcing projects to generate innovative uses and outlets for data

• Created high quality open data with no disclosure risk

• Published to an API

• Run an App Challenge

• See UKDS Data Quiz App – Google and Android
Dealing with ‘new and novel’ data

- **Varied Nature** of data of interest in social sciences
  - Surveys collecting biomedical/real-time data
  - Understanding urban environments - Smart Cities
- **Value at High Volumes**
  - Smart energy meters
  - Environmental sensing
  - Wearables and geo-health
  - Social Media
  - Surveillance
  - Transactional data
- **Varied provenance** and often unknown error
Methodological challenges for big data

- Data Integrity
- Reproducibility
- Provenance
- Quality
- Curation
- Preservation
- Long term access and value
- Context
- Ethics and legal frameworks
- Publication and Citation
- Licensing Conditions
Scaling up for big data

- Open Source Enterprise Hadoop system now at UKDA
- Data Service as a Platform (DSaaP)
- Open sources tools for semi-automated QA of data and interrogation – e.g. R tools
- Upskilling staff and social scientists in big data

- Project to help South Africa data archive ‘scale up’
- Significant role for institutions as they accommodate multi-faceted research data needs
- Big data summer school Jan 2017
- Blue print delivered summer 2017

Watch our big data video!
UK Data Service resources

- UKDS [webpages](#) and [video](#) on preparing data
- UKDS [webpages](#) on operating the ESRC Data Policy
- UKDS [webpages](#), book and video on RDM issues

- [Depositing Shareable Survey Data](#) brochure
- UKDS ReShare guide/checking guidelines

- UKDS [Collections Development Policy](#)
- UKDS [Selection and Appraisal Criteria](#)
- UKDS [Data Purchase Guidelines](#)

- **Call to action**: Use of DDI metadata in survey production process
Keep connected

- Subscribe to UK Data Service list: www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKDATASERVICE
- Follow UK Data Service on Twitter: @UKDataService, @UKDSRDM, @LouiseCorti
- Youtube: www.youtube.com/user/UKDATASERVICE
Contact

Louise Corti
Collections Development and Producer Relations team
UK Data Service
University of Essex
UK CO43SQ
corti@essex.ac.uk