

Administrative sources and their usage for statistical purposes

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION



Administrative sources and their usage for statistical purposes

- Administrative sources – overview
 - Scope
 - Type
 - Usage
- Statistical register – overview
 - Scope
 - Maintenance
 - Usage

Administrative sources

- *“In order to reduce the burden on respondents, the NSIs, other national authorities (...), and the Commission (Eurostat) shall have the right to access and use, promptly and free of charge, all administrative records and to integrate those administrative records with statistics, to the extent necessary for the development, production and dissemination of European statistics... .”*
(Article 17a Access, use and integration of administrative records, Regulation (EU) 2015/759 of the European Parliament and of the Council amending Regulation (EC) No 223/2009 on European statistics).
- **Administrative data sources** are data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service. They include administrative registers (with a unique identifier) and possibly other administrative data without a unique identifier.
- **Administrative data:** data coming from administrative sources

Administrative sources

Scope:

- administrative processes: population registration, taxation, social benefits, administrative permits, pensions;
- reduce the costs of data collection and the respondent's burden;
- statistical office should have access to administrative registers kept by public authorities;

Administrative sources

- administrative registers data to be transformed into statistical data;
- administrative data to be integrated;
- consistency regarding populations and variables to be assured (definitions, identification information).

Administrative sources

Type of administrative sources:

- registration data – population registers, demographical registration, register of addresses and residences, register on education; enterprise registers; regulatory authorities registers;
- taxation – financial statements, VAT, customs declaration etc
- administrative permits – building permits;
- social benefits - unemployment registration

Administrative sources

Usage

- Build statistical registers
- Maintain statistical registers
- Sampling frames
- Data imputation
- Substitute statistical data collection or combine with statistical surveys data

Administrative sources

Using administrative sources

- Reduces the burden on data suppliers
- Allows statistics to be compiled more frequently with no extra burden

Administrative sources usually offer better coverage of target populations, and can make statistics more accurate:

- No survey errors and non-response very low

Using administrative sources to produce statistics can sometimes be quicker than using surveys

Administrative sources

Direct Usage

1. Direct Tabulation
2. Substitution and Supplementation for Direct Collection

Indirect Usage

1. Creation and maintenance of survey frames
2. Construction of sampling designs
3. Editing and imputation
4. Indirect estimation and weighting
5. Data validation/confrontation

Administrative sources

DOMAINS	COMBINATIONS				NON- RESPONSE	TOTAL ⁸
	Admin/ register data only	Admin/ register & survey data	Survey data only	Not specified		
BUSINESS REGISTER	12	16	-	-	2	30
Turnover	2	15	12	-	1	30
New orders	-	10	19	1	1	31
STS Production prices/costs	-	14	16	-	1	31
Building permits	14	2	13	1	1	31
Employment	3	16	10	-	1	30
Annexes I-IV	1	23	4	-	2	30
Annex V	13	11	3	-	3	30
SBS Annex VI	16	8	2	1	3	30
Annex VII	13	7	4	3	3	30
Annex VIII	-	13	14	-	3	30
Annex IX	20	7	1	-	2	30
PRODCOM	-	10	13	1	2	26

Elaboration from Admin Data ESSnet WP1, Deliverable 1.2/2010. Database "Overview of Existing Practices in the Uses of Administrative Data for Producing Business Statistics in the EU and EFTA" (2011).

□

Administrative sources

- Relevance and completeness
- Timeliness and punctuality
- Comparability and coherence
- Accessibility and clarity
- Cost efficiency
- Low response burden

Statistical registers

Scope

- *Statistical Registers* are crucial for the whole process - they record the statistical units and a wide range of their variables are used for the identification of statistical units.
- Register is the *survey frame used to* identifies the statistical units of the population being observed, measured by a survey.
- Serves as a base for sampling, data collection and statistical processing

Statistical registers

Scope

- to record and maintain the *statistical units* and their characteristics (identifier, contact, classification, size category etc) for a scope as complete as possible.

According to the Business Registers – Recommendation Manual (Eurostat, 2010) the *register* is "a written and complete record containing regular entries of items and details on particular set of objects. Administrative registers come from administrative sources and become statistical registers after passing through statistical processing in order to make them fit for statistical purposes (production of register based statistics, frame creation etc.)."

Statistical registers

- **Statistical registers** are registers created for statistical purposes. They are typically created by transforming data from registers or administrative data sources. This transformation is often required to enable the registers or administrative sources to meet statistical definitions

Statistical registers

Statistical register plays the role of a data coordination tool by integrating data from several sources - statistical and administrative.

The main functions of the registers are:

- ✓ to collect and store information about the register population from one or more sources,
- ✓ to provide frame for collection and processing of statistical data, and
- ✓ to facilitate the analysis of the register population demography.

Statistical registers

The statistical units in BR are defined on the basis of three criteria:

- the legal, accounting or organisational criteria
- the geographical criteria and
- the production criteria.

List of statistical units of the production system:

- the enterprise;
- the enterprise group;
- the kind-of-activity unit (KAU);
- the local unit;
- the local kind-of-activity unit (local KAU).

Statistical business registers

- ***The enterprise** is the smallest combination of legal units that is an organisational unit producing goods or services, which benefits from a certain degree of autonomy in decision making, especially for the allocation of its current resources.*
- *An **enterprise group** is an association of enterprises bound together by legal and/or financial links*
- The ***kind of activity unit** (KAU)* groups all the parts of an enterprise contributing to the performance of an activity at class level (four digits) of NACE/ISIC and corresponds to one or more operational subdivisions of the enterprise.

Statistical registers

- The ***local unit*** is an enterprise or part thereof (e.g., a workshop, factory, warehouse, office, mine or depot) situated in a geographically identified place.
- The ***local kind-of-activity unit*** (local KAU) is the part of a KAU which corresponds to a local unit

Statistical registers

Identification characteristics

- Identity number
- Identity number(s) of the legal unit(s) of which the enterprise consist(s)
- Name and address
- Geographical location code
- VAT registration number or other administrative identity number

Statistical registers

Demographic characteristics

- *Date of commencement of activities*
- *Date of final cessation of activities*

Economic characteristics

- *Legal form*
- *Principal activity code*
- *Secondary activities*
- *Number of persons employed/Number of employees/Number of employees in full-time equivalents (FTEs)*
- *Turnover*

Statistical registers

Statistical register is built for statistical purposes (collect, store and maintain the units of a given population) – criteria to be met

- to solve the regular update of the units.
- to describe the attributes for identification and accessibility of population units.
- to describe the attributes for supporting the surveying process of the population.
- to manage the linkage of the register unit with the units of other registers connected to it.
- to contain and maintain the current and historical statuses of the population and the causes, effects and sources of alterations in the population.
- register data of population units have to be stored in a structured database.

Statistical registers

Maintenance - the register sources can be administrative registers, register surveys, statistical data transfers, feedbacks from surveys and other sources.

- The main sources of the statistical registers are the *administrative sources*.
- *Register survey* - The primary purpose of these surveys is to collect information to update register information. They can be used to control the quality of the register, to get information on the activity/inactivity of the units or the changes of their certain attributes.

Statistical registers

The third type of register sources is *the transfer of statistical data*. The register unit can have attributes updated from statistical sources.

For example, in the business register such attributes are the number of employees, the value of the turnover or the principal activity of the units. As only a small part of the business register units is observed by statistical surveys, survey data are not available to maintain these attributes for the whole register population.

For register quality, *feedbacks from the statistical surveys* can also be important sources.

During the data collection phase, accessing the respondents might reveal the error of the survey frame attributes, the changes of the address, name or other attributes of the respondents. The same feedback can be gathered from questionnaires returned.

Statistical registers

Commercial and other data sources

- private companies that manage utilities, for example, water or power supply, telephone companies,

Big data

Statistical registers

The maintenance of statistical registers

- is not an isolated operation;
- is part of a co-ordinated approach towards the joint development of statistical and administrative registers.

Statistical registers

Identification characteristics uniquely identify a given register unit;

Contact characteristics provide information to reach the unit for surveys;

Stratification characteristics support the classification of the units to range them into different strata for sampling, grossing and for the analyses;

Link characteristics facilitate the connection with other registers or sources.

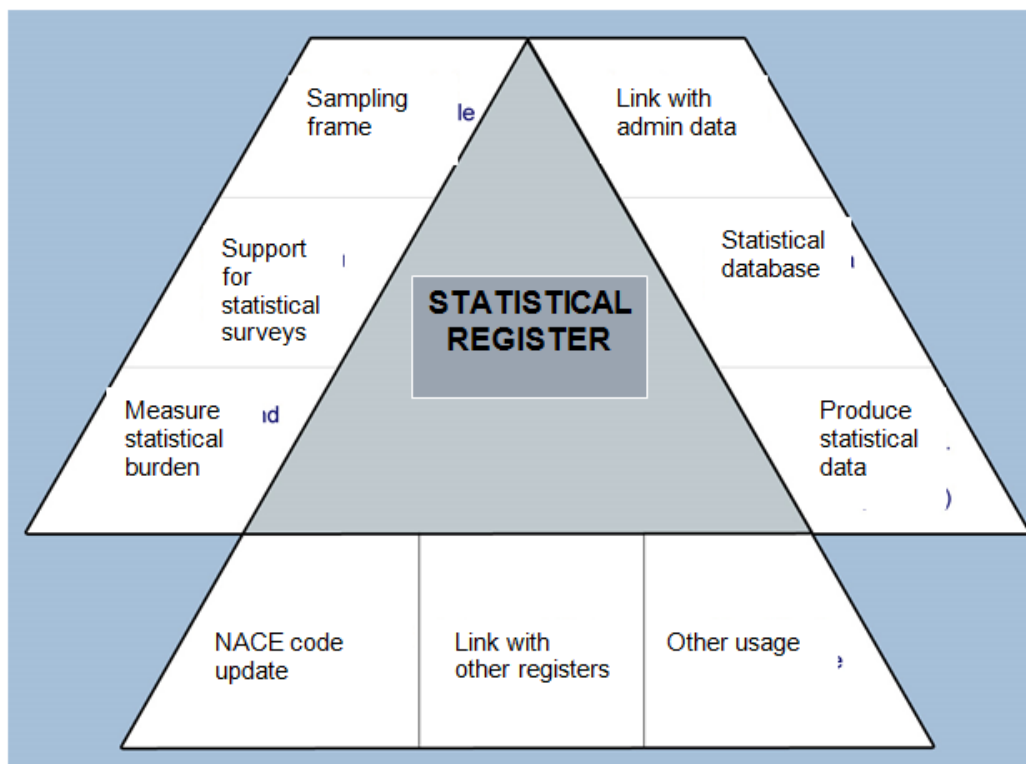
Statistical registers

- Frequency of updating
- Flow of maintenance and updating
- Role of staff in maintaining and updating
- Quality of statistical registers

Statistical registers

- new register should be in place at the beginning of each year;
- changes in units (during the year) and the dates of and reasons for those changes should be recorded;
- to store information about changes;
- at the end of the year, the register should be copied and stored;
- to keep copies of the registers (*"Member States shall make annually a copy that reflects the state of the registers at the end of the year and keep that copy for at least 30 years for the purpose of analysis"- BR Regulation*)

Statistical registers



Statistical registers

Usage

- To coordinate between various statistics (statistical units and elementary coordinated attributes);
- Population frame to ensure collection of statistical data coordinated in space and time;
- To control burden on respondents - compiling indicators on statistical burden;
- Linkage to administrative data by registering the link of administrative unit to a statistical unit;

Statistical registers

Usage

- micro data linkage - to increase the comparability of indicators on statistical units and to allow the production of additional information and analysis based on the already collected data;
- to create statistical information based on registers by aggregating administrative data to the level of the statistical unit – business demography.

