Settings, Practices and Data Access: Results of a Survey of UK Social Scientists

Jo Wathan
Centre for Census and Survey Research
University of Manchester
jo.wathan@manchester.ac.uk
http://www.cccsr.ac.uk/srf
Talking about…

Survey into the research facilities and work practices of researcher as it impacts on data access

• Context and purpose
• Survey findings
• Reflect on process
Why?

- Lot of change in access arrangements – formalisation and restriction
- Data services negotiate on behalf of users
- Relatively little known about researchers’ ability to meet requirements
- Important for ESDS, vital for SARs
UK Data Access Modes
Gov microdata at May 2010

• End User Licence data
  – e.g. LFS, Small Area Microdata

• Special Licence data – disseminated microdata
  – e.g. LFS with local authority district, Household SAR

• Secure Data Service
  – pilot academic remote access

• Virtual Microdata Laboratory
  – e.g. Controlled Access Microdata

• Other agreements possible...

Not identifiable on own or with published data
Not personal data

Approved Researcher access only under SRSA 2007
# What conditions?

Increasing security

<table>
<thead>
<tr>
<th></th>
<th>End User Licence</th>
<th>Special Licence</th>
<th>Secure Data Server</th>
<th>Virtual Microdata Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration required?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Extended application</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data stewardship standards imposed</td>
<td>✗</td>
<td>✓</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>Outputs checked before used outside system</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Publications checked?</td>
<td>✗</td>
<td>?</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>Travel to secure location</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>
Obstacles to use occur when...

<table>
<thead>
<tr>
<th>Obstacle</th>
<th>End User Licence</th>
<th>Special Licence</th>
<th>Secure Data Server</th>
<th>Virtual Microdata Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can’t wait a month for an application to be processed</td>
<td>✗</td>
<td>✓</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>Overseas researcher or collaborator</td>
<td>✗</td>
<td>✓</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>Cannot secure PC</td>
<td>✗</td>
<td>✓</td>
<td>?</td>
<td>✗</td>
</tr>
<tr>
<td>Unwilling to have work checked</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Find it difficult to block out time for research</td>
<td>✗</td>
<td>?</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>Unable or unwilling to travel to secure setting</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>
Aims of the survey

• Obtain unbiased descriptive information on:
  – the community of quants and microdata users
  – facilities available to them (as relates to access)
  – ability to adhere to licencing conditions

• Ethical and data quality issues involved with requesting self disclosure of licence breaches
Social Research Facilities Survey

- Funded under ESRC Census Programme
- RA: Chris Farmer
- mix: census & probability sample
- weighted to give representation of 10 disciplines
- target sample 1,000
- n=598
10 disciplines

- Economics & Econometrics
- Sociology
- Business and Management Studies
- Statistics & OR
- Psychology
- Geography
- Social Policy & Administration
- Politics/International Studies
- Education
- Accounting and Finance
Who are our target audience?

90% of academic social scientists have done research in previous 2 years

- Literature/theoretical = 68%
- Qualitative = 69%
- Quantitative = 59%
What quantitative data?

- Collect own individual/hhd data
- Collect own: other
- Published tables
- Macro data service
- Micro service
- Other

21% - census/survey microdata is relevant to their work
### Environmental obstacles?

<table>
<thead>
<tr>
<th>Obstacle</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mainly work away from institution</td>
<td>14</td>
</tr>
<tr>
<td>Share an office with others (Open plan)</td>
<td>29 (1)</td>
</tr>
<tr>
<td>No sole access to desktop computer</td>
<td>9</td>
</tr>
<tr>
<td>No sole use of lockable drawer</td>
<td>15</td>
</tr>
</tbody>
</table>

Base all respondents: n=597, 592, 595, 595
## Obstacles: practice

<table>
<thead>
<tr>
<th>Obstacle</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot usually predict how long data will take</td>
<td>38</td>
</tr>
<tr>
<td>Difficult to clear several days to work on analysis</td>
<td>80</td>
</tr>
<tr>
<td>Prefer to work at home when analysing data</td>
<td>38</td>
</tr>
<tr>
<td>Rework data many times until happy with it</td>
<td>80</td>
</tr>
</tbody>
</table>

