

# Statistical Disclosure Control

a.k.a. Statistical Disclosure Limitation

General introduction



## Contents

- Confidentiality
- What is Statistical Disclosure Control?
- Five key stages to confidentiality
  - Why is SDC needed?
    - (Increasing) need for SDC
    - General SDC issues
      - R-U map
  - Data characteristics and uses

## Confidentiality

- General definition of confidential data:
- Data that can not be published as such
  - By law (e.g., statistical law)
  - Sensitive data (what's sensitive?)
  - Respondent considers it confidential

## What is SDC?

Physical protection

- Entrance
- Network

Legal protection

- Oath

*Statistical Disclosure Protection*

- *Protection of statistical output*

## Five key stages

1. Why is confidentiality protection needed?
2. What are the key characteristics and uses of the data?
3. What disclosure risks need to be protected against?
4. Disclosure control methods
5. Implementation

## Why is SDC needed?

- Is SDC really needed in this specific case?
- What kind of information can be deduced?
  - Sensitive information?
  - Publicly known information?
    - Freely available?
- Group disclosure or statistics?



## Need for SDC

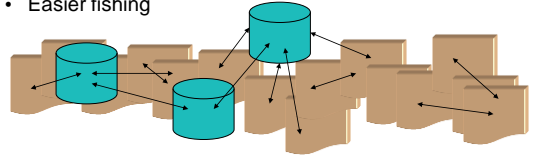
- Laws
  - International (EU)
  - National (The Netherlands)
- Agreement with respondent
  - Current response
  - Future response
- Agreement with owners of registrations

General introduction

7

## Increasing need for SDC

- Linking
  - Registrations
  - Surveys
- Dataming techniques
- Easier fishing



General introduction

8

## Increasing need for SDC

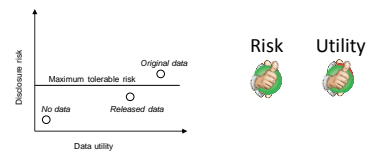
- Change of data characteristics
  - Availability
    - Administrative registrations
    - Powerful computers
  - Amount of detail
  - Actuality

General introduction

9

## General SDC issues

- Data should be safe: disclosure risk should be low
  - Zero risk is not really possible!
- Data should be useful: utility should be high



General introduction

10

## Data characteristics and uses

- Type of data
  - Full population
  - Sample
- Meta information
  - Sampling design
  - Response, coverage
- Type of variables
  - Categorical
  - Continuous
- Type of output
  - Microdata file
  - Tabular data

General introduction

11