Statistical registers

- mailing lists for statistical surveys;
- population of units for designing sampling schemes and for monitoring;
- the basis for grossing-up results from sample surveys to produce population estimates;
- instrument to prevent duplications and omissions in the collection process
- tool to insure the coherence between the results of different surveys;
- to measure and reduce response burden by co-ordination of samples .

Statistical registers

Roles of Statistical business register (UNECE, Guidelines on Statistical Business Registers)

	Role	Purpose
1	SBR live register	Provide the gateway between data from various input sources and statistical units
2	SBR register snapshot and frozen frame	Provide populations of statistical (and possibly administrative) units at fixed points in time
3	SBR survey frame	Provide a set of statistical units for a survey, valid for a specified reference period, with all characteristics required.
4	SBR survey support	Monitor survey response and measure and control response burden
5	SBR statistics	Produce statistics based directly on the SBR
6	SBR information source	Provide lists of enterprises and their locations and possibly other characteristics
7	SBR international data exchange	Facilitate coherence in international statistics
8	SBR in modernisation of statistical production and services	Promote integration of SBR within the production processes for economic statistics

Thank you for your attention !

Frameworks for the access and use of administrative data

Kaija Ruotsalainen

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- Preconditions for successful use of administrative data
- ESS – European Statistical System
- EU's statistical regulation
- Case: Legal framework in Finland
 - Data collection
 - Data processing
 - Confidentiality and data privacy
- Organisational frameworks

Preconditions for successful use of administrative data

- 1 Comprehensive and reliable nationwide administrative registers
- 2 Strong legal basis
- 3 Use of uniform identification numbers in the different registers
- 4 Acceptance of the system by the population
- 5 Good collaboration between statistical and administrative authorities

ESS - the European Statistical System

- A partnership of Eurostat, National Statistical Institutes and other national authorities responsible in each Member State for the development, production and dissemination of European statistics (EU, EEA, EFTA countries)
- ESS defines harmonised methodologies in collaboration with the MSs
- Data collected/compiled by the NSIs in the MSs
- Eurostat receives, treats and disseminates data on the EU, euro area and MS
- Available for all who are interested (journalists, decision makers, researches, citizens etc.)

The EU's statistical regulations

- [Revised regulation 223/2009 on European statistics \(759/2015\)](#)
 - Professional independence
 - The access to the administrative data within the own national administration was improved
- Reform of EU data protection rules
 - [Regulation \(EU\) 2016/679](#) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data

The EU's statistical regulations

- European Statistics Code of Practice
 - 15 principles concerning the institutional environment, statistical processes and outputs
 - It aims to ensure that statistics produced within the ESS are not only relevant, timely and accurate but also comply with principles of professional independence, impartiality and objectivity.
 - A set of indicators of good practice for each of the 15 principles provides a reference for measuring the implementation of the Code.
 - The Code was revised in 2011 and adopted by the ESSC on 28th September 2011.

The role of legislation

- The use of administrative data sets calls for the creation of the necessary legislation:
 - gives to the statistical authorities sufficient powers to extract data from administrative sources
 - enables information collected for other purposes to be handed over for the compilation of statistics and for research
- The legislation currently in force in Finland supports the statistical use of this material extremely well

Legal framework in Finland

- Statistics within ESS
 - EU legislation
 - National legislation
- Other national statistics
 - Statistics Act (1994, rev. 2004)
 - Personal Data Act (523/1999)
 - Act on the Openness of Government Activities (621/1999)

General guidelines of data collection (1)

- The Statistics Act defines general guidelines for data collection
 - to use primarily data generated in other contexts
 - data collection must cover only those data that are necessary for production of statistics
 - data must be collected in a manner that is economical and causes minimum inconvenience and costs to the respondents

General guidelines of data collection (2)

- The Statistics Act defines general guidelines for data collection
 - to consult with data suppliers
 - to inform data suppliers of matters that bear on the provision of the data
 - to give feedback to data suppliers

Obligation to provide data (1)

- The Statistics Act in Finland defines the obligation to provide data
 - central government authorities
 - enterprises, local government authorities,
 - persons performing public service duties
 - organisations providing education
- Data should be collected and stored without identification data whenever it is possible
- In practice: identification data is needed for efficient use of existing administrative data -> data linking

Obligation to provide data (2)

- Defines services free of charge
 - raw data obtained free of charge, except for the cost of data release involved in indirect data collection
 - feedback given to data suppliers is free of charge

Data collection from administrative registers

- About 95% of the data today come from administrative sources
- The flow of information runs in one direction only i.e. from administrative authorities to the statistical institute
- People must be able to trust that data given to one administrative authority do not go to another admin authority
- Register-based statistics offer various advantages in terms of data protection

Data processing (1)

- The Statistics Act defines data processing and compilation of statistics
 - data are duly protected during all stages of statistics production
 - no person's or business or professional secret is endangered
 - data processing takes place in accordance with statistical ethics, good statistical practice and international recommendations
 - the basic principle is that confidential data are only processed by those whose tasks demand it

Data processing (2)

- Personal Data Act defines the processing and use of personal data
 - Personal data must not be used or otherwise processed in a manner which is incompatible with the purposes for which the data have been collected.
 - Later processing for purposes of historical, scientific or statistical research is not deemed incompatible with the original purposes.
 - A personal identity number may be processed for purposes of historical, scientific or statistical research if it is necessary to identify the data subject.

Confidentiality of statistical data (1)

- The Statistics Act defines data secrecy and release
 - basic data confidential
 - may be released for statistical and research purposes, as a rule without identification data
 - may not be released under any other piece of legislation e.g. for administrative purposes
 - The data collected from administrative sources are confidential in the possession of statistical authorities even if these data are public in the possession of administrative authorities

Confidentiality of statistical data (2)

- exceptions for the confidentiality:
 - information describing the activities of authorities and the production of public services
 - some data in the Business Register
- Requirements of the Personal Data Act
 - systematic data collection
 - register descriptions on view on Statistics Finland's Internet pages and at the Library of Statistics
 - safeguarding of data privacy
- Right to check one's own data does not extend to statistical registers

Information to data subjects

- obligation to inform the respondents of
 - the purpose for which the data will be used
 - the procedures by which the statistics will be compiled
 - any particular principles governing dissemination of the resulting information
 - any case where interview data will be supplemented with administrative data
 - and other factors of relevance

Description of file

- Personal Data Act defines the description of personal data file - **transparency**
 - The name and address of the controller and, where necessary, the representative of the controller
 - The purpose of the processing of the personal data
 - A description of the group or groups of data subjects and the data or data groups relating to them
 - The regular destinations of disclosed data and whether data are transferred to countries outside the EU and the EEA
 - A description of the principles in accordance to which the data file has been secured

Benefits and challenges from the point of view of data protection (1)

Benefits

- Data subjects can live in peace. They are not harassed with unnecessary inquiries
- Reduction in staff handling the data
- Reduction in the number of external parties
- Data are seen only by the computer

Benefits and challenges from the point of view of data protection (2)

Challenges

- Administrative data offer the opportunity to extend the scope of information collected and in this sense increase its sensitivity
 - **importance of professional ethics**
 - **effective data security arrangements are needed**
- Extensive use of registers may arise a fear of infringement of personal privacy
 - **distribution of information**
 - **the good image of statistical authorities**

Organisational framework (1)

- Co-operation
 - Possibilities for using registers can be improved through effective co-operation with authorities
 - impact on the data content of registers
 - creating a better understanding of the use of administrative data for statistical purposes
- Meetings at DG level with ministries and other authorities
- Co-operation officers at NSIs
- Register pool/network

Organisational framework (2)

■ Communication

- Information about the use of administrative data sources
- Information through the media
- Information through the legislation governing administrative data sources
- Administrative authorities inform data subjects about the use for statistical purposes
- Letters and brochures are sent to the data subjects and interviewers inform them when interviewing

Organisational framework (3)

- the access to the administrative data is one of the key barriers to the effective use of administrative data for statistical purposes
- when the legal and policy problems are solved it is important to organise the data flows to Statistical office in a rational way
- some sort of agreement
- written agreement – has more power
- informal agreement – may include a risk, changes of staff etc.

Organisational framework (4)

- Agreements with the data owners
 - legal basis
 - contact persons
 - detailed description of data
 - frequency of data supply
 - quality standards
 - confidentiality rules
 - technical standards
 - provisions for payment for supply data
 - period of agreement
 - procedure for resolving disputes



**Thank you for your
attention!**

Administrative sources in statistical registers

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- Statistical registers
- Type of administrative sources
- Models for creating and maintaining Statistical Registers using administrative data

Statistical registers

“A register is a written and complete record containing regular entries of items and details on particular set of objects.” (UNECE / Conference of European Statisticians Statistical Standards and Studies, 2000)

***“Register”** is defined as a systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up to date, correcting or extending the register, i.e. keeping track of any changes in the data describing the units and their attributes”*

- register is structured list of units, which contains a number of characteristics for each of those units with a regular maintenance and updating mechanism;
- statistical register is a register that is constructed and maintained for statistical purposes based on statistical concepts and definitions;
- statistical register is built, maintained and used by statisticians.

