

# Using administrative data for Social Statistics

12<sup>th</sup> SPECA Working Group on Statistics

24-26 October 2017

Dushanbe, Tajikistan

Harald Utne, Statistics Norway

[harald.utne@ssb.no](mailto:harald.utne@ssb.no)

# Contents

- Statistical systems
- Concepts
- Preconditions for use of administrative sources in production of statistics
- Cooperation
- Quality issues (mainly in census presentation)
- Documentation and metadata
- Coordination
- Concluding remarks

# Statistical systems

- Based on aggregated data
- Based on individual data

# Statistical system based on aggregated data

- Many countries have a decentralized system for the production of statistics with several geographical levels
  - Municipal, County, Regional, National
- Method: produce statistics at each level and pass the results on to the next
  - Averages and number of observations (weights) from the regions
    - ♦ Example: Average wages and the number of employees
- Most statisticians in the national office work on aggregated data

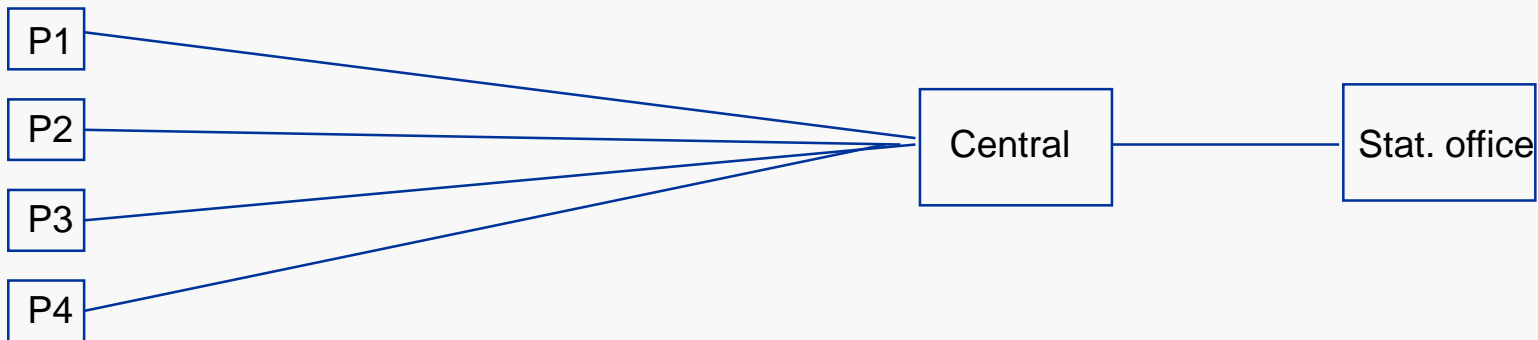
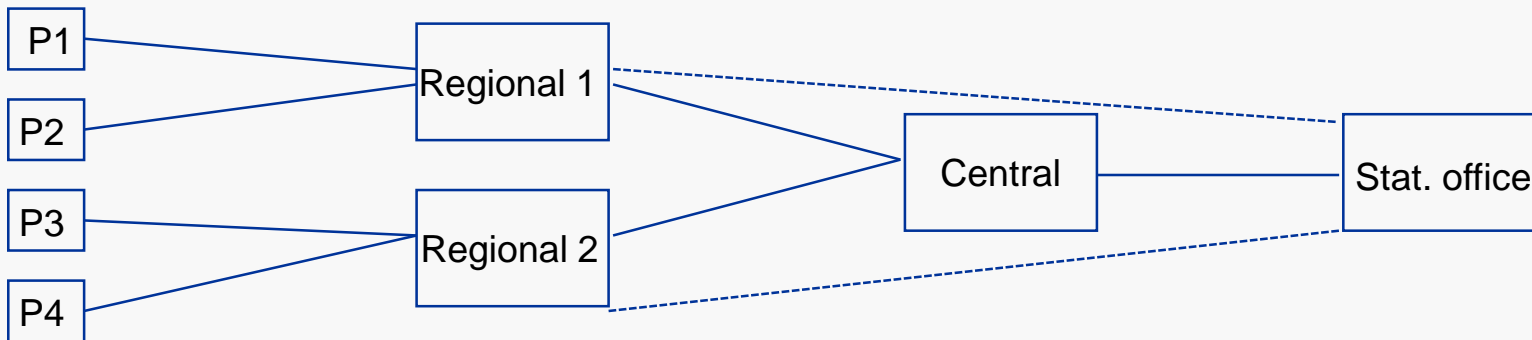
# Statistical system based on individual data

- Individual data for the whole population
  - From administrative registers
- Most statisticians in the national statistical office work directly with individual data

# Centralized or decentralized registers

Adm. registers

Stat. registers



# Centralized register system – advantages

- Centralized system for ID-numbers
- No double counting (unique ID-numbers)
- More unified registration
- More efficient data collection
- More easy to link data from several registers

