

# Research Centers and Confidential Data



Ron Dekker

Scientific Statistical Agency  
and NIWI Data Archives

# Task of the Agency

- Open up relevant data for research
  - Contract
  - Data Protection
- Case study
  - Micro data of Statistics Netherlands (CBS)
- Generalize to other data

# Contract I

- Agreement between Agency - CBS
  - Micro data on individuals and households
- Tailor-made
  - Right to use the micro data
  - Different degrees of access
    - e.g. university vs. commercial research institutes
  - Different rules for accessing the data
    - check the organization, the user, the project

# Contract II

- Contract between CBS - User (Organization)
  - No redistribution of data
  - Hand over publications
    - beforehand or afterwards
  - Local safety procedures (at user organization)
  - Tariff
    - very low, but prevents data ordering without use
- Secrecy Statement
  - By individual user (and his/her boss)

# Data Protection

- Remove direct identifiers
- Process on revealing and/or sensitive variables
  - Indirect identifiers
    - like region, data of birth
    - 3 levels of variables: very, normal, slightly revealing
    - CBS has software to check on unique cells in a table
      - ARGUS (= ‘Suspicious’)
      - one for micro data and one for tables
  - Randomization Techniques
    - Post Randomization Method (for integer data)

# Result

- Scientific Use File on CD-ROM
  - Protected against spontaneous recognition
- On site facilities
  - Access to original (unprotected data)
- Does it work?

Yes it does

large part of social science faculties,  
planning bureaus, ministries, research institutes  
use CBS data

# Other data

- Other micro data on individuals
  - Remove direct identifiers
  - Check on the very identifying indirect identifiers
    - date of birth
    - full (6 digit) zip code ---> 4 digit code
- Micro data on firms
  - CBS: Not allowed to leave CBS office
    - On site facility at CBS office
  - Other producers
    - Rely on the contract (not on data protection)

# Other data (cont'd)

- Registers & combined data (indiv. & firm)
  - CBS: Data are closing down
    - Perhaps work on site
- Geographical data
  - Not possible to use data protection techniques
- Besides contract and data protection techniques
  - Mutual trust between researchers and producers
    - added with formal code with rules for 'good' research



# New Instruments

- On site at a ‘trusted third party’
  - $\approx$  Research Data Center
    - 1 Center for different data (producers)
    - especially for small data producers
- Remote access
  - Logging of all your activity (as a researcher)
  - (Semi-) automatic control on the log file

# Conclusions

- Contracts and Data protection techniques are good
- Trust is better