Statistical registers

- Statistical register uses several sources of information and integrate them within the statistical register to assure more accuracy of the data.
- Administrative source are utilized in all phases:
 - **Creating statistical registers**
 - **Maintaining and updating statistical registers**
 - **Improving quality of statistical registers**
 - **Expending statistical registers**

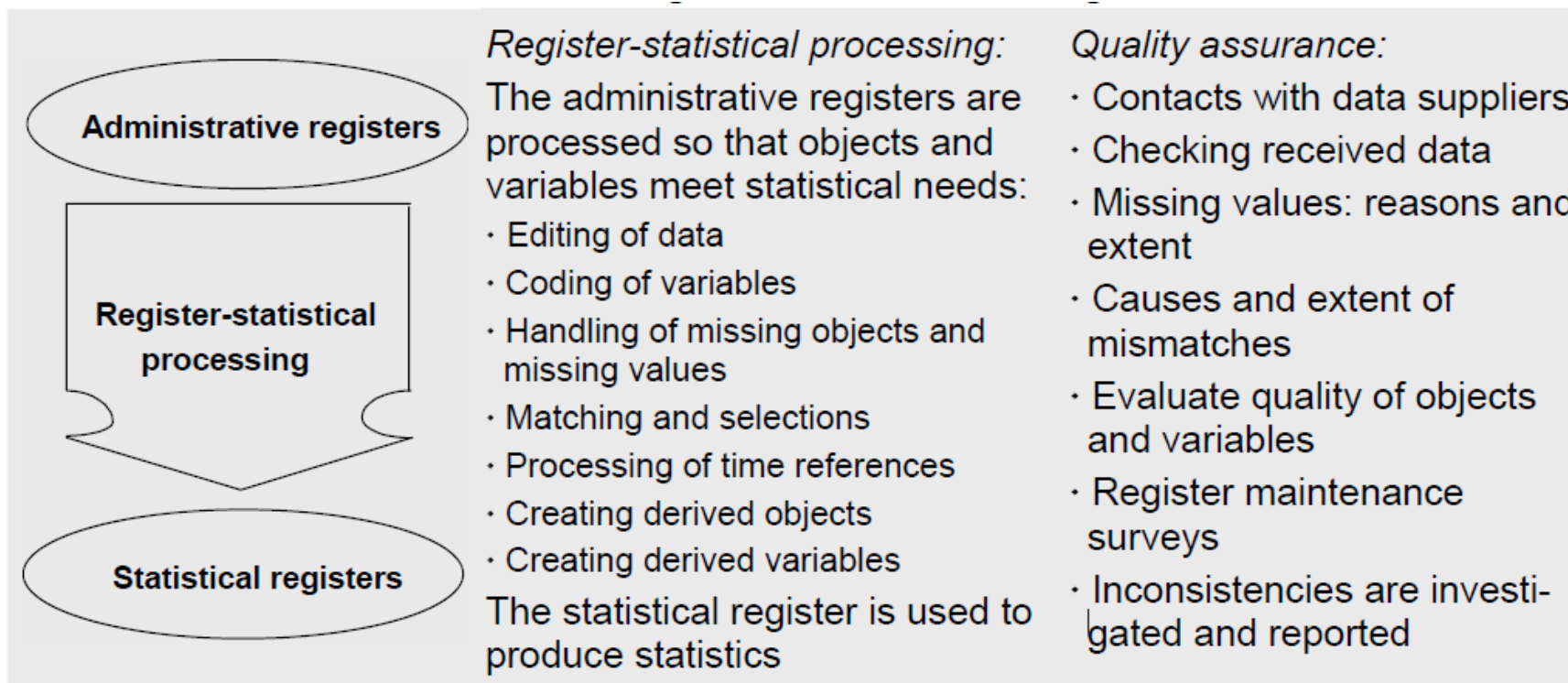
Type administrative sources

- Business registration/license register;
- Tax registers
- Company/trade associations and chambers of commerce registers
- Social security registers
- Labour and employment registers
- Government units registers
- Non-profit unit registers
- Industry association registers
- Agricultural administrative registers
- Central banks

Models for creating and maintaining Statistical Registers using Administrative Data

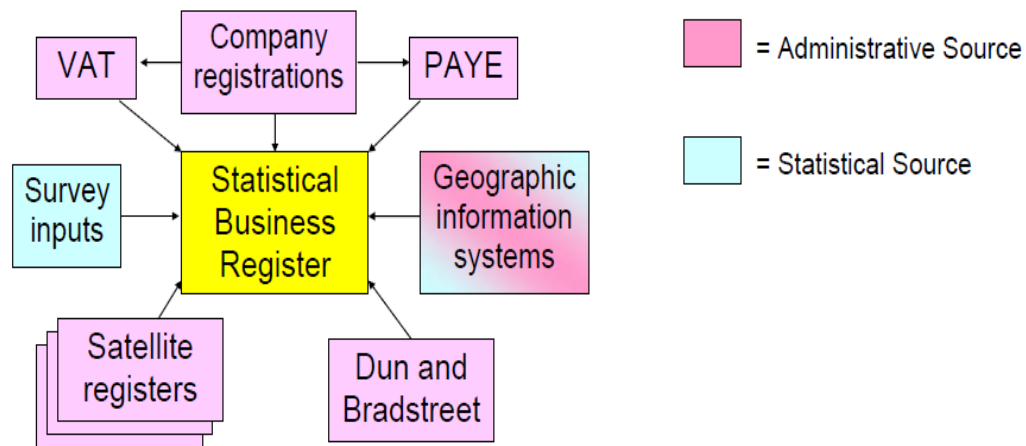
1. Combining Multiple Sources
2. Using Centralised Administrative Registers
3. Creating a Data-sharing Hub
4. Using Administrative Data via Satellite Registers
5. Register-based Statistical Systems

Models for creating and maintaining Statistical Registers using Administrative Data



Combining Multiple Sources

Model of Statistical Business Register Sources in the UK

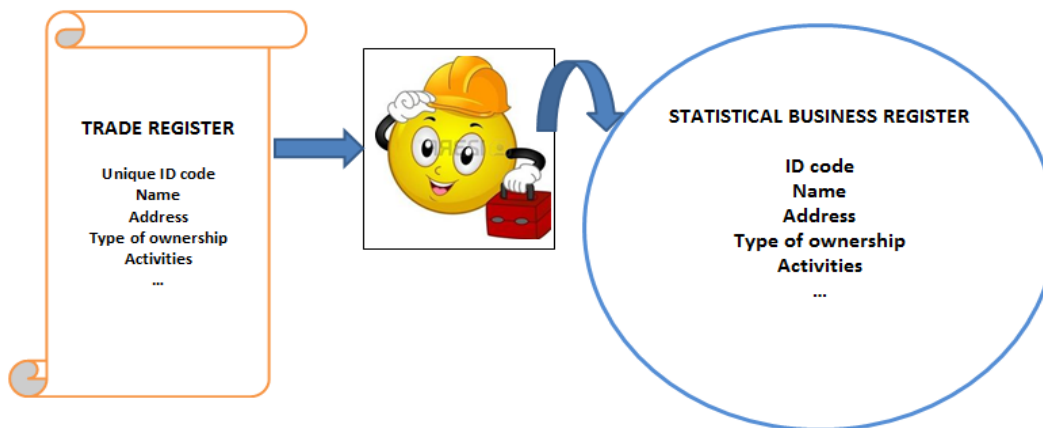


UNECE, *Using Administrative and Secondary Sources for Official Statistics, A Handbook of Principles and Practices*

Combining Multiple Sources

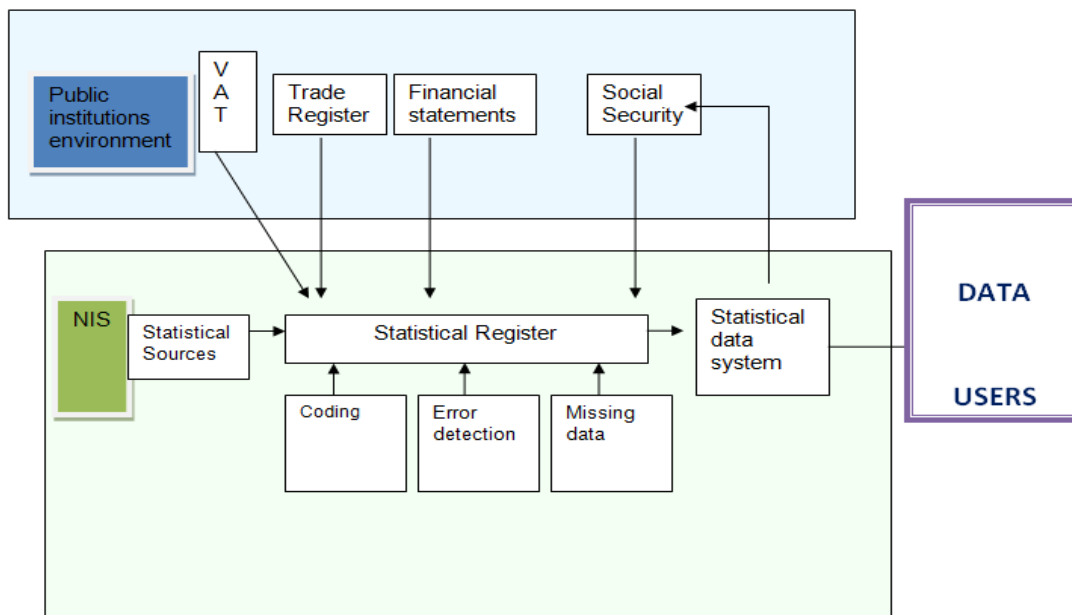
Model of Statistical Business Register Sources in Romania - creation

Romanian National Trade Register. The Trade Register records all acts, activities and the identity of the dealers involved for which registration is required by law, and any other acts or deeds expressly required by law.



Combining Multiple Sources

Model of Statistical Business Register Sources in Romania



Combining Multiple Sources

Steps in transforming administrative data into statistical register data

- Check the quality and completeness of the admin data;
- Preparing the structure of data;
- Identify the link between the administrative unit and statistical units
 - Identification number => direct link
 - Not unique identification number => link using name, address, activity code
- Improve the quality of statistical register

Using Centralised Administrative Registers

- a single interface through which the subjects of the register can interact with different government agencies;
- data from different sources matches;
- ensure that statistical needs are met: units, classifications, definitions

Using Centralised Administrative Registers

- Australian Business Register (ABR) was developed by the Australian Tax Office to administer various businesses taxes; it is maintained in close cooperation with the Australian Bureau of Statistics;
 - used by business and government of a complete and single source of business information for planning and service delivery;
 - use of consistent information exchange standards by business and government;
 - more-streamlined online interactions between business and government including a single business entry point.

