

Legal framework and current access facilities

Eric Schulte Nordholt

Contents

- Introduction
- Legal issues
- Practices
- The release of microdata
- Statistical services and information
- On-site
- Remote access
- Conclusions and remarks

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 (examples: American Factfinder of the United States Census Bureau, TableBuilder of the ABS)

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

Other name for MUC: Scientific Use File

Introduction (4)

Other options (sometimes called Secure Use Files):

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

- International Statistical Institute
(<http://isi.cbs.nl/ethics.htm>)
- American Statistical Association
(<http://www.amstat.org/about/ethicalguidelines.cfm>)
- European Statistics Code of Practice
(<http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>)
- UN Fundamental principles of official statistics
(<http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>)

Legal issues (3)

Ethical codes continued

- 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 (4)

Laws (EU)

- Commission Regulation (EC) No 223/2009 of the European Parliament and Council of 11 March 2009 on European statistics (establishes the legal framework for the development, production and dissemination of European statistics, including the rules on confidentiality)

Legal issues (5)

Laws (EU) continued

- Commission Regulation (EU) No 557/2013 of 17 June 2013 implementing Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics as regards access to confidential data for scientific purposes and repealing Commission Regulation (EC) No 831/2002

Legal issues (6)

Laws (The Netherlands)

(<https://www.cbs.nl/en-gb/about-us/organisation>)

- Law on economic statistics 1936
- Personal Data Protection Act 2001 → GDPR
(<https://www.eugdpr.org/eugdpr.org.html>)
- Statistics Netherlands Act 2004
 - autonomous government agency
 - free access to public administrative data
 - statistical confidentiality as legal obligation
 - microdata access for statistical and scientific research

Practices (1)

Research access to microdata - advantages

- Standardization of concepts and questions
- Standardization of results
- Feedback on data quality (co-operation in strategic research program)
- Cost reduction
- Reduction response burden

Practices (2)

Research access to microdata - disclosure problems

- Matching
- Response knowledge
- Spontaneous recognition

Practices (3)

Research access to microdata - solutions

- legal authority as exception to confidentiality
- restricted to researchers only, under contract
- safe data (masking, swapping, sub-sampling, synthetic data) versus safe settings (rules, contract, stand alone, logging)
- microdata often only from social surveys
- on-site access for economic surveys, registration data and special data (e.g. causes of death)
- remote access on a large scale

Practices (4)

Difference between social and economic data

- obligation to participate (most business surveys) or not (most household surveys)
- size of population
- sampling fraction
- shape of basic distributions
- public visibility of extreme cases
- vulnerability

Practices (5)

Difference between administrative and sample survey data

- sample size
- nonresponse
- measurement error
- cost
- timeliness
- ease of matching

The release of microdata (1)

Four examples of microdata:

1. Public use microdata files ('for everybody')
2. Microdata under contract ('for researchers')
3. Microdata for on-site analyses
4. Microdata for remote access

Example 2 is a Scientific Use File; examples 3 and 4 are Secure Use Files

The release of microdata (2)

Aspects of public use microdata files:

- Strict protection
- Educational purposes
- Data on persons (or households)

Aspects of microdata for researchers:

- Partial statistical partial legal protection (contract)
- Heavily used by research community (1994 – 2006)
- Data on persons or households

Availability of Statistics Netherlands public use microdata files and microdata for researchers via data archive (that has a different role than NSI)

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 have to sign a confidentiality warrant to the effect that they will not disclose individual information of respondents

The release of microdata (4)

Aspects of remote access:

- Combination of advantages of on desk and on-site
- Security risks are high, especially with remote execution (traditionally no intermediary between the researcher and the national statistical institute; therefore sometimes heavily disturbed or synthetic data are made available)
- Remote access is becoming very popular in several countries

Statistical services and information (1)

Started in 2002 to provide ministries with optimal statistical information (simple version of remote execution)

Now: special unit of Statistics Netherlands with about 100 employees

Ministries are still important, but other users get more and more attention

Statistical services and information (2)

Services:

- Research projects based on existing data at Statistics Netherlands (additional to the regular working program), results are published at the (Dutch part of the) website of Statistics Netherlands
- Statistical advice which data to use to answer which question and how to use them (incl. which classifications and definitions to choose)
- Remote access and on-site (costs for documenting data; for on-site also costs for a work station)

Statistical services and information (3)

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

On-site:

- Register data, survey data or a combination (linking by Statistics Netherlands or researcher)
- Researcher works at Statistics Netherlands
- Analyses results checked by two Statistics Netherlands' employees independently and sent electronically to the (institute of the) researcher

Statistical services and information (4)

Rules for remote and on-site:

- Agreement of DG is needed
- Contract with the organization and both the researcher and his superior have to sign a confidentiality warrant
- The researcher gets only access to the microdata needed for his project

Statistical services and information (5)

SDC checks:

- Prevent disclosure of individual personal or enterprise information
- Labour intensive work: the possibly unsafe information has to be checked manually

Statistical services and information (6)

Conclusions:

- On-site meets many needs
- No real disclosure problems encountered
- Remote access has the future for most researchers

On-site (1)

The On-site room

- Standard PCs in a Windows network
- No rights to access the rest of Statistics Netherlands' network
- No internet, printer, CD-ROM or USB
- MS-Office suite, SPSS, SAS, R
- Own software on request (own license)
- Fee for the use of the data (preparation) and the use of the PCs

On-site (2)

Approved researchers = persons employed by approved institutes

- **Institutions mentioned in the law on Statistics Netherlands**
 - Universities
 - Planning agencies
 - Other European NSIs and Eurostat
- **Institutions admitted (e.g. research departments of ministries and bonafide research bureaus) under the following conditions**
 - Legal corporate status
 - No direct administrative authority
 - Explicit research aim
 - Publish for public use
 - A good reputation

