Using Administrative and Secondary Sources for Official Statistics:

A Handbook of Principles and Practices

UNITED NATIONS

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

Using Administrative and Secondary Sources for Official Statistics

A Handbook of Principles and Practices

UNITED NATIONS
New York and Geneva, 2011

Note

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontier or boundaries.

Acknowledgement

The UNECE would like to acknowledge the valuable contributions of more than two hundred participants of an international training course on the use of administrative sources for statistical purposes, and particularly the inputs from the presenters named below, all of which have helped considerably to enrich this handbook.

Mr Driss Afza, MEDSTAT projects
Ms Sue Fendall, Office for National Statistics, United Kingdom
Ms Riitta Harala, Statistics Finland
Mr John C Hughes, Office for National Statistics, United Kingdom
Mr Ben Humberstone, Office for National Statistics, United Kingdom
Mr Pekka Myrskylä, Statistics Finland
Ms Kaija Ruotsalainen, Statistics Finland

ECE/CES/13

Foreword

Statistical organisations around the world are coming under increasing pressure to improve the efficiency of the statistical production process, and particularly to make savings in costs and staff resources. At the same time, there are growing political demands to reduce the burden placed on the respondents to statistical surveys. This is particularly the case where respondents are businesses, as many governments see reducing bureaucracy as a key measure to support and promote business development.

Given these pressures, statisticians are increasingly being forced to consider alternatives to the traditional survey approach as a way of gathering data. Perhaps the most obvious answer is to see if usable data already exist elsewhere. Many non-statistical organisations collect data in various forms, and although these data are rarely direct substitutes for those collected via statistical surveys, they often offer possibilities, sometimes through the combination of multiple sources, to replace, fully or partially, direct statistical data collection.

The degree of use of administrative sources in the statistical production process varies considerably from country to country, from those that have developed fully functioning register-based statistical systems, to those that are just starting to consider this approach.

Although several subject specific texts exist, there have, until now, been no general, international methodological guidelines to help those in the early stages of using administrative data. This handbook aims to fill that gap. It builds on material developed over ten years in the context of an international training course on the use of administrative sources for statistical purposes. That course has now been delivered over ten times, to audiences of official statisticians from throughout Europe, Western and Central Asia, and North Africa.

Each time the course has been run, it has been improved and enhanced by sharing experiences with, and receiving feedback from participants. It has also benefited greatly from the input of various expert guest presenters from Statistics Finland and the British Office for National Statistics.

Mr Steven Vale, UNECE, Course leader

Contents

FOREWORD III

CONTENTS III

NOTES VI

1. WHAT ARE ADMINISTRATIVE AND SECONDARY SOURCES? 1

2. THE ADVANTAGES OF USING ADMINISTRATIVE SOURCES 7

3. FRAMEWORKS FOR ACCESS TO ADMINISTRATIVE SOURCES 11

4. COMMON PROBLEMS AND SOLUTIONS 19

5. QUALITY AND ADMINISTRATIVE DATA 37

6. DATA LINKAGE AND MATCHING 43

7. USING ADMINISTRATIVE DATA IN STATISTICAL REGISTERS 59

8. USING ADMINISTRATIVE DATA TO SUPPLEMENT STATISTICAL SURVEYS 69

9. TOWARDS A REGISTER-BASED STATISTICAL SYSTEM 75

Notes

1) Note on References

This handbook includes many references to other papers, web sites and publications. To help those who want to follow-up these references, Internet addresses are given wherever possible. These were all checked at the time of writing, but that is no guarantee that they will still work at the time of reading. If the reader finds a broken link, please report it to support.stat@unece.org.

2) Note on Exercises

The exercises at the end of Chapters 6 and 7 are taken from the course on which this handbook was based. They are included here as practical examples to reinforce the theory presented in those chapters.

1. What are Administrative and Secondary Sources

1.1 Introduction

Before we start to consider the practicalities of using data from administrative and secondary sources, it is worth just taking some time to clearly define what these terms mean. Several definitions exist in the literature currently available, the most relevant of which are examined in this chapter. The chapter ends by proposing a relatively simple and broad definition, which is then used as the basis for the remainder of this handbook.