Using Centralised Administrative Registers

“French model”, highly centralised, the NSI is to directly manage a “hybrid” business register (SIRENE), serving for both administrative and statistical purposes,

- wide range of sources used to update it – which places under its control much of the generation process of administrative data itself

Using Centralised Administrative Registers

Benefits

- reducing the administrative cost to business of complying with Government rules;
- minimise the administrative impact on business;
- adoption of new practices by businesses to reduce operating costs;



European
Commission

Member	ABR Advisory Board	SBR Board	SBR Steering Committee
ACT Revenue Office (representing state and territory revenue offices)	✓		
Australian Board of Taxation		✓	
Australian Bureau of Statistics	✓	✓	✓
Australian Business Register	✓ (Chair)	✓	✓ (Chair)
Australian Chamber of Commerce and Industry		✓	
Australian Immigration and Border Protection Service	✓		
Australian Local Government Association	✓		
Australian Prudential Regulation Authority		✓	✓
Australian Securities and Investments Commission	✓	✓	✓
Australian Taxation Office	✓	✓	✓
Business Council of Australia		✓	
Council of Small Business of Australia (also representing the SBR Business Advisory Forum)		✓	
Department of Communications	✓	✓	✓
Department of Finance	✓	✓	✓
Department of Health		✓	✓
Department of Human Services	✓	✓	✓
Department of Industry and Science	✓	✓	✓
Department Prime Minister and Cabinet		✓	
Department of the Treasury	✓	✓ (Chair)	✓
Office of NSW Small Business Commissioner	✓		
Office of State Revenue New South Wales (representing the eight state and territory revenue offices)			✓
Office of State Revenue Queensland (representing the eight state and territory revenue offices)		✓	
Small Business Victoria (representing state government agencies)	✓		

Data-sharing Hub

A sort of a single centralised administrative

- tool for finding and matching data held by different agencies - contain basic identification data;
- gateway through which data from different organisations can be shared within the government sector.

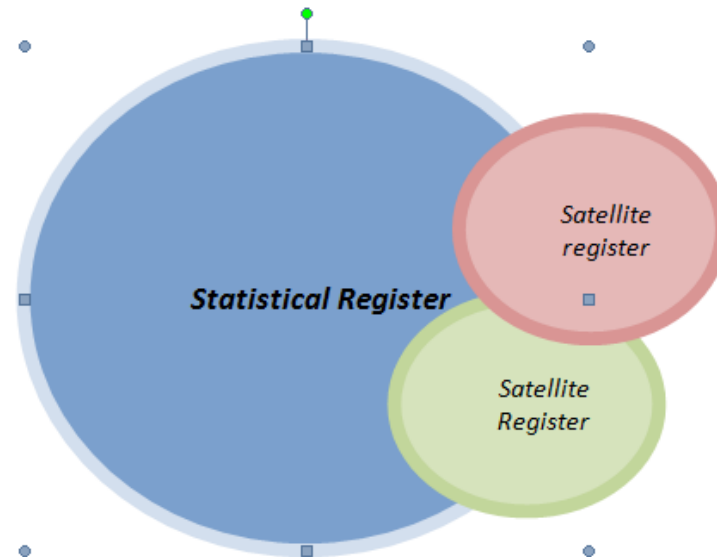
Administrative Data via Satellite Registers

Satellite registers are available to the national statistical system, contain information about units and variables of interest

- not an integral part of a statistical register, but can be linked to it;
- more limited in scope than that statistical register (more extensive coverage of units and/or variables for their scope);
- contain variables that are not found in the statistical register.

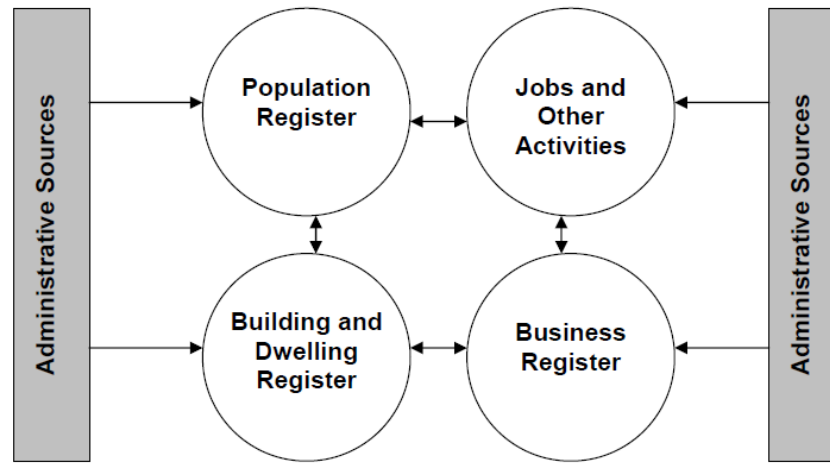
Tools for incorporating administrative data that are only relevant for a sub-set of units in a statistical register.

Administrative Data via Satellite Registers



Register-based Statistical Systems

- ***Nordic Register-based Statistical Systems***



Thank you for your attention !

- Exercise

Techniques for administrative data sources integration in statistical registers

Kaija Ruotsalainen

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- What is matching/data linkage?
- Matching keys
- Matching process
- Examples of matching methods
- Where and when to use?
- Integrated statistical data
- Examples from Finland
 - Business register
 - Short Term Business Statistics
 - Population and Housing Census

What is matching / data linkage?

- Linking data from different sources
- Exact matching - linking records from two or more sources, often using common identifiers
- Probabilistic matching - determining the probability that records from different sources should match, using a combination of variables

Matching keys

Data fields used for matching e.g.

- Id-numbers (persons, dwellings, businesses)
- Name (first names, family name, birth name)
- Birth/death date
- Age
- Sex
- Address
- Postcode
- Classification (e.g. ISIC, ISCO)
- Other variables (age, occupation, etc.)

Matching via

Data type	Primary identifiers (exact)	Secondary identifiers
Personal data	PIN	Name, Date of birth, Place of birth, sex
Business data	BIN	Enterprise name, Post code, Street name, Street number
Buildings (Location)	co-ordinate, address-id, building-id	Post code, Street name, Street number

Matching process (1)

- With the unique ids usually no problem
 - Possible changes if ids: the history of ids
- With textual data
 - Standardisation
 - Capital letters
 - Abbreviations and common terms are replaced with standard text
 - Common variations of names are standardised, synonyms
 - Postal codes, dates of birth etc. are given a common format
 - Elimination blanks

Matching process (2)

- Parsing
 - Names and words are broken down into matching keys
 - Improves success rates by allowing matching where variables are not identical
- Blocking
 - If the file to be matched against is very large, it may be necessary to break it down into smaller blocks to save processing time
- Scoring
 - Matched pairs are given a score based on how closely the matching variables agree
 - Scores determine matches, possible matches and non-matches

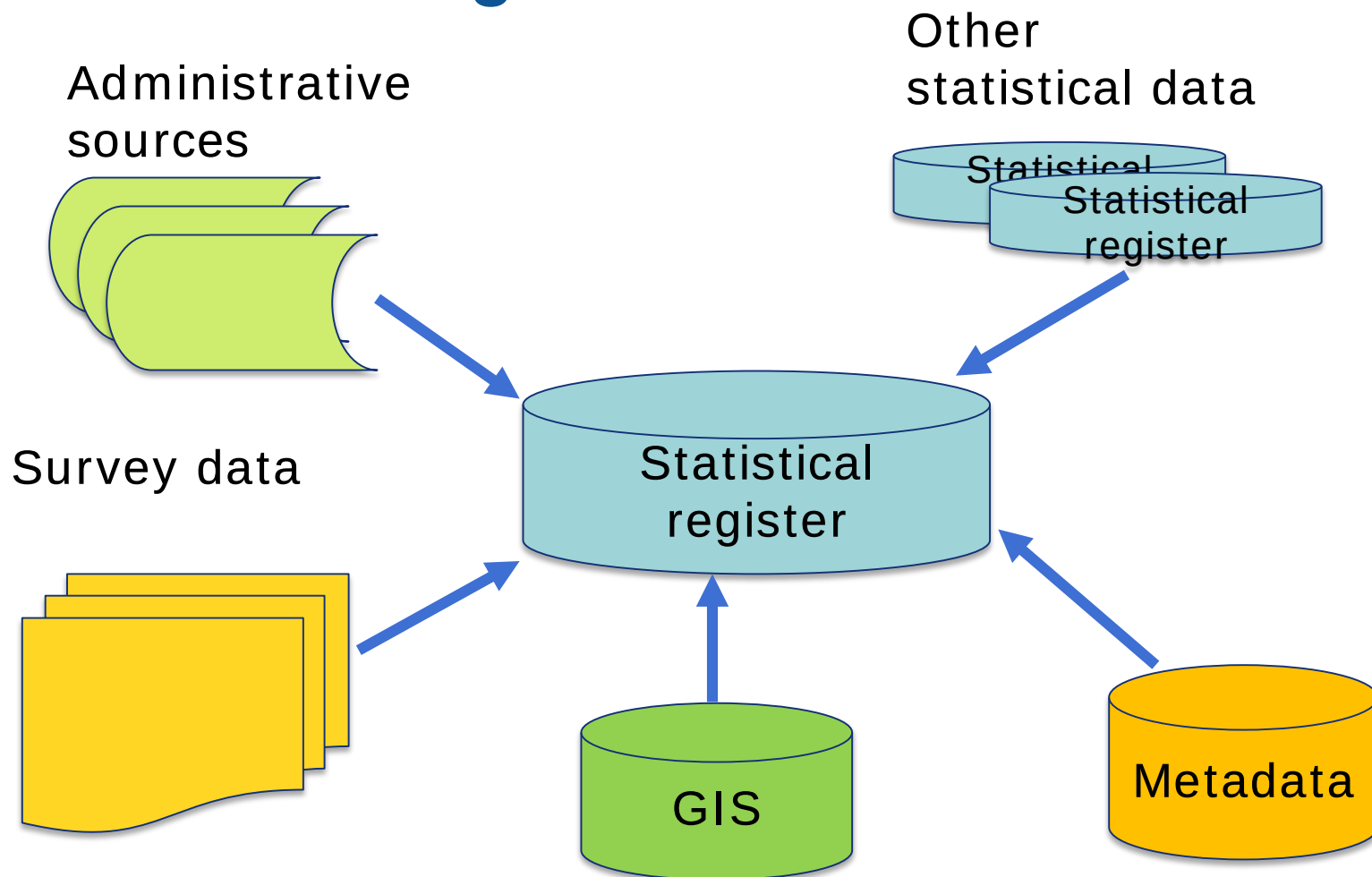
Matching methods: Some examples

- Richard Cadieux & Daniel R. Bretheim (2014): [Matching Rules: Too Loose, Too Tight, or Just Right?](#)
 - SPEDIS function – Determines the likelihood of two words matching, expressed as the asymmetric spelling distance between the two words.
- String Similarity
 - The Levenshtein Distance Algorithm
 - Jaro-Winkler Distance Algorithm
 - Adjacent Pairing Algorithm

Where the data matching is used?