# Nordic report on best practices for register-based statistics

- Nordic countries have been working with register-based statistics for fifty years
- Nordic report on best practices published by UNECE in 2007
  - [http://www.unece.org/fileadmin/DAM/stats/publications/Register\\_based\\_statistics\\_in\\_Nordic\\_countries.pdf](http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf)
  - [http://www.unece.org/fileadmin/DAM/stats/publications/Register\\_based\\_statistics\\_in\\_Nordic\\_countries\\_Russian.pdf](http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries_Russian.pdf)
- Aim: Collect main experiences from Nordic countries in **one** report
- Target group: Statisticians and decision makers in countries that are developing or planning for register-based statistics
- All kinds of register-based social statistics, but focus on censuses



# UNECE Guidelines on the use of registers and administrative data for population and housing censuses

- Draft report being prepared by UNECE task force
  - Presented at the Group of Experts for Population and Housing Censuses meeting in Geneva 4-6 October
- Expected endorsement by CES plenary session in April 2018
- More focus on combined censuses than the Nordic report

# Why register-based statistics?

- Budget restriction. Censuses or other full coverage surveys may be regarded as too expensive
- Register-based statistics: Use data already collected for administrative purposes
  - Data collection almost free of charge (for the NSO)
- Instead of recollecting data for statistical purposes, the NSO uses resources on improving existing data
  - Supplement and correct existing (administrative) data
- Compromise between costs, response burden and product quality
- Register-based statistics: Lower cost and lower response burden

# Use of administrative data in production of statistics

- Direct use of register data: Register-based statistics
- Register data used as a supplement
  - Sampling frames
  - Editing, coding and imputation for non-response
  - Additional information from registers
    - ◆ Reducing response burden and costs
    - ◆ More accurate data (income)

# Definition of register

- Systematic collection of unit level data organized in such a way that updating is possible

# Updating

Processing unit level information to keep track of changes in units and their attributes

- New units (new born, immigrants) added
- Exiting units (dead, emigrants) “removed”
  - Classified as “not active”
- Real changes in attributes
- Corrections
- Updating only for units that have undergone changes
- Updating when changes arise (“continuously”)
  
- A traditional census file is **not** a register: All data are collected on unit level for a point in time or a period
  - New data collected for next census
  - Not updated as a register

# Preconditions for use of administrative sources in production of statistics

- Legal base
- Public approval
- Unified identification system
- Comprehensive and reliable register system
- Cooperation among authorities

# Legal base

- National legislation must reflect the idea of re-use of administrative data for statistical purposes
- Legislation must give the NSO
  - Access to administrative data on a unit level, including identification information
  - Right to link data from different administrative sources
- Legislation must be up to data
  - Statistics Act of Norway from 1989
  - Replacing the former act of 1907
  - New act is under preparation

# Legislation on data protection

- Act of processing of personal data
- Legal rights of citizens must not be violated
- Data processed in NSO = Statistical data
  - Not for external use at an individual level
- In Norway: Principle of “One way traffic”
- Administrative data for statistical purposes
- Statistical data for
  - Statistical purposes
  - Research purposes (regulated by Statistics Act)
  - **Never for administrative purposes**



# Public approval

Attitudes may be different between countries concerning

- Existence and linking of administrative registers in general
- Linking of register data for statistical purposes, different views:
  - Lack of privacy: NSO knows “everything” about “everybody” (Big Brother) or
  - Data security better taken care of in centralized system
- Citizens may be reluctant to report data to the NSO if they know that the same data have already been collected for administrative purposes
- Common Nordic experiences
  - Work of register authorities (and NSOs) should be open and transparent

# Unified identification system

- Unified identification numbers for basic units (persons, businesses, buildings, dwellings)
- Possible to link without unified ID-numbers
  - Using names, addresses, sex, date of birth etc.
  - Software is available
  - More resources needed
  - Quality of linking must be assessed

# Comprehensive and reliable register systems

- The existence of relevant administrative sources
- Quality in administrative data must be assessed
- Decision to produce register-based statistics: evaluation based on national circumstances

# Cooperation among authorities

Between different administrative authorities (registers owners)

Between NSO and the registers owners

- Cooperation at all levels of the organizations
  - From top management to “daily” exchange of information
- Norway: National forum for register owners and users
- Register owners should inform the NSO about all kind of changes
  - In Norway regulated in the Statistics Act
- The NSO should inform the register owners about quality problems
  - But normally not on an individual level

# Requirements to data from administrative sources

- Unique identifiers
  - Identifying codes should not be changed during the lifetime of a unit
  - Example: Norwegian personal ID (from CPR) contains date of birth and sex
    - ♦ Changes in ID number may occur
  - Good solution: Establishes links when ID-number is changed
  - Best solution: Using ID codes containing no information in the administrative register
- Time references and data on events
  - Important to have dates of changes and events
    - ♦ Real point in time when possible: Births, deaths
  - Dates of actual changes and dates of registration
  - Distinguish real events from corrections in registers