Base: users of microdata in last 2 years n= 192
## Last analysis

<table>
<thead>
<tr>
<th>Practice</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linked data</td>
<td>38</td>
</tr>
<tr>
<td>Conducted multivariate analysis</td>
<td>61</td>
</tr>
<tr>
<td>Kept data after finished analysis</td>
<td>74</td>
</tr>
<tr>
<td>Shared data with UK colleagues</td>
<td>40</td>
</tr>
<tr>
<td>Shared data with overseas colleagues</td>
<td>17</td>
</tr>
<tr>
<td>Provided a copy of the data to journal</td>
<td>16</td>
</tr>
</tbody>
</table>

Base: respondents using microdata in last two years n= 205
## Microdata users’ competencies at institution

<table>
<thead>
<tr>
<th>Obstacle</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot save data to sole access space</td>
<td>4  (5)</td>
</tr>
<tr>
<td>Cannot encrypt data</td>
<td>12 (48)</td>
</tr>
<tr>
<td>Cannot password protect data</td>
<td>9  (21)</td>
</tr>
<tr>
<td>Cannot store outputs in lockable drawer</td>
<td>8  (9)</td>
</tr>
</tbody>
</table>

Base microdata user in last 2 years n= 189
Figure 4: Whether they would do the following activities if it enabled them to access more detailed microdata

- **Travel away from your institution and work on analysis at a specified location**
  - Yes, for routine analysis: 60%
  - Yes, in exceptional circumstances: 50%
  - No: 90%

- **Submit a data application in a month long application procedure**
  - Yes, for routine analysis: 50%
  - Yes, in exceptional circumstances: 60%
  - No: 80%

- **Submit your data analysis to be approved by a data service (e.g. ONS, UK Data Archive)**
  - Yes, for routine analysis: 40%
  - Yes, in exceptional circumstances: 60%
  - No: 70%

- **Work on analysis in a safe setting at your institution**
  - Yes, for routine analysis: 30%
  - Yes, in exceptional circumstances: 70%
  - No: 80%

- **Work on analysis solely at your institution in your office**
  - Yes, for routine analysis: 20%
  - Yes, in exceptional circumstances: 80%
  - No: 90%
## Cumulative impact

<table>
<thead>
<tr>
<th></th>
<th>Special Licence</th>
<th>Secure Data Server</th>
<th>Virtual Microdata Lab</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not have sole use pc</td>
<td>9%</td>
<td>9%</td>
<td>9%</td>
<td>9</td>
</tr>
<tr>
<td>Overseas researcher or collaborator</td>
<td>24%</td>
<td>24%</td>
<td>24%</td>
<td>33</td>
</tr>
<tr>
<td>Will not undertake month long application</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
<td>54</td>
</tr>
<tr>
<td>Unwilling to ever have work checked</td>
<td></td>
<td>12%</td>
<td>12%</td>
<td>66</td>
</tr>
<tr>
<td>Unable or unwilling to travel to secure setting</td>
<td></td>
<td></td>
<td>11%</td>
<td>77</td>
</tr>
<tr>
<td>Total</td>
<td>54%</td>
<td>66%</td>
<td>77%</td>
<td></td>
</tr>
</tbody>
</table>
Other headlines

• Generally content with resources at their institution
  – 69% IT support
  – 63% data storage/archiving facilities

• 51% of supervisors report all students have sole access to desk and PC
Conclusions

• Move towards SDS is likely to be popular
• Extensive applications and data checking are also obstacles
• Environmental issues are not typically an obstacle to use
• Work patterns and preferences are!
Reflections on undertaking the survey

Unusual to have a probability sample of entire community
- Excludes non-staff students
- Excludes non HE
- Excludes users in ‘odd’ disciplines

Expensive and time consuming!
- Registration data constraints restrict use
- No good register of academics
- Good response rate because of extensive follow up
Interested?

- [email](mailto:Jo.wathan@manchester.ac.uk)
- [Website](http://www.ccsr.ac.uk/srf)
- Working paper in review
- Data in review
- Invitation to support users