- Statistical registers
- Statistics from multiple source models
- Register-based statistical systems
- Using administrative data for estimation
- Reduce response burden
- Build efficient sampling frames
- Pre-filled questionnaires
- Using administrative data for non-responders
- Impute missing data

Statistical registers



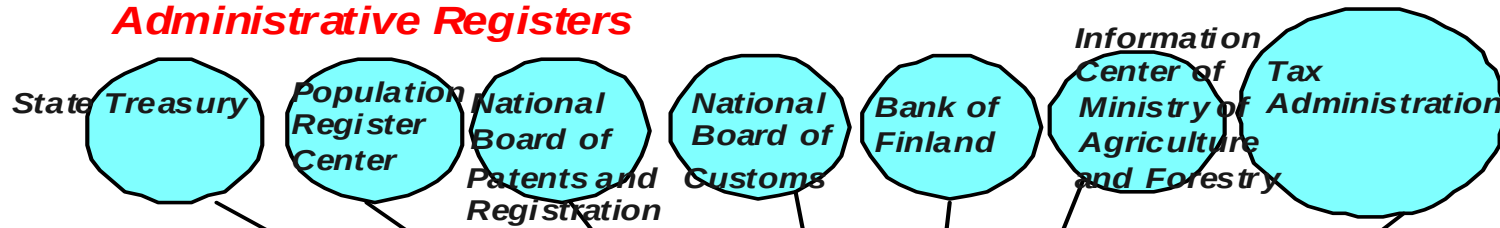
Multiple Data Source model

- Traditionally one statistical output was based on one statistical survey
- Very little integration or coherence
- Now there is a move towards more integrated statistical systems
- Outputs are based on several sources

Example: Statistical register, Finnish BR

Data sources of Business Register

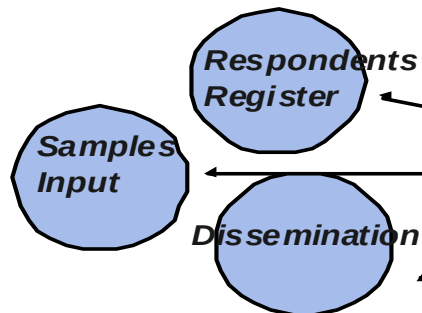
Administrative Registers



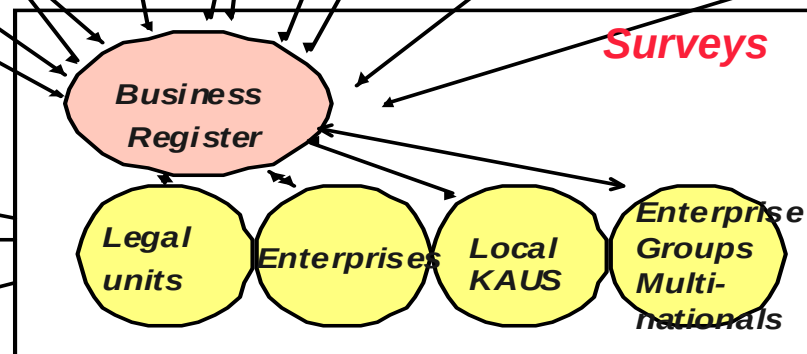
Commercial Registers



Statistics Finland



Surveys

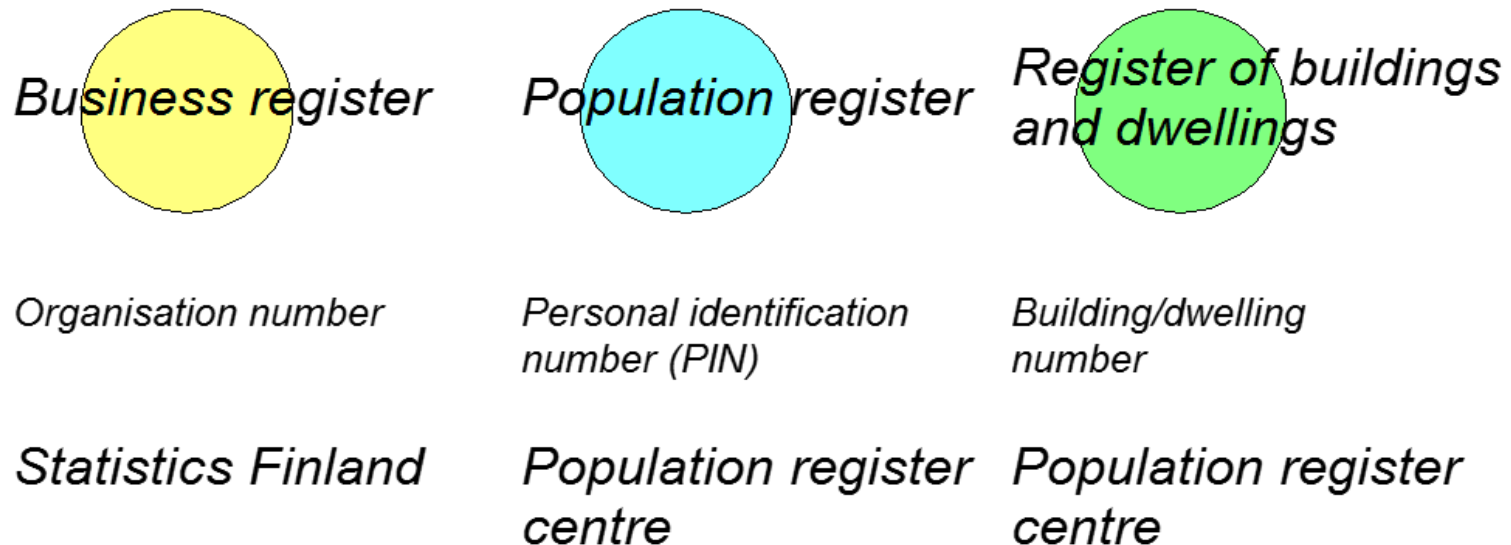


Linking Administrative and Survey Data example: STS - Short Term Business Statistics

- The monthly VAT (Value added tax) and PAYE (pay as you earn) data file of the Tax Administration
 - total coverage of units (200 000 statistical units, 150 observations, 40 variables)
 - each release covers 6 months
 - delay: one-two months
- The BR: units, classification variables, samples input
- Direct surveys: 2000 largest enterprises from industry, construction, trade and other services
- Regional STS -> to use local kind of activity unit data from the BR to disaggregate monthly enterprise level VAT data to a regional, establishment level

