UK Data Archive
www.data-archive.ac.uk

esds
Economic and Social Data Service

Implementing selection and appraisal policies at the UK Data Archive

K. Schürer
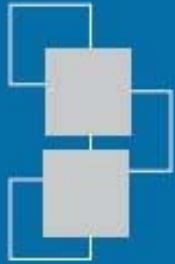
UK Data Archive and Economic and Social Data Service
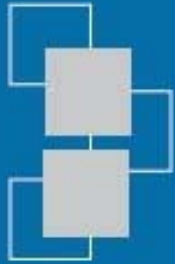
DCC and DPC Workshop, Oxford

4th July 2006

UKDA history & overview

- Archive established in 1968 (as 'Data Bank')
- Funded by (then) SSRC to provide a service to UK HE sector
- Initial focus on government survey data
- New distributed service established 1 Jan. 2003
    - Economic and Social Data Service (ESDS)
- Mixed data types and formats
    - Specialist Qualidata unit and History Data Service
- Still predominately funded to provide service for HE/FE sectors
    - ESRC, JISC, University of Essex
    - Project funding (EC, JISC, MRC, AHRB, etc.)
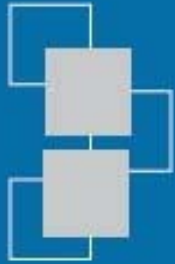- Since 2005 designated as 'Place of Deposit' by TNA

Who produces social science data?

- Government agencies
- HE/FE sector
- Private sector
- Within HE/FE not just ESRC funded
  - MRC, NERC, AHRB, Wellcome, Leverhulme
- Increasing number of large digitisation projects
  - JISC, NOF
- Increasing tendency for government agencies to contract out survey work to private sector (NatCen)
- University sector tend not to get Government contracts
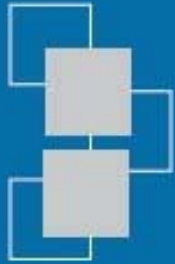- Devolution
- Local Government

Types of data

- Quant and Quali
- Surveys
- Censuses
- Administrative data
- Also increasing amounts of 'non-survey' type data
    - Images
    - Sound
    - Video
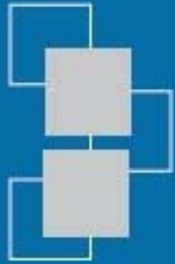   - Mixed media

Who owns data?

- Nearly all Government data technically owned by the Queen – Crown Copyright
- HE/FE sector generated data, host institution

But host of IPR, data copyright and consent issues.

Selection vs appraisal

Selection guided by 'use' & 'users'

- UKDA seeks to identify and acquire material within broad areas
- *Discipline coverage:* at the broadest level, data and other electronic resources relating to society, in particular data about individuals or groups of individuals. This includes strategic social science and economic datasets e.g. unemployment statistics, major household surveys.
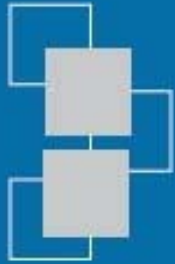- *Geographic coverage:* data across a broad geographic coverage focusing on the United Kingdom and cross-national datasets but including material from other countries where appropriate and in particular where these provide opportunities for comparative research e.g. European data.
- *Temporal coverage:* there are no restrictions on temporal coverage, although pre-1945 accessions are acquired through AHDS History.
- *Time series and panel data:* data are sought which create or add to a time series and/or panel survey.
- *Thematic coverage:* to create a coherent body of materials relating to a particular discipline or field of enquiry e.g. health.
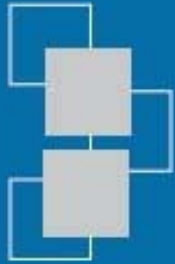
The National Archives

- The state's own administrative and policy processes
  - formulation of policy and management of public
  - resources by the core executive
  - management of the economy
  - external relations and defence policy
  - administration of justice and the maintenance of security
  - formulation and delivery of social policies
  - cultural policy
- Interaction of the state with its citizens and its impact on the physical environment
  - the economic, social and demographic condition of the UK, as documented by the state's dealings with individuals, communities and organisations outside its own formal boundaries
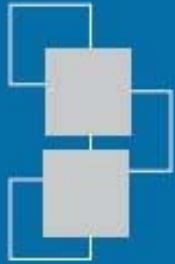 - impact of the state on the physical environment

UKDA/ESDS

- Ensure they fall within the 'Scope of Collections'
- Assess their content, long-term value and the level of potential interest in their re-use. Factors influencing this evaluation include:
    - the geographic and/or temporal scope is significant;
    - the subject coverage of the data is broad and may be of interest across the social science disciplines;
    - the data are not available in any other form;
    - accession into the Archive makes the resource more accessible;
    - a dataset adds to or is made more valuable by existing holdings (eg part of an existing series);
    - a dataset fills a gap in the existing holdings;
    - there is research and/or teaching activity in the subject area covered by the data;
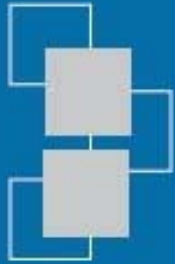   - data for which longevity and access would otherwise be threatened.

Research funded by ESRC, BA and other agencies

- contractually obligated to 'offer' their research for deposit and preservation
- pre-2000 UKDA accepted what it was offered – subject to assessment of physical data and supporting documentation quality
- post-2000 more systematic evaluation scheme implemented:
  - Acquisition Review Committee established, meeting every 2 weeks
  - Formal evaluation of submission forms; assigned to staff with research backgrounds
  - Focus on theme, scope breadth, usage potential quality of research, consent and confidentiality
  - Accept/reject and assigned levels of processing
- post-2003 service level agreement to acquire fixed minimum number (quota) of studies per year
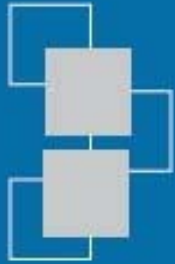
Economic and Social Data Service

Datasets Implementation Policy

- UKDA created comprehensive web based guidance for data creators/depositors to support the ESRC's datasets policy

- Joint database for UKDA/ESRC staff to record and monitor ALL data creation awards

- Enables UKDA to liaise with data creators early in the research process/data 'life-cycle'
  - one-on-one pro-active contact
  - workshops for chosen data-rich thematic programmes
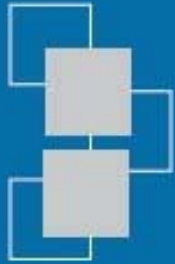
- Role of ARC is broadly to determine if data may be viably managed, preserved and distributed to potential secondary users
- Factors influencing this evaluation include:
  - are data of a type with which the UKDA has expertise or may easily obtain expertise or expert advice?
  - can data formats be converted to suitable dissemination and preservation formats?
  - is the level and quality of documentation at an appropriate standard to enable a secondary analyst to make informed use of the data and to allow preservation overtime?
  - are there problems with the sample size, are data very localised or of peripheral interest for users (social scientists)?
  - would the data be more effectively handled by another archive or data centre?
  - are there insurmountable legal obstacles e.g. copyright clearance from all parties has not been achieved?
  - does the depositor wish to impose too stringent access conditions or wishes to place an indefinite embargo on use?
  - costs

Prospects for the future?

UK Data Archive
www.data-archive.ac.uk

- Legal issues
  - Move toward greater openness and transparency
  - Freedom of Information Act, 2000
  - Yet greater concerns with confidentiality
  - Data Protection Act, 1994
- Technical issues
  - e-Science and data grid
  - Institutional repositories
  - 'Self-archiving'
- Political issues
  - RCUK statement on research outputs