DATA REVIEW AT THE UK DATA ARCHIVE
UK DATA SERVICE & ESSEX REPOSITORY (PILOT)

VEERLE VAN DEN EYNDEN

UK DATA ARCHIVE
UNIVERSITY OF ESSEX

IDCC13 – DATA PUBLICATION, PEER REVIEW AND REPOSITORY ACCREDITATION
AMSTERDAM, 17 JANUARY 2013
WHO WE ARE & WHAT WE DO

- curator of digital data in the social sciences and humanities
- enhance and provide access to data for research and teaching
- provide detailed user guides and documentation
- access to data via search, download, online visualisation
- support researchers, data producers and data users
- raise standards in data management
- raise standards in data security
- partner on R&D data projects
SOURCES OF DATA

- government survey data
- individual academics - research grants (ESRC and BA funded)
- market research agencies
- public records and historical data
- international statistical time series
- qualitative, quantitative and cross-disciplinary data
- access to international data via links with other data archives worldwide

- 5000 datasets in the collection
- 230 new datasets each year
- over 22,000 registered users
- approx. 60,000 downloads worldwide per year
- 3000+ user support queries
RESEARCH DATA @ESSEX

Piloting a research data management and sharing infrastructure for the University of Essex

Team: UK Data Archive, Information Systems & Services (ISS), Research and Enterprise Office (REO), Records Management

Working with pilot departments:
- Biological Sciences (proteomics, bio-imaging)
- Essex Business School (management)
- Language & Linguistics (2nd language acquisition, socio-linguistics)
- Computing & Electronic Systems (artificial intelligence)
RESEARCH DATA @ESSEX ACTIVITIES

- Policy & advocacy
  - UoE Research Data Management (RDM) policy
  - ESPRC policy framework roadmap
- Support & training
  - DMP guidance
  - DM training
- Data repository
  - EPrints data instance
  - user interface and workflow customisations
  - metadata profile & standards (DDI, INSPIRE)
  - digital object identifiers (DOI)
DATA REVIEW

• Economic and Social Data Service (ESDS) 2002-2012
  • Acquisitions Review Committee
• UK Data Service (UKDS) 2012-2017
  • UKDS Collections Development Committee (independent)
  • Data Appraisal Group
ESDS ACQUISITIONS REVIEW

- Data offered to ESDS, via Data Review Form
  - data topic, methods, size, format, units, ethical/legal issues
  - incl. representative samples of data and documentation
- Acquisitions Review Committee meets 2-weekly to evaluate the quality and potential re-use value:
  - accept data into main ESDS collection for curation & long-term preservation
    - processing standards: A*, A, B, C depending on condition of data and documentation & anticipated level of re-use
  - accept data into self-archive system ESRC Data Store (UKDA-store) for short-term management and access
  - unable to accept data
ESDS ACCEPTANCE CRITERIA

• Within scope of UK Data Archive collection
  • social sciences & history
  • UK & cross-national
  • time series & panel data
  • for research & teaching

• Longterm value and re-use potential
  • broad coverage (geographic, theme, discipline, sample size,....)
  • not otherwise available
  • more accessible through acquisition
  • popular research trend
  • fits in / complements existing collections

• Data requested (ESDS Advisory Committee, users)

• Data from ESRC funded research

• Viable for preservation
  • acceptable file format
  • well documented
Criteria for non-acceptance of data

- value of data in publications
- legal obstacles (copyright, IPR)
- ethical constraints (consent, anonymisation)
- depositor wishes unnecessarily stringent access conditions
UKDS CURATION CATEGORIES

• *Data collections selected for long-term curation*
  • long term secondary analysis potential
  • processed, curated, preserved by UKDS

• *Data collections selected for short-term management*
  • backed-up, made accessible and discoverable online
  • self-archiving into ESRC Data Store

• *Data collections selected for delivery only*
  • data from 3rd parties accessed via APIs/web services

• *Data collections selected for discovery only*
  • in other (institutional) repositories, harvest metadata
UKDS APPRAISAL CRITERIA (draft)

