

# Confidentiality in data integration

**Eric Schulte Nordholt**  
Senior researcher and project leader of the Census  
Statistics Netherlands  
Division Social and Spatial Statistics  
Department Support and Development  
Section Research and Development

[ESLE@CBS.NL](mailto:ESLE@CBS.NL)

Presentation in Luxembourg  
December 2008



## Contents

- UNECE Task Force on Confidentiality and privacy aspects of statistical data integration
- Principles on confidentiality aspects of statistical data integration
- Introduction to the ESSnet on Integration of Surveys and Administrative Data (ISAD)
- Activities of WP1
- Content
- Practical examples



## UNECE Task Force on Confidentiality and privacy aspects of statistical data integration (1)

### Objective of the Task Force

Provide a common framework for assessing and mitigating legislative and other confidentiality aspects of the creation and use of integrated datasets for statistical and associated research purposes

### Task Force participants

Brian Pink (Task Force Chair), international organisations (Eurostat, UNECE) and NSIs (Canada, Finland, the Netherlands)



3

## UNECE Task Force on Confidentiality and privacy aspects of statistical data integration (2)

### ECE fundamental principles of official statistics number 6

Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes



4

## UNECE Task Force on Confidentiality and privacy aspects of statistical data integration (3)

### Definition of data integration

Data integration is concerned with integrating unit record data from different administrative and/or survey sources to compile new official statistics which can then be released in their own right. In addition these integrated data sets may be used to support a range of economic and social research not possible using traditional sources.



5

## UNECE Task Force on Confidentiality and privacy aspects of statistical data integration (4)

### Contents and benefits of data integration

Integration of data may include exact matching, probabilistic matching and/or statistical matching. The benefits of integrated data sets include:

1. production of new or enhanced statistics
2. production of more disaggregated information for measures where some information currently exists
3. ability to carry out research using integrated data sets
4. potential to improve or confront existing data sources
5. potential to reduce respondent burden



6

## UNECE Task Force on Confidentiality and privacy aspects of statistical data integration (5)

### Key conclusion of the Task Force

Integration of statistical data sets has become a normal part of the operations of a number of statistical offices and is generally most advanced in those countries where a heavy reliance is placed on obtaining statistical information from administrative registers

### Aim of the work of the Task Force

Provide some clarity and consistency of application of integrating data from different sources for statistical and related research purposes



7

## Principles on confidentiality aspects of statistical data integration (1)

### Principle 1

Data integration must comply with a country's legislative requirements.

### Principle 2

National Statistical Institutes (NSI's) must undertake data integration strictly exclusively for accomplishment of their official mandates.

### Principle 3

Data integration should only be considered when the public good is served, notably when it can be demonstrated to reduce costs, increase quality or minimise compliance load.



8

## Principles on confidentiality aspects of statistical data integration (2)

### Principle 4

Data integration benefits must clearly outweigh any privacy concerns about the use of data and risks to the integrity of the official statistics system.

### Principle 5

Data integration must not occur when it will materially threaten the integrity of the source data collections.

### Principle 6

Data must not be integrated where any undertaking has been given to respondents that would preclude such action.



9

## Principles on confidentiality aspects of statistical data integration (3)

### Principle 7

Integrated data must only be used for approved statistical or related research purposes.

### Principle 8

The number of unit records and data variables to be included in a linked dataset must be no more than required to support the approved purpose(s).

### Principle 9

Data integration must be conducted openly, including a brief outline of each data integration project being published on the NSI website.



10

## Principles on confidentiality aspects of statistical data integration (4)

### Guidelines

Principles are clarified

### Business case

Purpose(s), Benefit to Official Statistics, Other benefits, Retention of the integrated dataset, Data sources, Alternatives, Stakeholders, Retention of names and addresses, Frequency of reviews and Privacy impact assessment (if legally necessary)



11

## Introduction to the ESSnet on Integration of Surveys and Administrative Data (ISAD) (1)

### Participating national statistical institutes

Austria, Czech Republic, Italy (project leader: Mauro Scanu), the Netherlands and Spain

### Content

A review of recent literature was performed and an investigation was carried out on current practices and methods adopted in the five participating countries (AT, CZ, ES, IT and NL) and other countries



12

## Introduction to the ESSnet on Integration of Surveys and Administrative Data (ISAD) (2)

### Objective of WP1

Collect information about the state-of-the-art in  
the area methodologies for integration of surveys  
and administrative data

### Website

<http://cenex-isad.istat.it/>



13

## Activities of WP1

1. Literature review on probabilistic record linkage
2. Literature review on statistical matching
3. Literature review on micro integration processing
4. Practical experiences
5. Survey on the use of integration methodologies in the different ESS countries



14

## **Content (1)**

### **The first three chapters**

**Methodological literature reviews on probabilistic record linkage, statistical matching and micro integration processing**

**Collection of references to enable orientation through the literature on methods for integration**



15

## **Content (2)**

### **Chapter four**

**Examples of integration experiences: detailed information on many different projects that need a joint use of two or more sources in the participating NSIs of this ESSnet**



16

## Content (3)

### Chapter five

Results of a survey on the use and/or development of integration methodologies in the ESS countries

### Annex

Survey with the glossary



17

## Practical examples

- EU-SILC
- Business demography data
- SBS
- Business Register
- Population and Housing Census
- Population statistics
- LFS and TUS



18