Constructing a Preservation Policy: 
the case of the UK Data Archive 

Matthew Woollard 

Head of Digital Preservation 
and Systems, UKDA
Requirements influencing policy

- organisational and business need
- user requirements
- standards
- legislation
- technology and technical environment
- security
UKDA Mission Statement

“to support high quality research, teaching and learning in the social sciences and humanities, by acquiring, developing and managing data and related digital resources, and by promoting and disseminating these resources as widely and effectively as possible.”
SN 1 - Family, Work and Leisure in the London Region, 1970: Main Study

*Internal User*

**Title:**
Family, Work and Leisure in the London Region, 1970: Main Study

**Subject Categories:**
Community and urban studies - Society and culture
Family life and marriage - Social stratification and groupings
Leisure, tourism and sport - Society and culture

**Depositor(s):**
Willmott, P., Institute of Community Studies

**Principal Investigator(s):**
Willmott, P., Institute of Community Studies
Young, M., Institute of Community Studies

**Sponsor(s):**
Nuffield Foundation
Leverhulme Trust
Centre for Environmental Studies
Post Office
Sports Council

**Abstract:**
The purpose of this survey was to examine the relationship between work, home and leisure in order to explain changes in family patterns; to provide demographic, social/behavioural and economic data for the analysis of family life.

**Main Topics:**
Attitudes/Behavioural Questions
Present and preferred occupation: type, hours, holidays, career structure, autonomy at work, whether physically or mentally bring or boring. Journey to work.
Present and preferred residence: size, district, tenure, type, garden.
<table>
<thead>
<tr>
<th>Column</th>
<th>Title</th>
<th>Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/21</td>
<td>Pressed</td>
<td>Not pressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 464</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sometimes pressed, sometimes not</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 352</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 414</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I/I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 9 6 6 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not applicable (i.e. not working)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 9 2 8 full or part time</td>
</tr>
<tr>
<td>2/22</td>
<td>Physically tiring</td>
<td>Not applicable (i.e. not working)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 9 2 8 full or part time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not applicable (i.e. not working)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 9 2 8 full or part time</td>
</tr>
<tr>
<td>Pressed</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>Valid</td>
<td>468</td>
<td>36.7</td>
</tr>
<tr>
<td>Not Pressed</td>
<td>382</td>
<td>29.9</td>
</tr>
<tr>
<td>Sometimes pressed, sometimes not</td>
<td>417</td>
<td>32.7</td>
</tr>
<tr>
<td>Pressed</td>
<td>9</td>
<td>.7</td>
</tr>
<tr>
<td>Incomplete information or N/A</td>
<td>1276</td>
<td>100.0</td>
</tr>
</tbody>
</table>
### Workplace

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central area</td>
<td>211</td>
<td>10.9</td>
<td>10.9</td>
</tr>
<tr>
<td>Elsewhere Inner London (ILEA)</td>
<td>154</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Outer GLC</td>
<td>366</td>
<td>19.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Outer Metropolitan Area</td>
<td>385</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Outside London Region</td>
<td>7</td>
<td>.4</td>
<td>.4</td>
</tr>
<tr>
<td>Works at home</td>
<td>27</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>No fixed place, travel around, &quot;all over&quot;</td>
<td>121</td>
<td>6.3</td>
<td>6.3</td>
</tr>
<tr>
<td>Insufficient Information / N/A (i.e., not working)</td>
<td>655</td>
<td>34.0</td>
<td>34.0</td>
</tr>
<tr>
<td>Total</td>
<td>1928</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**Note:** Change from precodes 8 and 9.
Standards used to inform

- Rights (Legal Protection of Databases, Re-use of PSI)
- BS ISO 14721 OAIS
- BS ISO 18492 Long term preservation of electronic information
- Information security standards (BS ISO 27001 and 27002)
- Records management standards (BS ISO 15489)
- Media/hardware storage and maintenance
Standards used to inform

- Archival standards (BS 5454)
- Metadata standards
- File formats

Also:
- Other policies (TNA, English Heritage...)
- Other guidelines (TNA, AHDS)
- Planning tools
Planning tools

- TRAC (Trustworthy Repositories Audit and Certification) Criteria and Checklist
- the NESSTOR catalogue of criteria
- the Digital Preservation Coalition’s Handbook
- the DRAMBORA toolkit
- ‘Core requirements for Digital Archives’ (Center for Research Libraries)
- Data Seal of Approval published by DANS.
Life-cycle approach

- Legal and regulatory framework
- Pre-ingest / ingest
- Archival storage
- Systems (hardware) / IT architecture
- Data Management / metadata
- Access
Key changes in new policy

- Increased effort on authenticity, integrity and reliability;
- adherence to archival/records management standards;
- new concepts surrounding versioning and editions;
- definitions for archival packages;
- verification of AIPs;
- improved data management (i.e., non-resource discovery metadata);
- preservation of resource-discovery metadata;
- planning open implementation of preservation metadata.
Acquisition of data...
from students' phd, theses, research organisations, market research organisations, academics, etc.

Data arrives
in card or tape form with explanatory documents

Documents entered and checked
eg. codebook complete?
questionnaire included?
licence form signed?

Data entered and checked
eg. wild codes?
pregnant males?
too many cases?
multi-punched?

Data converted to standard storage form
eg. data put on tape with back-up tapes created for security

Neat and tidy codebook compiled

Description written for inventory

SSRC SURVEY ARCHIVE

User query

Data query
eg. how do I get about getting a copy of the NOP political data?

Data located from file

Parameter cards prepared for run on data file

Produce copy of data and codebook
data are provided in a format compatible with the users' computer installation.

Produce copy of codebook

Check inventory
and obtain from other sources if necessary

Locate relevant surveys
if data of interest application form sent to enquirer

Text query
have you got any data on tourism?