Accessing Confidential National Microdata

Presented by
Reza Afkhami

DwB Training Course, Bucharest, 13 February 2013
Research Data Centres

- Provide onsite and/or remote access to data considered to be too detailed, confidential or sensitive to be provided through standard access mechanisms.

- Typically associated with National Statistical Institutes, government departments and national Data Archives.

- Depending upon the RDC, available datasets are administrative data, social survey, census and business microdata.
Data Access Spectrum

MORE CONFIDENTIAL, MORE SECURE

No release | Research data centres | Restricted Licence Files | Scientific Use Files | Public use files

Source data | Identifiable business, social, health data | Detailed social data, cross-government data access | Anonymised personal data

LESS CONFIDENTIAL, EASIER ACCESS

Modified from Ritchie, 2006
Research Data Centres – examples (1)

- Denmark
  - Statistics Denmark
- Estonia
  - Statistics Estonia
- France
  - CASD/GENES
- Germany
  - Federal Statistical Office
  - German Microdata Lab, GESIS
  - Research Data Centre of the German Federal Employment Agency
- Netherlands
  - Statistics Netherlands
Research Data Centres – examples (2)

- Portugal
  - Statistics Portugal
- Slovenia
  - Statistical Office of the Republic of Slovenia
- United Kingdom
  - Virtual Microdata Laboratory, Office for National Statistics
  - Secure Data Service, UK Data Archive
Typical Application Process

• Application form
  ➢ Detailed project proposal and intended publications
  ➢ Previous relevant experience and publications

• Signed agreements / contracts

• Application sent to data owners / data access committee for approval

• Training in legal aspects, how to make outputs safe, how to use the access facility, consequences of breaching the terms and conditions
Data Access

Access varies depending upon the RDC and the data concerned

- Remote execution
- Remote access at researcher’s own institution
  - Usual office desktop
  - Thin-client
  - Safe room
- Onsite in RDC safe room

Outputs are usually vetted by RDC staff to ensure that they are safe for publication
Challenges

• Cost to the researcher (travel, access)

• Variation between RDCs in:
  
  ➢ Data classification
  ➢ Legal gateways and obstacles
  ➢ Cost
  ➢ Application process and criteria
  ➢ Length and complexity of approval process
  ➢ Data access method and remote access policy
  ➢ Data matching/linking rules
  ➢ Output checking ‘rules’
UK SECURE DATA SERVICE
About the **Service**

- For the academic research community in the UK
- Delivered by the UK Data Archive
- Funded by the Economic and Social Research Council
- Supported by the Office for National Statistics
- Government and ISO 27001 accredited
About the Service

Working with data owners, the Service provides researchers with secure access to data that are too detailed, sensitive or confidential to be made available under the standard licences operated by the UK Data Archive.
Application process

- Register with and order through the Economic and Social Data Service
- Complete application form – vetted by us and data owners
- Attend mandatory face-to-face training
- Complete Service User Agreement that outlines terms and conditions of use of the service and requires institutional signature
- Receive member logon
Training

- an introduction to the training, the service and other similar services

- data security and personal responsibility, including legal background, security model, breaches and penalties

- Statistical Disclosure Control - how to make statistical outputs safe and what principles are used, including hands-on exercises

- using the Secure Data Service - about the service and how to use it, how to obtain outputs and how to use the service safely
Using the Service

- UK academics – institutional desktop access
- Non-resident researchers – safe room onsite at UKDA
- Secure remote access (log in via a web browser to launch a remote desktop) – can’t download, copy and paste, print
- Access to standard software (SPSS, Stata, R, Word, Powerpoint etc)
- Conduct all research within the secure environment
- Final outputs are checked by the team for Statistical Disclosure Control before being released
How the Service is secure

• secure data technology and procedures
  ➢ Including ISO 27001

• training and convenience for approved researchers

• meaningful penalties for breaches
  ➢ UK Statistics Act provides legal gateway to provide detailed ONS data to Approved Researchers and also the ability to prosecute for misuse
  ➢ ESRC backed penalties (e.g. including withdrawal of current and future funding)
The data

Depending upon the data, typically include more detailed geography, industry, occupation, health and demographic variables than data that can be downloaded.

