Licensing Data for Research: The case of the UK Data Archive

Louise Corti
Director, Collections Development
UK Data Service

Licensing Forum
TNA, Kew
22 March 2017
What is the UK Data Archive

• Department of the University of Essex
• Est. 1967 by the Economic and Social Research Council (ESRC) as a ‘Data Bank’ for science science
• 50 years of curating and providing access to data
• Data and support 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?

- Funded by the Economic and Social Research Council (ESRC)
- 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
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
Common approach: data archives worldwide

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

Explore online

Data access policy

Thesauri

ICSSR Data Service
Some statistics about our UK Service

- **7,064** datasets in the collection
- **1046** qualitative and mixed methods collections
- **400** new datasets and new editions added within last 12 months
  - **99** qualitative in last 12 months
- **25,000** registered users
- **60,000** downloads worldwide per annum
- **4000+** user support queries per annum
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
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 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
Collections Development approach

Trawling

Line-catching
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
Progressive data sharing in the UK

- Long term relationship with ONS
- Supporting the ESRC Research 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
Assessment for new deposits

- Data Appraisal Group assesses data
- Use Appraisal and Selection Criteria – agreed audit trail, research integrity, ethical review
- Decision made:
  - 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>
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

Open where possible, closed when necessary
Access conditions – end user licensing

<table>
<thead>
<tr>
<th>Open</th>
<th>Safeguarded</th>
<th>Controlled</th>
</tr>
</thead>
<tbody>
<tr>
<td>• available for download/online access under open licence without any registration</td>
<td>• available for download/online access to logged-in users who have registered and agreed to an End User Licence</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>
Strategy for enabling safe access

Fives safes enables safe access to data that meet the needs of data protection. Yet fulfils the demands for open science and transparency

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

5 Safes Animation
Non compliance

- Breach procedures
- Removal of individual access to data; maybe permanent
- Individual or institutional suspension from ESRC data services
- Individual or institutional sanction from ESRC funding

- For government data that fall under the Statistics Act 2007, penalties (S39) may include a fine and/or imprisonment
- Sanctions being sought against the licence holder by the data owner
- Self-reported unintentional breaches are penalised with discretion
Depositor Licence types

Concordat
- Formal deposit/access relationship
- ONS, NRS, NISRA
- Covers all ONS datasets

Open
- OGL – no disclosure risk
  - ONS UK aggregate census data
  - Teaching data - ONS surveys
- Creative Commons BY – C & NC
  - Teaching data
  - Bespoke

Standard licence
- Serves all safe-guarded and controlled datasets
- Appendix sets up specific clauses
Safeguarded data – conditions of access

- Most common licence 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 project for which the data are to be used
- Download data in choice of formats
Not just open data …**high quality** open data

- Pathways to access for data held by UK Data Service
- Gaining ODI Platinum certification for open data
- Running an App Challenge
Experience of Open Data Institute (ODI) data certification

2015-10-28 by Louise Corti

Louise Corti, from the UK Data Service, reflects upon the value they gained in the process of attaining ODI Open Data Certificates.

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. The project involved us creating a truly open dataset certified by the ODI to be made available via a new test open API.

The UK Data Service uses the UK Data Archive based at the University of Essex to manage its data curation services. As a national trusted digital repository, the archive conforms to best practice benchmarks for digital preservation via certification under the ISO standard for information security (27001) and the Data Seal of Approval (DSA). While these certifications require transparency and documentation about procedures and practices for managing, curating and publishing data, the questions posed for ODI open data certification place greater emphasis on the presence of detailed machine-actionable information.
The problem of dead depositors…

- Promises of bespoke access conditions by an individual
- Tendency to choose restrictive access
- Difficult to manage long-term
- Renegotiation project
- Advise access via an Access Committee not individual
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
Summary of successful data ingest

- Use standard deposit licenses or data sharing agreements and End User Agreements – not bespoke
  - Open
  - Safeguarded
  - Controlled

- Work to an agreed timetable for data publishing
- Use robust and explicit quality assessment techniques
- Use standard deposit metadata and data description
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...
Links

- NSI Concordat – on request
- Our standard licence
- Our Collections Development Policy
- Our Selection and Appraisal Criteria
- Deposit web pages
- Our Depositing shareable survey data
Contact

Louise Corti, Karen Dennison, Susan Cadogan
Collections Development team
UK Data Service
University of Essex
collections@ukdataservice.ac.uk