Redesigning and formalising a national data archive’s Collection Development Policy

Louise Corti
IASSIST, May 2005
Typical sources of data

UKDA - data for research and teaching purposes and used in all sectors and for many different disciplines:

Pre-2000

- initiated 1967 – acquiring academic survey data
- From 1974 official agencies - mainly central government
- individual academics - research grants via Research Councils and charities
- market research agencies
- public records/historical sources
- access to international data via links with other data archives worldwide, via IASSIST/IFDO community
Collection held at UKDA

- 5,000+ datasets in the collection
- 300+ new datasets are added each year
- 9000 registered users
- 38,000+ datasets distributed worldwide per year
- history data service in-house (AHDS History)
- data disseminated at no cost for higher education where usage is not commercial
Transitions from original data archive

- Data archive to multi-data value-added support services
  - AHDS History
  - Qualidata unit
  - Census Registration Service
  - Economic and Social Data Service (ESDS)
  - NCeSS – supporting e-social science
  - cross-council research programmes
  - other funded projects that relate to resource discovery/metadata
  - preservation function for other data services
Broadening the scope

- 1994 – specialised history data service
- 2000 – specialised qualitative data service
  - text interview transcripts/focus groups
- 2001 - mixed methods collections
  - survey data and qualitative
- 2001 - links to UK census data
- 2003 – Economic and Social Data Service (ESDS)
  - international statistical time series via ESDS International - first data purchase
- 2005
  - audio recordings of qualitative interviews
  - video clips
- 2006 anticipated
  - more commercial data sources
  - web sites
  - more sensitive government micro data
  - links to European data via Nesstar Euro portal
ESDS Overview

• core archiving services:
  – collections development strategy
  – ingest activities - including data and documentation processing, metadata creation
  – data dissemination services
  – long-term preservation

• plus four specialist data services to provides access and support for key economic and social data:
  – ESDS Government
  – ESDS International
  – ESDS Longitudinal
  – ESDS Qualidata

• provides:
  ✓ dedicated web sites
  ✓ data and documentation enhancements
  ✓ user support
  ✓ training
The large-scale government surveys

Excellent formalised mutual relationship with Office of National Statistics

- General Household Survey; ONS Omnibus Survey
- Labour Force Survey
- Health Survey for England/Wales/Scotland
- National Food Survey/Expenditure and Food Survey; Family Expenditure Survey; Family Resources Survey
- British Crime Survey
- Survey of English Housing
- British Social Attitudes
- National Travel Survey; Time Use Survey
  - some series date back to 1961
  - acquired free and in timely manner from ONS or from NATCEN
  - many delivered via Nesstar
Official national archive status

- from 1 January 2005 UKDA became an official Place of Deposit for government records

- discussions with the National Archives (TNA) concerning the custodianship and long-term preservation of important government records

- the Public Records Act (1958) determines what are public records to be selected and scheduled for permanent preservation

- applies to records over 30 years old - under FOI earlier records can be made available

- UKDA also now holds the only preservation copies of key government data series in the country

- evaluation process - UKDA conforms to various British Standards (most notably BS5454)

- Means wider access to those outside academic/policy spheres and will need to monitor future requests
Longitudinal Data

• Acquire key studies primarily UK Research Council:
  - British Household Panel Survey (BHPS)
  - British Birth Cohort studies:
    • National Child Development Study (NCDS)
    • 1970 British Cohort Study (BCS70)
    • Millennium Cohort Study (MCS)
    • English Longitudinal Study of Ageing (ELSA)
  - also possible forthcoming Medical Research Council population cohort data
  - And new forthcoming longitudinal qualitative study with £3m investment
Other top ten core survey

• Government departments and academic funding:
  
  - British Election Studies
  - Census datasets
  - Home Office Citizenship Survey
  - International Passenger Survey
  - Psychiatric Morbidity Surveys
  - Road Accident Data
  - Scottish Household Survey
  - Survey of Mortgage Lenders
  - Workplace Employee Relations Survey
  - Youth Cohort Study of England and Wales
Qualitative Data Collections

- data from National Research Council (ESRC) individual research grant awards
- data from ESRC Programme research grant awards
- data from ‘classic’ social science studies
- other funders/sources
- focus on DIGITAL Collections, but also facilitate paper-based archiving
- a few delivered via ESDS Qualidata Online
Acquisitions Review Process

