

# A practice guide for microdata anonymization

Session 1.2

Thijs Benschop and Matthew Welch

UNECE/Eurostat Work Session on Statistical Data Confidentiality  
The Hague, The Netherlands, 29–31 October 2019


# Background

- ▶ Requirements for SDC practice for microdata
  - Technical tools (sdcMicro<sup>1</sup>, mu-Argus<sup>2</sup>)
  - Guidelines (theory/knowledge and practice/experience)
- ▶ World Bank / International Household Survey Network
  - contributed funding for the development of sdcApp (GUI for sdcMicro) and sdcMicro
  - conducted experiments to evaluate anonymization techniques
  - consolidated knowledge and experience in a practice guide

<sup>1</sup> Matthias Templ, Alexander Kowarik, Bernhard Meindl (2015). Statistical Disclosure Control for Micro-Data Using the R Package sdcMicro. Journal of Statistical Software, 67(4), 1–36. <doi:10.18637/jss.v067.i04>

<sup>2</sup> mu-Argus version 5.1.3 <http://neon.vb.cbs.nl/casc/mu.htm>

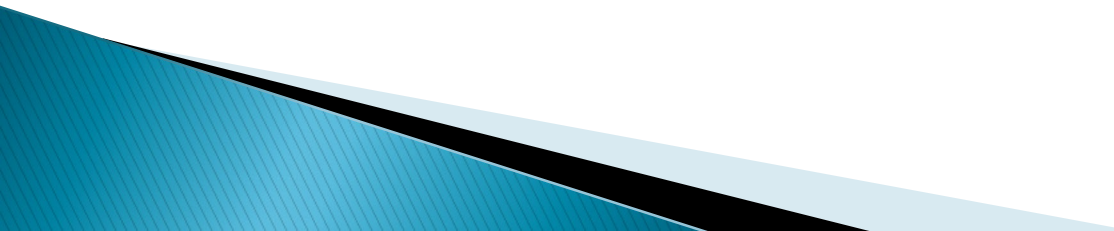
# Practice guide

- ▶ Handbook making practice of SDC accessible
  - ▶ Consolidated knowledge on SDC for (survey) microdata
    - Literature review
    - Experiments on real data
    - Practice in other countries
    - Case studies
  - ▶ Target audience
    - producers of official statistics in developing countries
    - useful for any data producer who wants to release microdata
  - ▶ Practical applications in R with sdcMicro
- 

# Three guides

- ▶ Three guides for Statistical Disclosure Control for Microdata
  1. A Theory guide
  2. A Practice Guide for sdcMicro
  3. sdcApp manual
- ▶ All three guides available on ReadTheDocs and GitHub (also as pdf)
  - [sdctheory.readthedocs.io/](https://sdctheory.readthedocs.io/)
  - [sdcpractice.readthedocs.io/](https://sdcpractice.readthedocs.io/)
  - [sdcappdocs.readthedocs.io/](https://sdcappdocs.readthedocs.io/)

# Theory guide

- ▶ Introduction to SDC process and theory
  - ▶ Self-contained guide covering all aspects of SDC process
  - ▶ Includes references to seminal papers
  - ▶ Many examples of practitioners provide insights in SDC practice
- 

## TABLE OF CONTENT

Introduction

Glossary and list of acronyms

Statistical Disclosure Control (SDC): An Introduction

Release Types

Measuring Risk

Anonymization Methods

Measuring Utility and Information Loss

The SDC Process

Case Studies (Illustrating the SDC Process)

Appendices

Acknowledgements

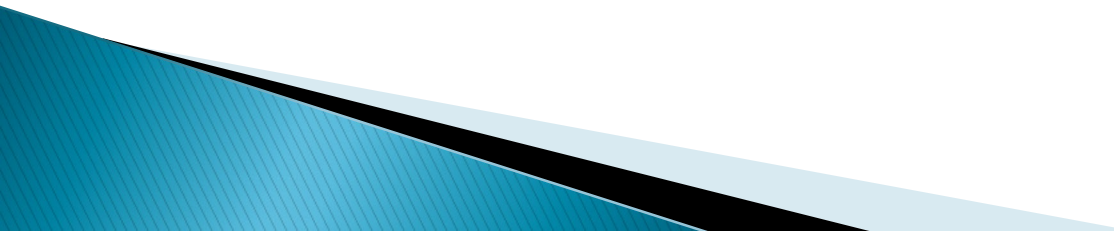
# Statistical Disclosure Control for Microdata: A Theory Guide


## Table of Content

- [Introduction](#)
  - [Building a knowledge base](#)
  - [Using this guide](#)
  - [Statistical software tools for technical implementation](#)
  - [Outline of this guide](#)
- [Glossary and list of acronyms](#)
  - [List of Acronyms](#)
  - [Glossary](#)
- [Statistical Disclosure Control \(SDC\): An Introduction](#)
  - [Need for SDC](#)
  - [The risk-utility trade-off in the SDC process](#)
- [Release Types](#)