# Organizational aspects

- Statistics Norway and register owners have signed contracts on delivery of data from administrative registers
  - What registers covered
  - Units and variables covered
  - Frequency for updating/delivery
- Organization within NSO
  - Good practice: One unit in NSO responsible for contact with each register owner (coordinating unit)
  - In some countries: Separate data collection units

# Use of multiple sources

- Data from one single register may not be sufficient for statistical purposes
  - Differences in definitions
  - Limited coverage
  - Errors, for instance inadequate updating
- Using multiple sources in creation of new variables (derived variables)
  - Example: Employed persons
- Correction of information (editing, imputation)

# Quality issues

- Quality of administrative registers (input)
- Quality of linkage and transformation of data in NSO (statistical registers)
  - Also depending on human and technical resources for data processing
- Quality of register-based statistics (output)
- Quality frameworks developed
  - Austria, the Netherlands, Estonia
- More on quality in census presentation



# Documentation of administrative sources

The NSO needs a lot of information about the registers

- Laws, regulations and practices in register keeping
- Contents, metadata
- NSO should have (electronic) access to questionnaires/forms, instructions for registration etc.

# Documentation of changes over time

Changes important for register-based statistics

- Changes in definitions of administrative concepts
- New sources, new estimation methods
- Changes in classifications
- External changes that affect the statistics indirectly
- All changes should be stored in the NSO
  - Preferably in a calendar
  - Information organized by time of change, register, variable

# Coordination within NSO

- Organizational
  - Appoint units responsible for each statistical registers
    - ◆ Including contact with register owners
- Technical
  - Standardized naming conventions (same variable = same name)
  - Standardized data formats
- Methodological
  - Common methods for dealing with non-response, imputation etc

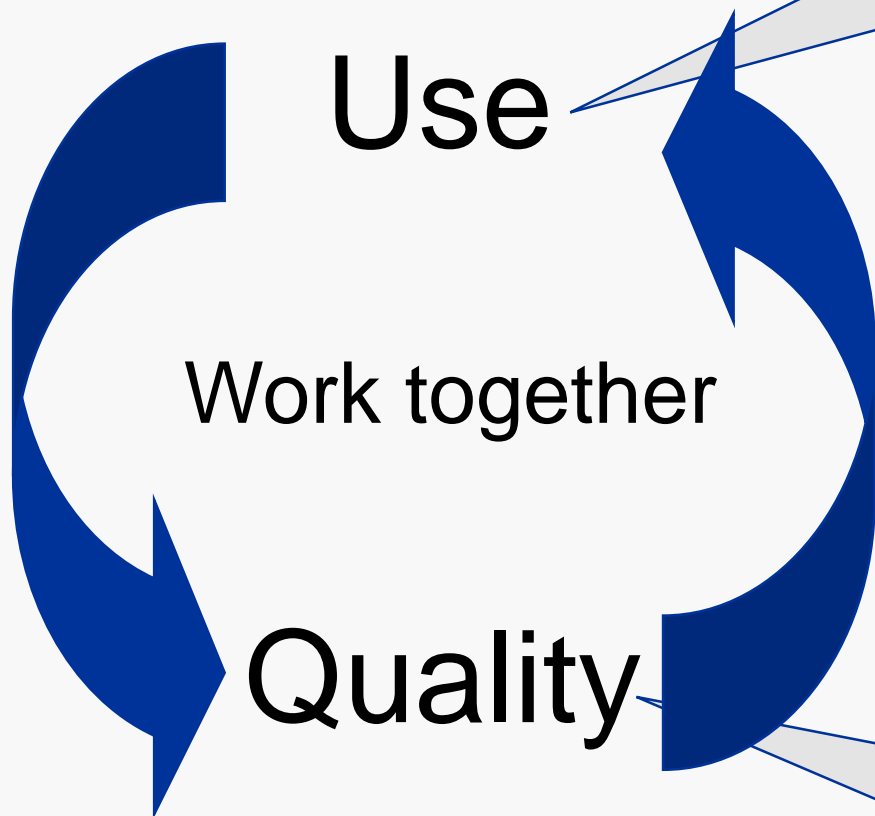
# Content related coordination

- Standardized populations
  - Same populations in all relevant statistics
- Standardized variables
  - For variables used in several statistical registers
- Standardized groupings of important classifications
- Classification and definition data bases recommended
  - Efficient storage of standards, classifications and definitions

## Concluding remarks

- Base registers and unique identifiers necessary to build an integrated system
- Register-based statistics cannot satisfy all user needs
  - Sample surveys as a supplement
- NSO should give active support to administrative registers
  - Establishment phase
  - Technical support – reporting quality problems
  - “Political” support – call for resources
- Registers as “common goods” inside the NSO
  - Data used for many purposes: Efficient metadata system
  - Responsibility (within the NSO) must be clearly defined
- Many users of register data: Better register quality

# ”The good circle”



Public sector  
Statistics  
Private sector

Updated information  
Quality secured  
Full coverage  
Documented (transparency)