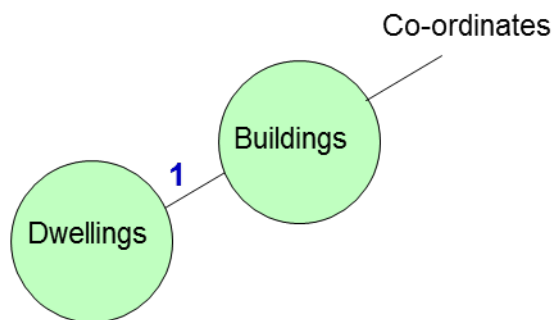
The basic registers

Population Information System



The basic units of the register-based census in Finland - and links between them

*Buildings and
dwellings
in PIS (CPR)*

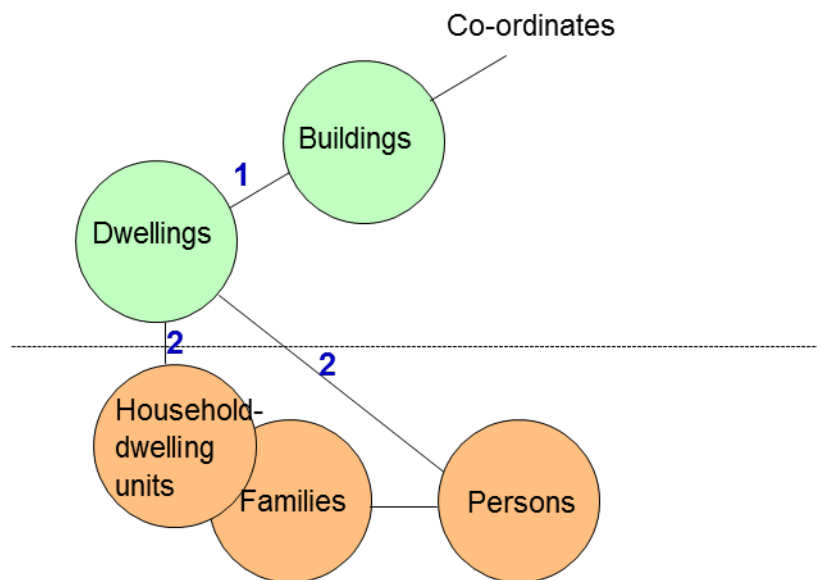


1 Building code

The basic units of the register-based census in Finland - and links between them

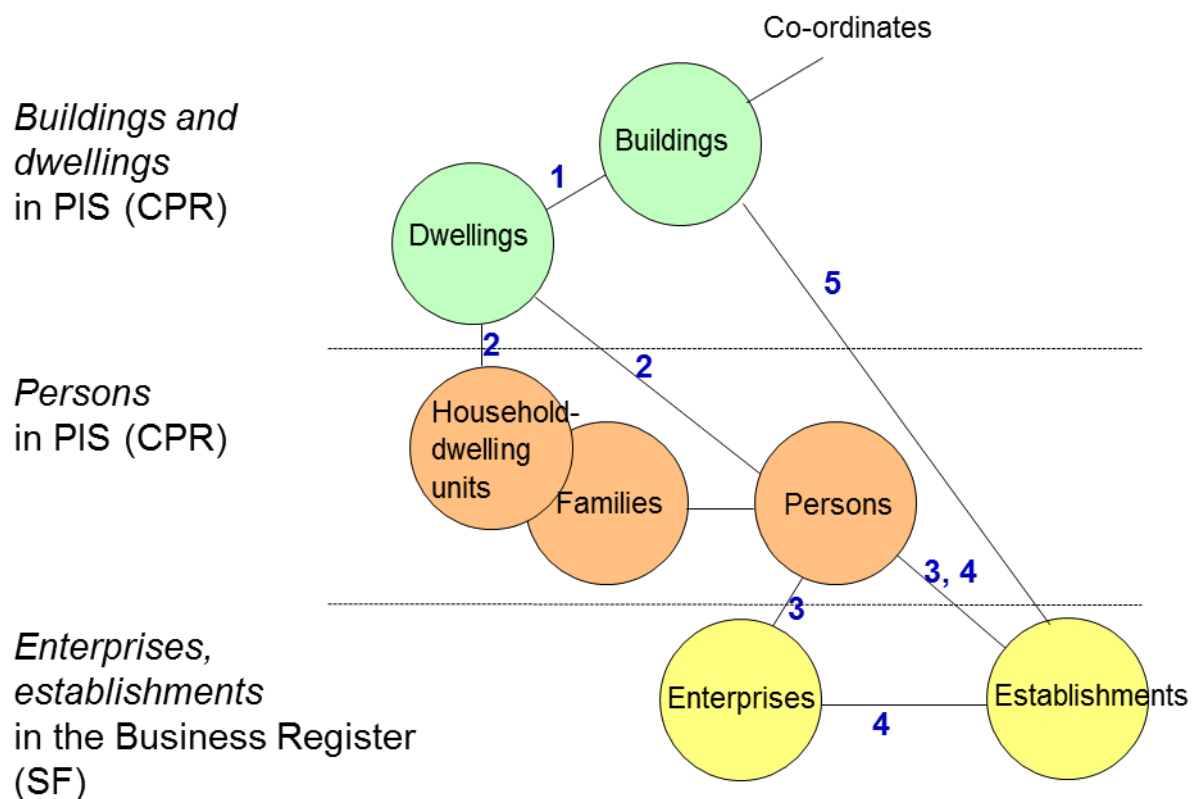
Buildings and dwellings in PIS (CPR)

Persons in PIS (CPR)



1 Building code
2 Dwelling code

The basic units of the register-based census in Finland - and links between them



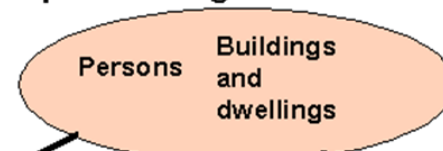
- 1 Building code
- 2 Dwelling code
- 3 Business-id
- 4 Establishment number
- 5 Address

Integrated Data sources of the Finnish Census

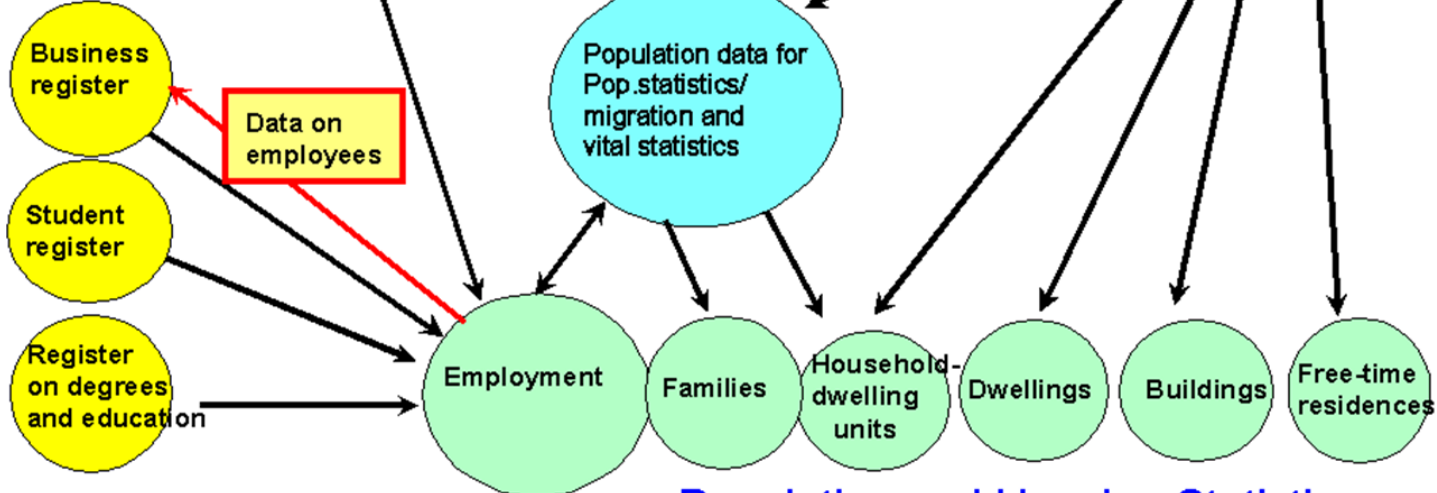
Administrative registers



Population Information System of the Population Register Centre



Statistics Finland



Population and Housing Statistics



**Thank you for your
attention!**

Combine statistical surveys and statistical register data

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

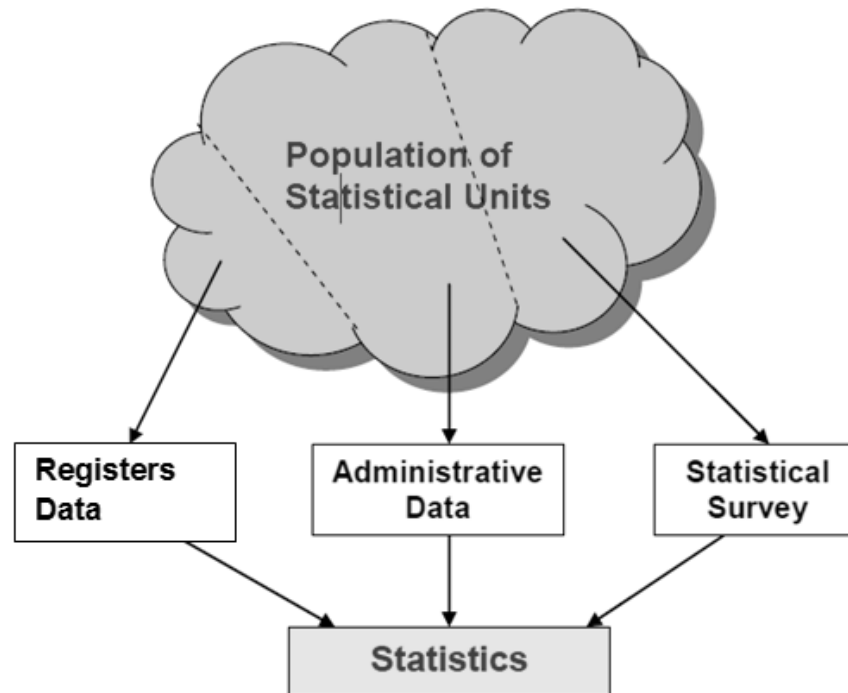
Content

- Overview
- Structural business statistics
- Trade by enterprise characteristics
- Foreign affiliates statistics

Overview

- mixed-source approach to be used to produce statistics at lower cost and/or better quality;
- accessible to statisticians;
- consistent, reliable, timely datasets

Overview



Overview

Relevance

Statistical registers are the basis for sampling frame

Accuracy

No sampling error

Timeliness and punctuality

Production time shorter than for statistical surveys; data are already available

Comparability and coherence

Variables already validated, the same statistical unit

Structural Business Statistics

Structural Business Statistics (SBS) describe the structure, activity, competitiveness and performance of economic activities within the business economy.

SBS cover industry, construction, trades and services.

The main indicators within SBS are generally collected and presented as monetary values or as counts.

SBS contain a comprehensive set of basic variables describing the economic performances of businesses, employment characteristics and business demography.

Structural business statistics

- Number of enterprises
- Number of local units
- Number of persons employed
- Number of employees
- *Turnover*
- Value added at factor cost
- Total purchases of goods and services
- Gross investment in tangible goods
- Personnel costs

Structural business statistics

Number of enterprises

“A count of the number of market enterprises as defined in Council Regulation (EEC) No 696/93 registered to the population concerned in the business register.”

Number of local units

A count of the number of local units as defined in Regulation (EEC) No 696/93 registered to the population concerned in the business register.”

Number of persons employed is defined as the total number of persons who work in the observation unit, inclusive of working proprietors, partners working regularly in the unit and unpaid family workers working regularly in the unit (SBS).

The number of employees is defined as those persons who work for an employer and who have a contract of employment and receive compensation in the form of wages, salaries, fees, gratuities, piecework pay or remuneration in kind

Structural Business Statistics

The register should record the actual numbers of persons employed and employees, both as head counts and the latter also in FTEs.

Definition: The structural business statistics definitions variables should be used.

Trade by economic characteristics

- globalised world with economies interconnected; it is important to know traders and their characteristics;
- statistical data on international trade Intra and Extra-EU (Intrastat and Extrastat) are linked with business register information on enterprises at the individual enterprise level;
- analysis of the impact of trade on employment and production.