<https://sdctheory.readthedocs.io/>

# A Practice Guide

- ▶ Provides user guidance in applying SDC using *sdcMicro* from *R* command-line
  - ▶ Covers all steps in the SDC process that can be carried out with *sdcMicro* in *R*
  - ▶ Includes intro to *R* and code snippets
  - ▶ Includes case studies fully documented in *R*
- 

 SDC Practice Guide

latest

Search docs

TABLE OF CONTENT

Introduction

Glossary and list of acronyms

Statistical Disclosure Control (SDC): An Introduction

Release Types

Measuring Risk

Anonymization Methods

Measuring Utility and Information Loss

SDC with *sdcMicro* in R: Setting Up Your Data and more

The SDC Process

Appendices

Acknowledgements

CASE STUDIES

Case Studies (Illustrating the SDC Process)

# Statistical Disclosure Control for Microdata: A Practice Guide

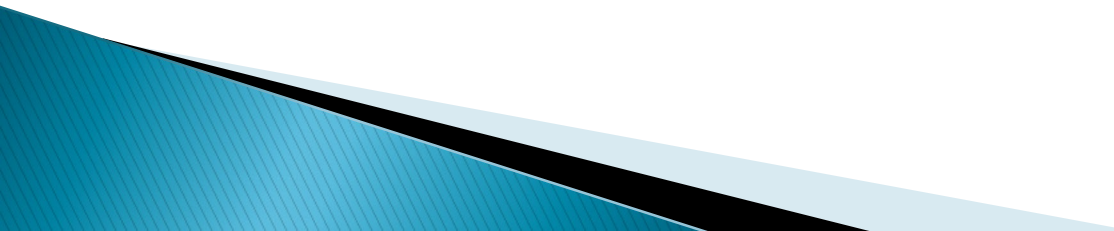
## Table of Content


- Introduction
  - Building a knowledge base
  - Using this guide
  - Outline of this guide
- Glossary and list of acronyms
  - List of Acronyms
  - Glossary
- Statistical Disclosure Control (SDC): An Introduction
  - Need for SDC
  - The risk-utility trade-off in the SDC process
- Release Types
  - Conditions for PUFs
  - Conditions for SUFs
  - Conditions for microdata available in a controlled research data center
- Measuring Risk
  - Types of disclosure

<https://sdcpractice.readthedocs.org/>



# sdcApp manual

- ▶ Provides user guidance in applying SDC using the *sdcApp*
  - ▶ *sdcApp* is the GUI for *sdcMicro* for users not familiar with *R*
  - ▶ Includes complete installation instructions
  - ▶ Covers all steps in the SDC process that can be carried out with *sdcApp*
  - ▶ Includes case studies fully documented using *sdcApp*
- 

 sdcMicro GUI manual

latest

Search docs

TABLE OF CONTENTS

Introduction

Installation and updating

Introduction to sdcApp

Loading and Preparing Data

Setup anonymization problem

Risk measurement

Anonymization methods

Utility measurement

Export data and reports

Reproducibility

Undo

Case Studies (Illustrating the SDC Process)

[Docs](#) » sdcApp manual

[Edit on GitHub](#)

## sdcApp manual

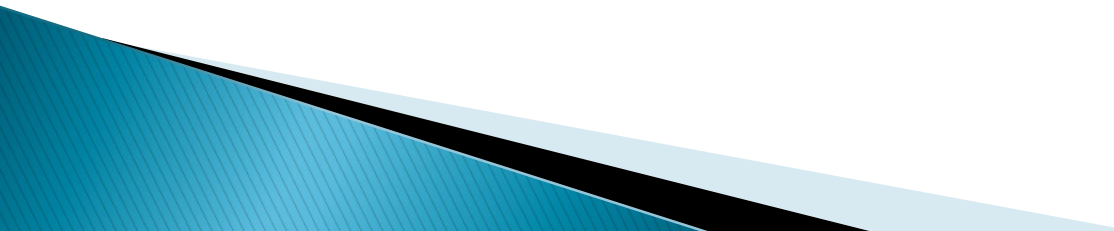
This is documentation and guidance for *sdcApp*, a user interface for the *sdcMicro R* package, which provides tools for Statistical Disclosure Control (SDC) for microdata, also known as microdata anonymization.

### Table of contents

- [Introduction](#)
  - [What is sdcApp?](#)
  - [Statistical Disclosure Control \(SDC\)](#)
  - [What is the purpose of the manual?](#)
  - [Background literature on SDC for microdata](#)
  - [Outline of this guide](#)
  - [References](#)
- [Installation and updating](#)
  - [Installing R and RStudio](#)
  - [Installing sdcMicro package](#)

<https://sdcappdocs.readthedocs.org/>

# Using the guides

- ▶ CC BY 4.0 license
  - ▶ Free to use
  - ▶ Input, corrections, comments
    - ReadTheDocs living documents
    - GitHub
  - ▶ Future work
- 

# Feedback and questions

- ▶ Contacts

- Matthew Welch, [mwelch@worldbank.org](mailto:mwelch@worldbank.org)
- Thijs Benschop, [tbenschop@worldbank.org](mailto:tbenschop@worldbank.org)

- ▶ Any comments and feedback welcome