- pre-2000, UKDA accepted what it was offered
  - subject to physical data quality and matched with collection supporting documentation assessment

- post-2000, more systematic evaluation
  - new ARC committee meeting every 2 weeks
  - formal evaluation of submission forms (limited info) assigned to staff with research backgrounds
    - focus on theme, scope breadth, usage potential, quality of research, data, consent and confidentiality
    - Accept/reject and assigned levels of processing
      - new data creation guidance
      - new depositor guidance
      - new forms and licences

- post-2003 service level agreement to acquire fixed minimum number (quota) of studies per year
### Types and number of acquisitions for ESDS core dissemination

<table>
<thead>
<tr>
<th>Origin of acquired data</th>
<th>N</th>
<th>p.a</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESRC-funded project</td>
<td>45</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>other academic project</td>
<td>10</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>market research companies</td>
<td>10</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>UK Government departments and agencies</td>
<td>20</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>95</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

*Excludes acquisitions that fall under the value-added functions for International, Longitudinal, Qualitative and Large-scale Government data*

- In reality UKDA accept more, especially government data. Limited processing staff. Reject large proportion, often discipline based: education, psychology. Problems for operating the ESRC Datasets Policy
Evaluation criteria

• assessing content, long-term value and the level of potential re-use interest:
  - users waiting with requests
  - consent and confidentiality permit data sharing
  - geographic and/or temporal scope
  - subject coverage of the data is broad, of interest across the social science disciplines
  - makes the resource more accessible than would be otherwise
  - complements or fills a gap in the existing holdings
  - research and/or teaching activity in the subject area covered by the data
  - data for which longevity and access would otherwise be threatened.

• can they be viably managed, preserved and distributed to potential secondary users:
  - ESDS has expertise, can obtain expertise or expert advice
  - data format can be converted to suitable dissemination and preservation formats
  - level and quality of documentation reaches an appropriate standard to enable a secondary analyst to make informed use of the data
  - level of data purchase, if any
Research Council Acquisitions

- ESRC Datasets Policy to preserve and share data from ESRC-funded research
- funding allowed to prepare data for archiving
- all award-holders must offer data for deposit to the ESDS within three months of the end of the award
- any potential problems should be notified to the ESDS at the earliest opportunity
- final payment will be withheld if dataset has not been deposited within three months of the end of the award, except where a waiver has been agreed in advance
No systemic ‘policing’ of Policy

- ESDS remit to offer advice to data creators and review offers of ESRC data

- ESRC have never provided lists of data being created
  - though do collect this information electronically on grant submission forms

- for UKDA - random acquisition process
  - some chasing of data we want
  - and declines for data we want, and advance deposit waivers sometimes requested

  - provide guidance and support to data creators via phone, web and workshops

- no official line on how much money can be costed into a grant for data and documentation preparation
  - often projects end abruptly, key staff leave and data are left in a mess
Datasets Implementation Policy

- UKDA/Qualidata proposed more formalised process in 2000
- much more linked to the data/research lifecycle
- joint database for ESDS and ESRC staff to record and monitor ALL data creation grants
- will be ‘green-lighting’ potentially desirable projects
  - receive all successful application forms
  - require advance review by ARC
  - create more work upfront
- will be able to liaise with data creators early in the research process
  - one-on-one pro-active contact
  - workshops for chosen data-rich thematic programmes
- but will not be rigidly selective for those data that have been initially red-lighted
Contract - Natural Environment Research Council

- more systemic Data Policy but less general outreach and training work
- all grant holders are required to offer to deposit with NERC a copy of datasets resulting from their research
- six official Designated Data Centres (DDCs)
- data support and collection centred around a specific Research Programme
- ongoing budgeted data management support working with data creators over life of project
- most data acquired, unless problems
- metadata template for data submission (primarily excel), drawn up for each programme, standards based
New joint ESRC/NERC Programme

- UKDA won award for Data Support Service for multi-million Research Programme on Land Use and Rural Economy

- service run from UKDA but staff split across four sites specialising in different kinds of data

- opportunity to:
  - combine best features of each Council’s approaches to data policies
  - to encourage interdisciplinarity
  - encourage wider knowledge about data sources; better resource discovery

relu.esds.ac.uk
Contact

www.esds.ac.uk/aandp/create/introduction.asp

acquisitions@esds.ac.uk

Susan Cadogan
Julie Missen
Louise Corti (Head of section)