Trade by enterprise characteristics

Statistical data survey

- Monthly survey data collected for entities that export and import

Variables collected:

- Value and quantity of goods delivered/exported
- Value and quantity of goods introduced/imported

By countries, type of goods, transport mode

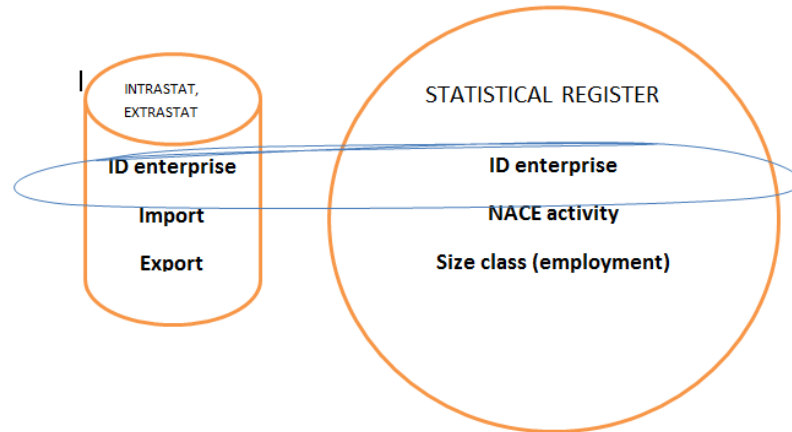
Statistical Business Register

- Economic activity: NACE code
- Size class: small, medium, large enterprise (according to employment)
- Region

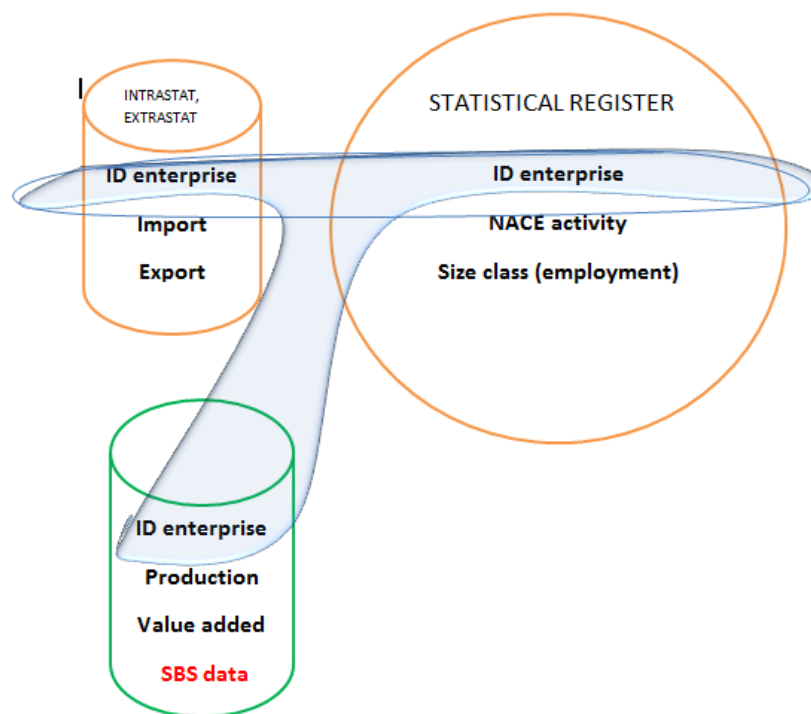
Trade by enterprise characteristics

Survey data

Statistical register
data



Trade by enterprise characteristics



Foreign affiliates (FATS)

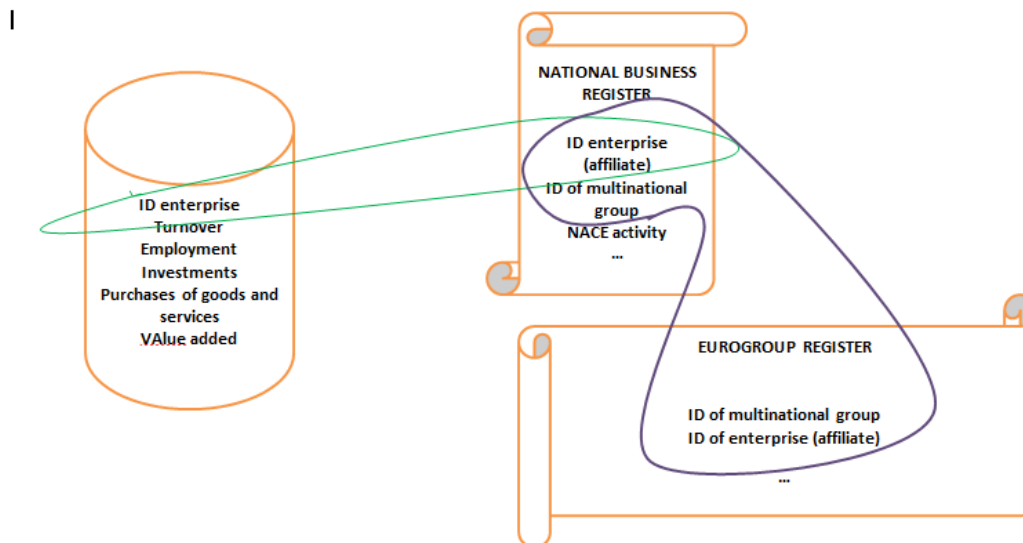
- In Romania FATS data are produced by combining data collected in the survey and statistical registers
- Information requested by FATS regulation: turnover, employment, investments, purchases of goods and services, value added

for enterprises that are affiliates of multinational enterprises

Foreign affiliates

- Sources of information:
 - **Annual business survey**
 - *Annual financial statements*
 - **National Business Register**
 - **Euro-group Register**

Foreign affiliates



Thank you for your attention !

- Group exercise



Evaluation of the data quality

Kaija Ruotsalainen

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- Quality Assurance Framework in ESS
- Criteria for evaluating quality
- Case studies in Finland: data quality
 - data on persons
 - data on buildings and dwellings
 - data on main type of activity

Quality Assurance Framework in ESS

- Quality Assurance Framework of the European Statistical System
 - serves as guidance on how to implement the European Statistics Code of Practice

Criteria for evaluating quality

- **Relevance**
 - the degree to which statistics meet the needs of current and potential users
- **Accuracy**
 - the closeness of statistical estimates to true values
- **Timeliness**
 - this reflects the length of time between data being made available and the event or phenomenon they describe
- **Punctuality**
 - the time lag between the date that data were actually released and the target release date

Criteria for evaluating quality

- Accessibility
 - the physical conditions in which users can obtain data
- Clarity / interpretability
 - metadata, information on data quality
- Coherence / consistency
 - data from different sources
- Comparability
 - comparability over time
 - comparability through space
 - comparability between domains

Quality measurements in practice

- The quality of incoming data
 - can be judged against the criteria listed above
 - e.g. data sets coming to Stat Finland are checked:
 - to be in readable format
 - to have correct keys (identifications)
 - to have asked variables
 - to have right values
 - to be compared with external source
 - to be compared with previous year/month data
 - if something unclear or there are big changes => contact to the data owner

Quality measurements in practice

- The quality of data processing
 - quality can be affected by different processes:
 - data matching and linking
 - data editing and imputation
- ⇒ to keep a copy of the raw data to refer back if necessary

Quality measurements in practice

- The quality of statistical outputs
 - moving from survey to administrative sources will have an impact on output quality
 - positive for some quality criteria and negative for others

Case study: Finland

Data quality

- data on persons
- data on buildings and dwellings
- data on main type of activity

Reliability (1)

- Studies to research and monitor the reliability of register-based data were carried out well ahead of the decision to adopt a register-based census system in Finland.
- A major reliability survey was carried out in conjunction with the first entirely register-based population census in 1990.
- These register sources were compared with the results of a sample questionnaire survey which comprised around two per cent of all buildings, dwellings and persons in the country.
- The results indicated the proportion of responses where the questionnaire data deviated from the register data, but not which of these two sources provided the correct information¹⁰

Reliability (2)

- Respondents may give a different figure
 - for the floor area of their dwelling than indicated in the building permit
 - a person who has more than one job may well opt for a different choice than the register keeper
 - a student who has a job will always be defined as gainfully employed on the basis of register data, yet that student might well not report having a job at all.
- It has been shown that the difference between register-based and questionnaire-based data is no greater than the difference between data from two questionnaire surveys.

Data quality: data on persons (1)

- In general, the Population Information System can be considered very exhaustive as regards persons.
- In order that person obtain a personal identification number, he/she has to be registered in the Population Information System.
- It is practically impossible to live in Finland without a personal identification number.
- It is needed so that one can work legally, open a bank account, have dealings with authorities and so on.

Data quality: data on persons (2)

- Annual quality checks are carried out to monitor the reliability of address data, for instance, by the Population Register Centre.
- In the connection of Labour Force Survey some data are checked in Population Information System. In year 2012: 98,9 % correct address (legal place of residence)
- Migrant survey 2014: When trying to contact the sampled persons, it was confirmed that 9 % of the population with foreign background had moved abroad and 3 % could not be found, total population overcoverage 0,7 – 1 %
- Signs-of-Life 2015 (any income, social benefits, etc): overcoverage 0,7 % of the total population

Data quality: data on persons (3)

- valuable information to Statistics Finland of the quality of the source data for the use this data for statistical purposes
- e.g. data on occupation is used only as an auxiliary data for those who have moved during a year
- for compiling statistical data of tenure status we use additional data on sales of real estates/flats

Population base in Finland

- National population definition in population statistics includes those Finnish and foreign citizens who are living permanently in Finland.
- As a general rule permanent living is measured by the stay / intention to stay in the country lasting at least 12 months.
- Persons permanently in Finland equal the persons with registered place of domicile in the country (home municipality).

Population base in Finland

- In principle a place of domicile is given to
 - all Finnish citizens who live in Finland.
 - a person immigrating to Finland if s/he has a intention to stay in the country lasting at least 12 months.
- Possible undercount
 - not all immigrants register their stay even a registered Finnish domicile entitles residents to certain benefits
 - the undercoverage mostly consists of immigrants' late registration
 - Statistics Finland can estimate the total numbers of undercoverage of different groups based on the number of residence permit applications etc.