Provided by the UK Office for National Statistics

- Business data
- Social survey data

ESRC-funded research

- British Household Panel Survey
- Understanding Society
- Millennium Cohort Study linked to the National Pupil Database
Data catalogue

Annual Business Inquiry, 1973-2008: Secure Data ...
The Annual Business Inquiry (ABI) is the financial information survey conducted by the Office for National Statistics (ONS). This is a statutory survey conducted under ... ORDER FROM ESDS

Annual Population Survey, 2004-2010: Secure Data ...
The Annual Population Survey (APS) is a major survey series, which aims to provide data that can produce reliable estimates at local authority level. Key topics covered ... ORDER FROM ESDS

British Household Panel Survey, Waves 1-18, ...
The British Household Panel Survey (BHPS) is conducted by the ESRC UK Longitudinal Studies Centre (ULSC), together with the Institute for Social and Economic Research ... ORDER FROM ESDS

Business Spending on Capital Goods, 2000-2004 and ...
The Business Spending on Capital Goods survey is an annual survey of business organisations in the UK that collects data about acquisitions and disposals of capital, ... ORDER FROM ESDS

Annual Inquiry into Foreign Direct Investment, ...
The Annual Inquiry into Foreign Direct Investment (AFDI) collects financial data concerning Foreign Direct Investment (FDI). The survey is conducted in two parts: the ... ORDER FROM ESDS

Annual Survey of Hours and Earnings, 1997-2009: ...
The Annual Survey of Hours and Earnings (ASHE) is one of the largest surveys of the earnings of individuals in the UK. Data on the wages, hours of work, and pensions ... ORDER FROM ESDS

Business Expenditure on Research and Development, ...
The Business Expenditure on Research and Development is an annual survey whose main purpose is to supply data for policy purposes on Science and Technology, of which ... ORDER FROM ESDS

E-commerce Survey, 2000-2008: Secure Data Service ...
The E-commerce Survey is an annual survey designed to measure the extent of use of Information Communication Technologies (ICTs) (i.e. computers, internet ... ORDER FROM ESDS
Office for National Statistics Virtual Microdata Laboratory

- ONS and Other Government Departments’ data:
  - Business surveys and register
  - Social surveys
  - Price indices
  - Census CAMS
  - Longitudinal Study
- Logon to virtual desktop with access points in ONS and at various safe rooms in the UK
- VML is primarily for government researchers, private sector and overseas academics. SDS provides access for UK academics.
- Non-resident researchers with successful applications for ONS data under DwB can use the onsite safe rooms at the SDS or the VML
Contact information

E securedata@data-archive.ac.uk

W securedata.data-archive.ac.uk

T +44 (0)1206 874968

VML: maus@ons.gsi.gov.uk
Access to French microdata

- French microdata are available on different forms:
  - Public Use Files
  - Scientific Use Files
  - Confidential Files

- Almost every data collected for official statistics are available:
  - Except data covered by another secret: fiscal, national security, ...
Public Use Files

- Available at http://www.insee.fr/fr/bases-de-donnees/fichiers-detail.asp
- Description in French
- Mainly civil status data and Labour Force Survey
Scientific Use Files

- Data from household surveys available through French Data Archive: Reseau Quetelet

- [www.reseau-quetelet.cnrs.fr](http://www.reseau-quetelet.cnrs.fr)

- Website in English, but data description in French (to be translated ..)

- No SUF from business surveys
Confidential Files (1)

- Available from Business and Household Surveys and employer-employee data
  - Except data covered by another secret: fiscal, national security, ...

- Researchers submitted to the same penal sanctions as the statisticians

- Access through French national secure center: CASD in Genes
Confidential Files (2) : The SD-Box

Le boitier : la SD-Box™
Confidential Files (3)

- Procedures in 4 steps:
  - Opinion of then French Statistical Confidentiality Committee, chaired by a judge and composed by researchers, statisticians and representatives of surveyed persons
    - Evaluation of the scientific project
    - Need of the confidential data
  - Agreement by the statistical services
  - Decision by the French national Archives
  - Formation to the use of CASD, taking place in Paris
Confidential Files (4)

- Duration of the process: 3-4 months, to be shortened in the future

- Use of SD Box not free
  - Except for DwB sponsored projects

- More information available on:
  - [http://www.cnis.fr/cms/Accueil/activites/_trois_comites/Comite_du_secret_statistique](http://www.cnis.fr/cms/Accueil/activites/_trois_comites/Comite_du_secret_statistique)
    - Page in French, but request form in English
    - Page in English
Thanks for Listening

Contact: contact@dwbproject.org

Website: http://www.dwbproject.org/