Statistical Disclosure Control – The Dutch Point of View

Eric Schulte Nordholt
Statistics Netherlands

e.schultenordholt@CBS.nl
Contents

- Introduction
- Legal issues
- The release of microdata
- Remote access
- Conclusions
Introduction (1)

Traditional output of a statistical office:
- Tables
- Graphs

But:
- Growing need for information
- Growing need for microdata
- Computing power
- Possibilities for analysis
Introduction (2)

First step:
- Larger tables
- Confidentiality risks (τ-ARGUS)
- Online publication of ‘tabular’ / aggregated data
  - First SDC protected, then safe tables (example: Dutch Statline)
  - First tables on request, then SDC protected (example: American Factfinder)
Introduction (3)

Second step:
- PUF (Public Use Files) Severe / heavy protection with μ-ARGUS (no fun for analyses)
- MUC (Micro data files Under Contract) moderate protection with μ-ARGUS; for Universities and recognized Research Institutes only; reliable researchers
Introduction (4)

Other options:
Work on-site in secure area (safe setting) or remote analyses
– Contract with researcher and university or research institute
– In addition to the standard statistical software packages special (e.g. own) software can be installed on request
– Control on each result taken home / published

National Statistical Institute (NSI) is bound by privacy protection (privacy authority)
Legal Issues (1)

Overview
– Ethical codes
– Laws (EU)
– Laws (The Netherlands)
Legal Issues (2)

Ethical codes
– American Statistical Association
– UN Fundamental principles of official statistics
– Conference of European Statisticians of the UNECE: principles and guidelines of good practice for managing statistical confidentiality and microdata access (http://www.unece.org/stats/documents/tfcm/1.e.pdf) and principles and guidelines on confidentiality aspects of statistical data integration (http://www.unece.org/stats/publications/Confidentiality_aspects_data_integration.pdf)
Legal Issues (3)

Laws (EU)


Legal Issues (4)

Laws (The Netherlands)
– Law on economic statistics 1936
– Law on protection of personal data 2001
– Law on Statistics Netherlands / CCS 2004
  – autonomous government agency
  – free access to public administrative data
  – statistical confidentiality as legal obligation
  – microdata access for statistical and scientific research
The release of microdata (1)

Four examples of microdata:

- Public use microdata files (‘for everybody’)
- Microdata under contract (‘for researchers’)
- Microdata for on-site
- Microdata for remote access / execution
The release of microdata (2)

Aspects of public use microdata files:
- Strict protection
- Educational purposes
- Data on persons or households

Aspects of microdata under contract:
- Partial statistical partial legal protection (contract)
- Heavily used by research community (1994 – 2006)
- Data on persons or households
The release of microdata (3)

Aspects of on-site:

- Researchers work in a secure area of the statistical institute
- Researchers can apply the standard statistical software packages and also bring their own programmes
- Researchers swear an oath to the effect that they will not disclose the individual information of respondents
The release of microdata (4)

Aspects of remote access / execution

- Combination of advantages of on desk and on-site
- Security risks are high, especially with remote execution (no intermediary between the researcher and the statistical institute)

Dutch variant of remote execution:
- Test dataset with metadata provided
- Set-ups tested on test dataset by the researcher
- Final set-up executed on real data and results sent to the researcher by Statistics Netherlands

- Remote access is becoming very popular in several countries
Remote access (1)

Remote access pilot in 2005 at Statistics Netherlands

Advantages of remote access
- at own institute
- 24/7 availability
- ability to play around with the data, without confidentiality checks until final output
- controlled safe settings

Disadvantages of on-site
- only at premises SN
- only working hours
- no direct contact with colleagues
- special offices needed
Remote access (2)

Only authorised users from selected research institutes allowed (under contract)

On site:
– Users cannot enter or leave Statistics Netherlands unaccompanied

Remote access:
– Biometric identification
  – Public Key Infrastructure (PKI) Certificates
  – Username and password
Remote access (3)

Detailed microdata stay at Statistics Netherlands

On-site:
- Network separate from production
- No internet
- No printer
- No CD-ROM or USB
- Desired output checked by Statistics Netherlands staff

Remote access:
- Network separate from production
- Citrix connection (on special PCs at the institute of the researcher)
- Desired output checked by Statistics Netherlands staff
Remote access (4)
Remote access (5)

Welkom

Welkom op de Remote Access pagina.

MEDEDELING:

De komende 2 maanden (juli en augustus) zal er, vanwege groot schaal ICT onderhoud bij het CBS, ieder weekend onderhoud worden gepleegd aan de Remote Access omgeving. Vanaf vrijdagavond 18:00 uur tot zaterdagavond 20:00 uur kan de faciliteit niet gebruikt worden.

Due to large-scale ICT maintenance, the coming 2 months (July and August) each weekend maintenance work of the Remote Access environment will be carried out. Therefore the Remote Access environment can not be used from Friday 18:00 o'clock until Saturday 20:00 o'clock.

Meldingen Center

Het Meldingen Center toont alle informatie of foutmeldingen die op kunnen treden.
Remote access (6)

Pilot:
University of Tilburg
Access for some staff members of Statistics Netherlands

Experiences:
Positive, no real problems
Performance like own desktop PC

Current situation:
Most Dutch universities have remote access
Many research institutes have remote access
Conclusions

Remarks
– Remote access is a promising counterpart of on-site
– Confidentiality issues ‘under control’
– Check of output labour intensive but needed
– Since 2006: remote access on a larger scale
– It has replaced MUCs to a large extent

Future options
– ESSnet on Decentralised and Remote Access to Confidential Data in the ESS attempts at combining on-site facilities
– Data Without Boundaries project: NSIs and Data archives work together
– OECD, Initiatives outside Europe (e.g. North America, Australia)
– Linking up with the world: connect remote access facilities between countries (The Netherlands, Denmark, Sweden and Finland are ready, will Germany follow?)