UK Data Service responses to changes in the data landscape

Dr Hersh Mann
User Support Manager
UK Data Service / UK Data Archive

IASSIST 2017
Lawrence, Kansas
24 May 2017
Overview

• What is the UK Data Service?
• Facilitating access to (sensitive) data
• ONS Approved Researcher Scheme
• Consultation and Special Licence review
• Impact on UK Data Service and researchers
• Conclusions
What is the UK Data Service?

• A comprehensive resource funded by the ESRC

• A single point of access to a wide range of secondary social science data

• Data support for research and teaching

• Training and guidance on a range of research data issues
Types of data

**Survey Microdata**
- Major UK surveys
- Cross-national, business data, longitudinal

**International Macrodata**
- Multi-nation aggregate databanks

**Census**
- Aggregate data for 1971-2011
- Microdata for 1991-2011

**Qualitative/Mixed Methods**
- Range of multimedia qualitative data sources
- Mix of qualitative and quantitative data
Levels of access

- **Open**
  - Freely available to all

- **Safeguarded**
  - End User Licence (EUL) – data may be downloaded immediately after user and project registration
  - Special Conditions – e.g. additional licence agreements
  - Special Licence – detailed application form and permission from data depositor

- **Controlled**
  - Secure Access data - detailed application procedure, depositor approval, face-to-face training, remote access only, outputs subject to statistical disclosure control (SDC)
Differences between licences

• Data are anonymised
• EUL microdata for major Government surveys will typically have aggregated categories or some of the collected information will be withheld
  • Age is banded e.g. 18-29, 30-39, 40-49 and so on
  • Small geographies such as local Authority District are combined to Government Office Region
  • Ethnicity, country of birth, wealth, income, prescription medication, attitudes, behaviours and so on, could be sensitive
• Special Licence and Secure Access data collections enable researchers to tackle important research questions at higher levels of detail
Data are anonymised

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Region</th>
<th>Age</th>
<th>Education</th>
<th>Jobstat</th>
<th>Jobstart</th>
<th>Hholdsze</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>99</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>99</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>99</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>99</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>99</td>
<td>99</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>99</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>1</td>
<td>99</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>99</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>99</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Example - spatial variables

- A local government researcher in London would want to use microdata that can be analysed by Local Authority District. Additional licences for restricted data can facilitate this.
Approved Researcher (AR) Scheme

- The AR Scheme facilitates access to Office for National Statistics (ONS) data that cannot be published openly
  - Detailed application to demonstrate researcher is ‘fit and proper’
    - Knowledge, experience, security, ‘public good’
  - ONS Special Licence data were supplied by the Economic and Social Data Service and the UK Data Service under the AR Scheme for about a decade
- A consultation on the AR Scheme was started in 2015
Key recommendations

• Researchers must have particular qualifications and/or research experience

• Accredited researchers will be approved for five years

• Researchers will be required to successfully complete Safe User of Research data Environments (SURE) training
  • ONS Special Licence (SL) data were withdrawn from the UK Data Service holdings as a follow-up to this review. The review of ONS SL is ongoing.
Office for National Statistics surveys affected by the Special Licence Review

- Labour Force Survey
- Annual Population Survey
- Integrated Household Survey
- General Household Survey/General Lifestyle Survey
- Living Costs and Food Survey
- Crime Survey for England and Wales
- Life Opportunities Survey
- ONS Omnibus/Opinions Survey
- Wealth and Assets Survey
- National Survey of Bereaved People
Finding data

Series

UK Data Service series record for.

Community Life Survey

SERIES ABSTRACT

The Community Life Survey was first commissioned by the Cabinet Office in 2012-2013. It is an annual household survey conducted with adults resident in England, covering a range of topics including measures that are key to understanding society and local communities. These include volunteering, charitable giving, neighbourhood, civic engagement, social action and subjective well-being. The survey incorporates a small number of priority measures from the Citizenship Survey, which ran from 2001-2011. These measures were incorporated in the Community Life Survey so that trends in these issues could continue to be tracked over time.

DATA ACCESS

+ [GN 33475 | COMMUNITY LIFE SURVEYS, 2012/2013](#)

+ [GN 33476 | COMMUNITY LIFE SURVEYS, 2012/13 - SPECIAL LICENCE ACCESS](#)
Working with ONS

• The SL Review has affected every aspect of our work from pre-ingest, ingest, access, support
• Generating usage reports, the withdrawal of c.150 SL studies, provision of new supporting documentation, survey of projects using SL data from the last 3 years, statistical disclosure control, moving variables from SL to Secure Access equivalent or a new study.
• Streamlining work has been scheduled
• These changes have had a large impact on the work of researchers. SL model is well known. Might other depositors follow?
SDC review

- Anonymisation tests performed on EUL data
  - A range of demographic and employment variables used
  - SDCMicro in R. Results were positive. Some increased risk associated with non-banded age variable.
- Intruder testing performed on SL data i.e. attempts to identify respondents
  - Could some widely-used variables be made available under EUL with edits to other variables?
- A careful balance between research need and security. ‘Personal information’ classification has serious implications.
Access and user support

• Researchers who could previously download SL variables now have to use the UK Data Service Secure Lab for any new data
• This requires successful completion of SURE training
• Outputs they produce are subject to SDC
• Increased short-term workload for our staff
• Long-run impact is uncertain
  • Could use of data decrease?
  • Will researchers meet future access requirements?
Early conclusions

• Be responsive
  • Systems must be in place to react quickly to change

• Build close working relationships
  • Share information

• Understand your users and help your users understand your challenges as well

• More change is imminent
  • Digital Economy Bill
Acknowledgments

- The review of ONS SL data has been a team effort. Special thanks go to Karen Dennison, Sharon Bolton, and Margaret Ward for helping me organise this presentation.
Questions

ukdataservice.ac.uk/help/

Follow us at:
ukdataservice@jiscmail.ac.uk
twitter.com/ukdataservice
facebook.com/ukdataservice