On-site (3)

Enterprise data:

- Old strict law on economic statistics 1936 forbids 'everything'
- Strong need for research
- Long negotiations with all parties
- On-site access options since 1998
- Since 2005 part of Statistical services and information unit

On-site (4)

On-site international

Some initiatives:

- **USA / Canada**
- **Australia / New Zealand**
- **Europe**
- **Eurostat**

Examples are there to be followed

Remote access (1)

Denmark was the first country to introduce remote access, others followed (e.g. the Netherlands, Sweden, Finland, Estonia, Hungary)

New Nordic model for researchers joint access to data from the Nordic Statistical Institution (see also <http://www.unece.org/stats/documents/2015.10.confidentiality.html>):

Social microdata (excluding health data) from six Nordic countries can be accessed via the remote access facilities from Denmark, Finland or Sweden

Remote access (2)

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 of SN
- 😊 only working hours
- 😊 no direct contact with colleagues
- 😊 special offices needed

Remote access (3)

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

On-site:

- Username and password
- Users can neither enter nor leave Statistics Netherlands unaccompanied

Remote access:

- Username and password
- Instead of biometric identification, we now send TAN (Transaction Authorisation Number) codes to mobile phones

Remote access (4)

Detailed microdata stay at Statistics Netherlands

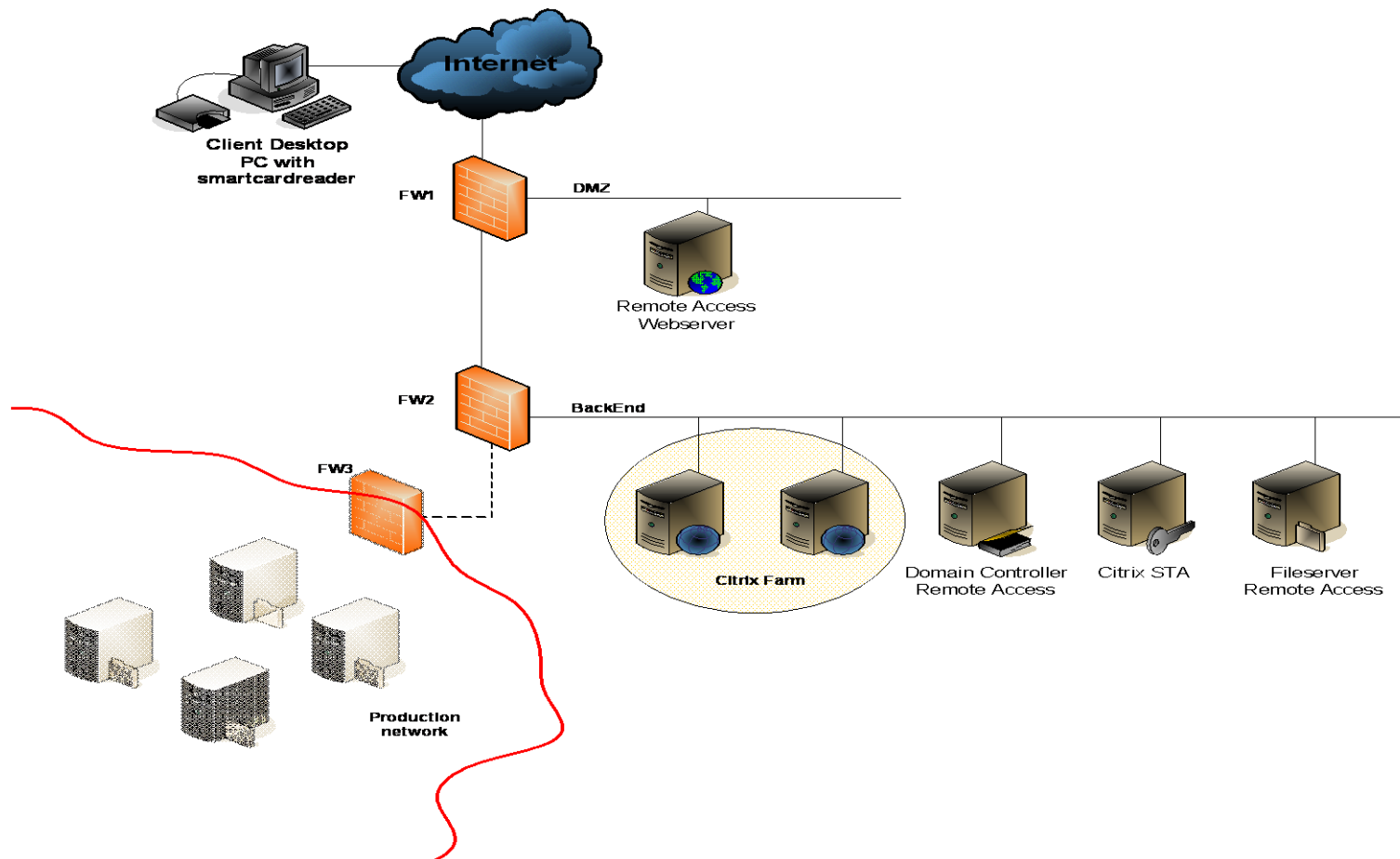
On-site:

- 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 (5)



Remote access (6)

Experiences of pilot with University of Tilburg:

- Positive, no real problems
- Performance like own desktop PC

Current situation:

- Most universities have remote access
- Many research institutes have remote access

Conclusions and remarks

- Legal framework allows data access
- Building a relation of trust
- Confidentiality issues 'under control'
- PUFs and MUCs become less popular
- Remote access has become a counterpart of traditional on-site
- Check of output is labour intensive but needed
- Linking up with the world: connect remote access facilities between countries