Essential data criteria (relevance)
• potential for secondary use
• or potential for teaching and learning use
• or replication and validation of research use
• needed for research and policy
• high quality, reliable and up-to-date
• good temporal and spatial coverage
• scientific value for social and economic sciences
• historical value
• data are well documented
UKDS PRIMARY CRITERIA (draft)

- New data source: transactions data (admin records, commercial records), tracking records
- Data with international value / for international research:
  - authoritative source of high quality statistics
  - reference databases used globally to support decision-making and policy formation
  - harmonised and comparable between countries
- Longitudinal or consistent time series data
UKDS PRIMARY CRITERIA (draft)

- Data needed for comparative research, e.g. across the UK and its devolved areas
- Disclosive microdata otherwise not available to the research community (e.g. business data, Secure Data Service)
- Contractual obligation for UKDS to acquire data
- Unique, unrepeatable data that are costly to reproduce
- Data directly requested by the dedicated user community

National Data Strategy
SUPPORTING CRITERIA (*draft*)

- Data in a suitable format for reuse (e.g. no specialised software needed or such software made available) or can be converted
- Data have sufficiently open access conditions
- Widely cited data
- Anonymised data
NEGATIVELY INFLUENCING SELECTION

• Legal and ethical constraints prohibiting archiving, sharing and/or re-use
  E.g. no consent to share/reuse data, IPR, copyright, data protection violation
• Restricted or limited sample size or sample representativeness
• Lack of documentation and contextual information
• Format unsuitable for re-use
POPULAR DATASETS (2008-2012)

• Data series
  • Quarterly Labour Force Survey
  • Health Survey for England
  • British Household Panel Survey
  • British Social Attitudes Survey
  • 1970 British Cohort Survey

• International macrodata
  • World bank World Development Indicators, 1960-
  • IMF International Financial Statistics, 1948-
  • IMF Government Finance Statistics, 1972-
  • IMF Direction of Trade Statistics, 1980-
  • IMF Balance of Payment Statistics
POPULAR DATASETS 2008-2012

• Individual datasets
  • British Crime Survey
  • Gender Difference, Anxiety and the Fear of Crime
  • GIS of the Ancient Parishes of England and Wales, 1500-1850

Our data in use: case studies using ESDS data
ESRC DATA STORE CRITERIA (draft)

- Data resulting from ESRC awards
- Contractual obligation from funder to researcher

Negatively influencing selection
- Legal and ethical constraints prohibiting sharing / re-use
  E.g. no consent to share/reuse data, IPR, copyright, data protection violation
- Restricted or limited sample size
- Lack of documentation and contextual information
- Format unsuitable for re-use
INSTIT. REPOSITORY PRIMARY CRITERIA

Appraisal by data owner & IR

• Data cited / referenced in a publication
• Funder requirement to preserve/share data
  E.g. ESRC, BA, AHRC, MRC
• Institutional legal requirement to retain/share data
  E.g. Freedom of Information, Environmental Information
  Regulation
• Scientific value – potential for use in current or future
  research
• Learning and teaching value
• No legal/ethical constraints prohibit preservation/sharing
  E.g. consent, data protection, IPR
SUPPORTING CRITERIA

- good data documentation and metadata
- suitable format for reuse (e.g. no specialised software needed or such software made available) or can be converted
- openness of access conditions
- easy to ingest without significant processing/preparation
- data in good condition (readable, undamaged,...)
- cost of ingest
QUALITY ASSURANCE OF DATA

Ensure high quality, accurate, well organised, accessible data with long-term validity

- Data processing to standards
- Processing standards: A*, A, B, C depending on condition of data and documentation & anticipated level of re-use
- Data management training and guidance for data producers / researchers
OUR RESOURCES

Best practice guidance on managing and sharing research data (online and published guide)

incl. training resources, examples, demos, templates, tools
OUR DATA MANAGEMENT GUIDANCE

- planning for data sharing
- ethical and legal aspects of data sharing and re-use
- data copyright
- documentation and metadata to understand and use data
- data formats, formatting and quality control for long-term preservation
- storage and back-up of data and files
- security and controlled access to data
- strategies for research centres and large projects

www.data-archive.ac.uk/create-manage
QUESTIONS ??