1.2 Traditional Definitions

Administrative sources have traditionally been defined as collections of data held by other parts of government, collected and used for the purposes of administering taxes, benefits or services. Perhaps the most comprehensive of the traditional definitions was set out by Gordon Brackstone of Statistics Canada in his 1987 paper "Statistical Issues of Administrative Data: Issues and Challenges"[1]. Brackstone identified four distinguishing features of administrative data:

1. The agent that supplies the data to the statistical agency and the unit to which the data relate are different (in contrast to most statistical surveys);
2. The data were originally collected for a definite non-statistical purpose that might affect the treatment of the source unit;
3. Complete coverage of the target population is the aim;
4. Control of the methods by which the administrative data are collected and processed rests with the administrative agency.

This definition is broadly in line with that proposed by the Statistical Data and Metadata eXchange (SDMX) initiative[2]:

> *"A data holding containing information collected and maintained for the purpose of implementing one or more administrative regulations."*

During 1996-97 an internal Eurostat task force examined ways to better coordinate work relating to the use of administrative sources across different domains of statistics. This task force used a simple typology of data sources to consider how administrative sources should be defined. Firstly all data sources were divided into primary sources (data collected for statistical purposes) and secondary sources (all other data). A traditional or "narrow" definition of administrative sources comprises just public sector non-statistical sources, whereas a wider definition would also include private sector sources.

[1] Brackstone G J: "Statistical Issues of Administrative Data: Issues and Challenges", in "Statistical Uses of Administrative Data -An International Symposium", organised by Statistics Canada, 23-25 November 1987 (Proceedings published by Statistics Canada, Ottawa, December 1988).
[2] See: www.sdmx.org

The wider approach is consistent with the definition of administrative data adopted by the Conference of European Statisticians in the publication “Terminology on Statistical Metadata”[3]:

> *“Data collected by sources external to statistical offices.”*

The narrow and wider definitions can be shown graphically as follows:

Figure 1.1 - Narrow definition

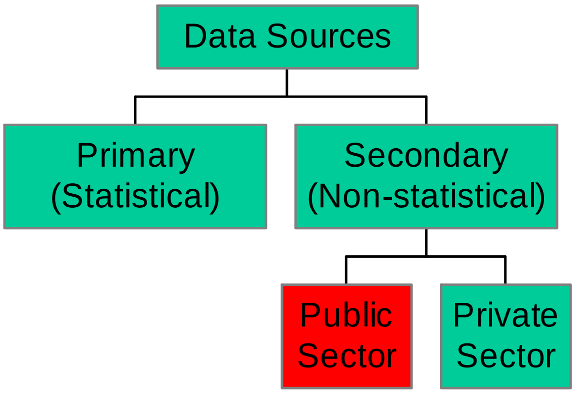
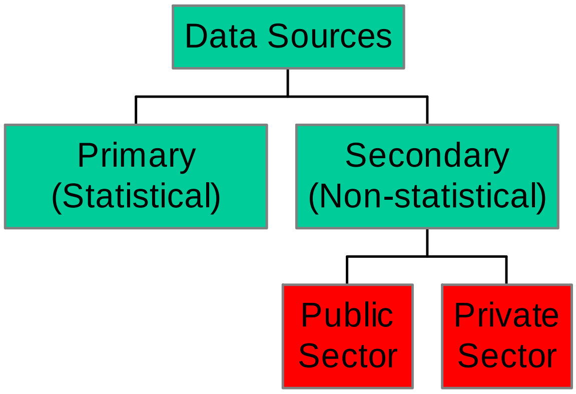


Figure 1.2 - Wider definition



Thus under the narrow definition, administrative sources are a sub-set of secondary sources, whilst under the wider definition these terms are synonyms.

There are a growing number of reasons for favouring the wider definition, including:

- ***Increasing privatisation of government functions:***

In several countries, regulatory functions that used to be carried out by government departments or agencies are being transferred to private or semi-private organisations. Typical examples are usually in the health, education or public utilities sectors, where former state monopolies are increasingly being replaced by private companies or non-profit institutions.

Registration functions, including the operation of administrative registers on behalf of government departments are also under consideration for privatisation in several