- Possible overcount
 - late/non-emigration
 - The overcoverage can be examined with register data
 - sign of life: persons with no signs of economic activities in other registers, can be removed from the population data

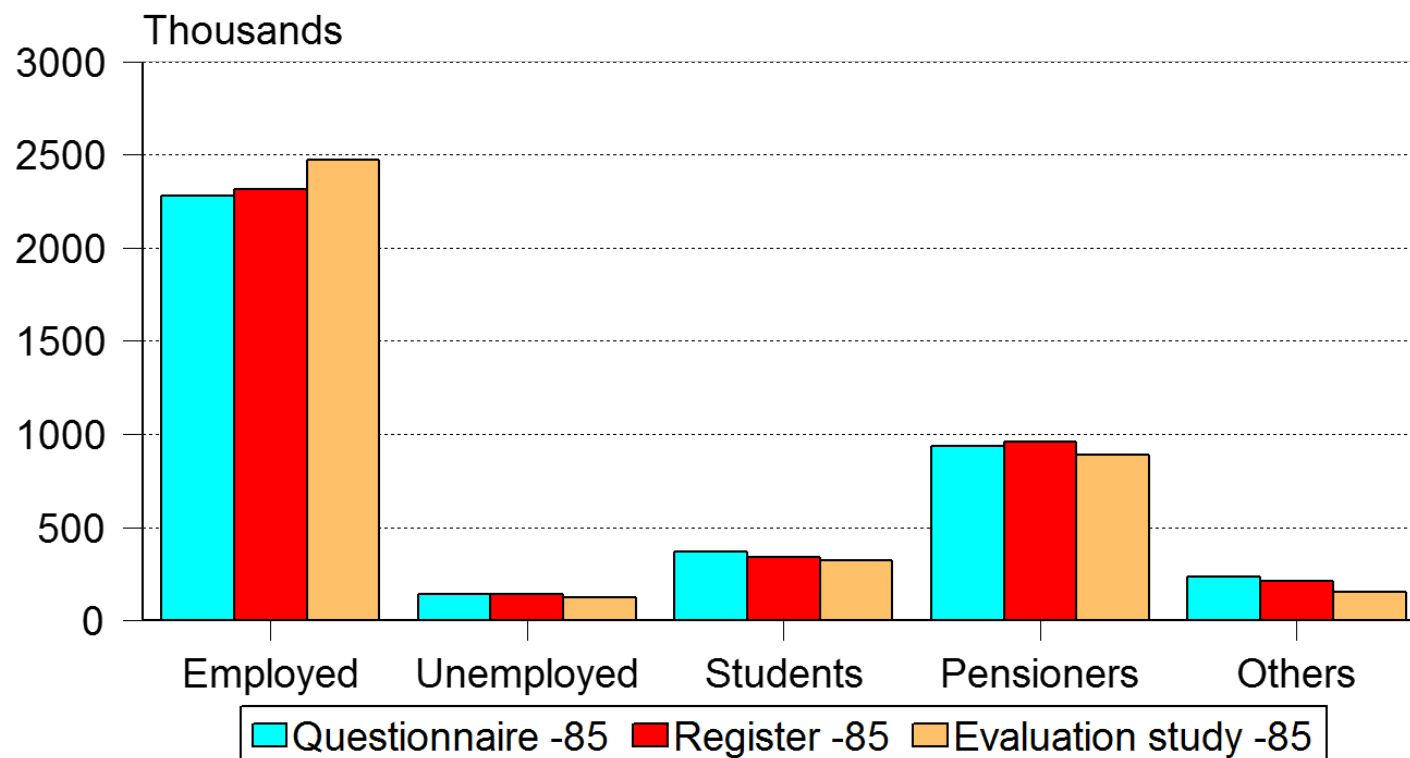
Data quality – activity status

- Municipal pilot study based on 1980 Population Census
- Register-based statistics in connection with 1985 census
- Evaluation study of the 1990 census
- Continuous quality assessment
 - Labour force survey as reference material
 - Two purposes:
 - monitoring of the level of the results
 - monitoring of the extent to which the methods produce data classified in the same manner

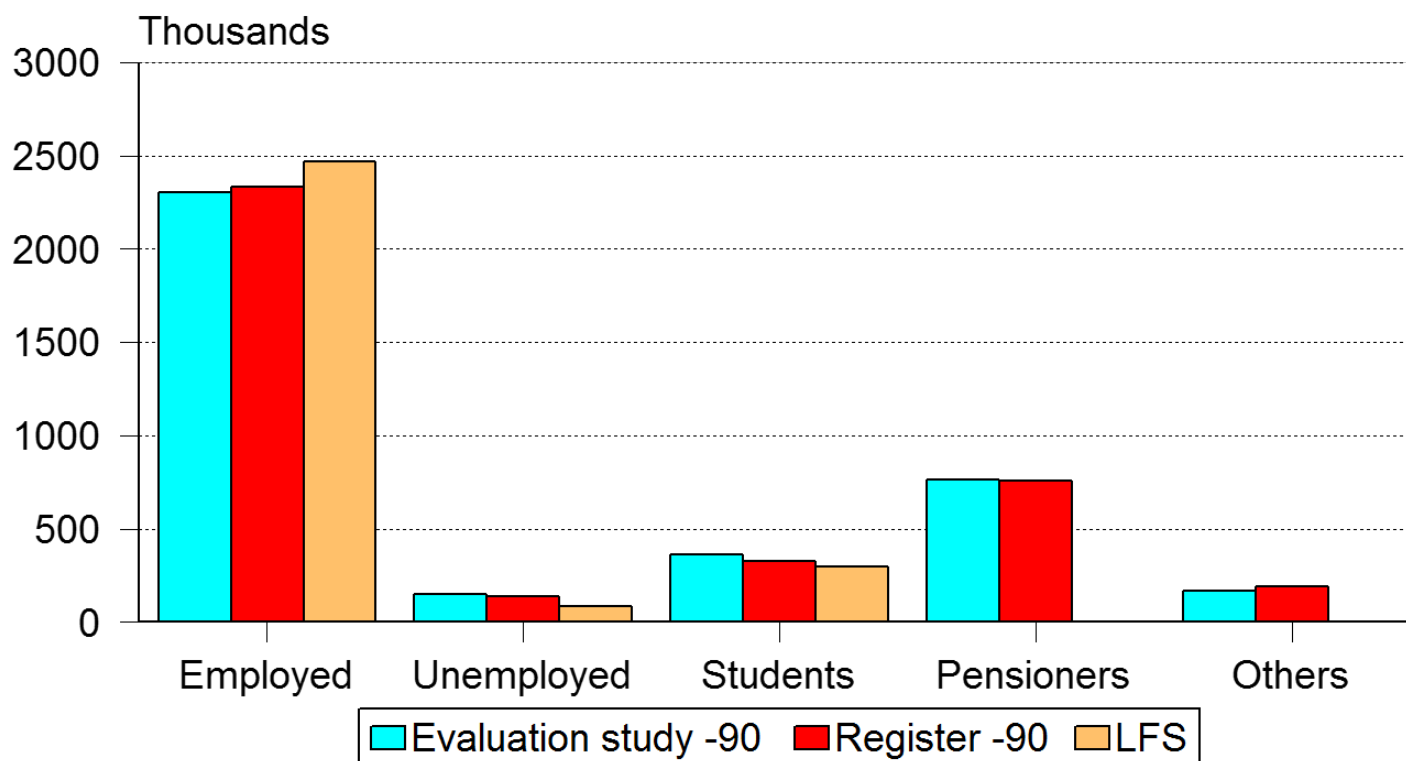
Data quality – activity status

- Monitoring of the extent to which the methods produce data classified in the same manner
 - to identify errors in data processing
 - to identify situations requiring a change in decision rules
 - to check the level of results

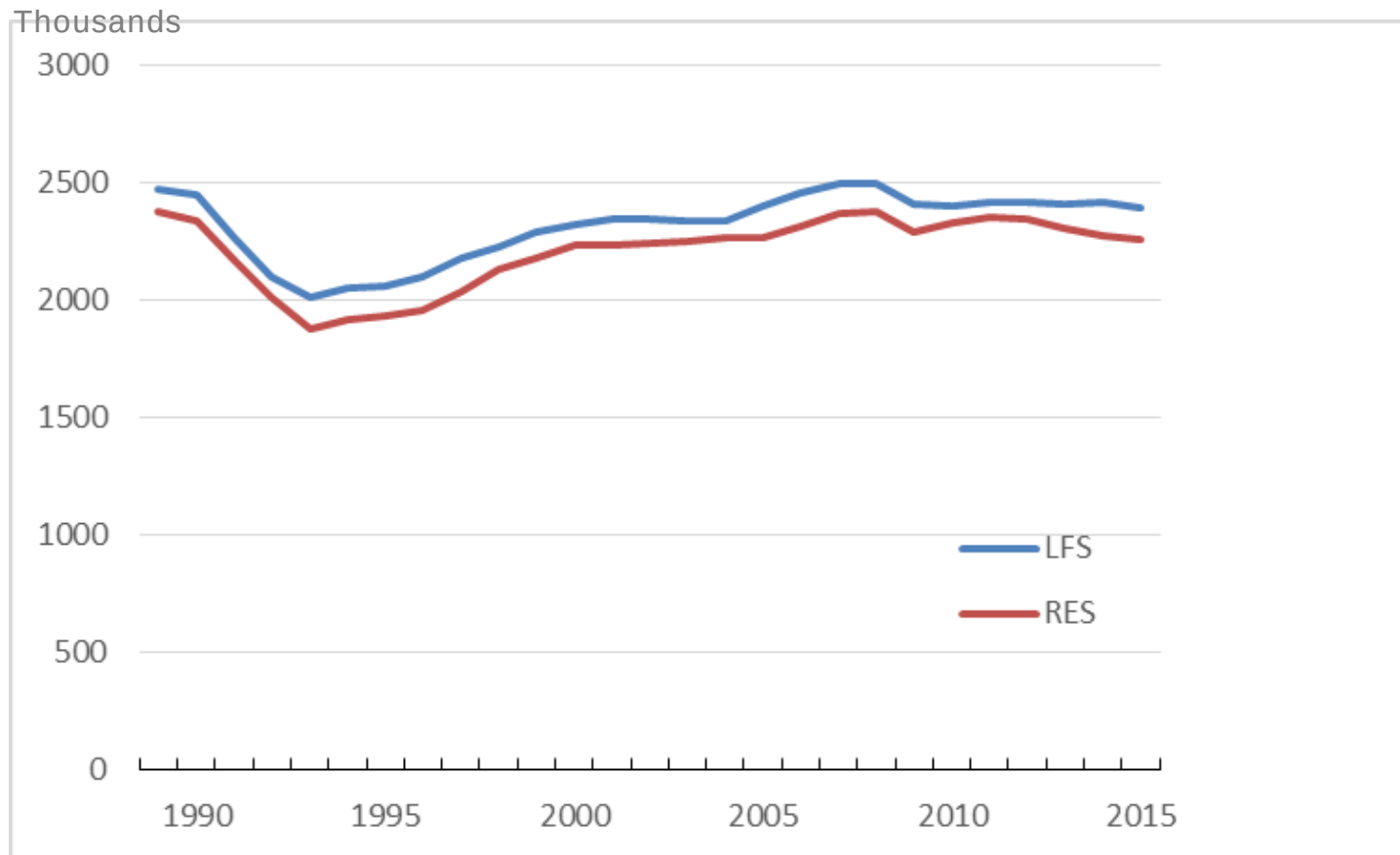
Activity status by questionnaire and by register 1985



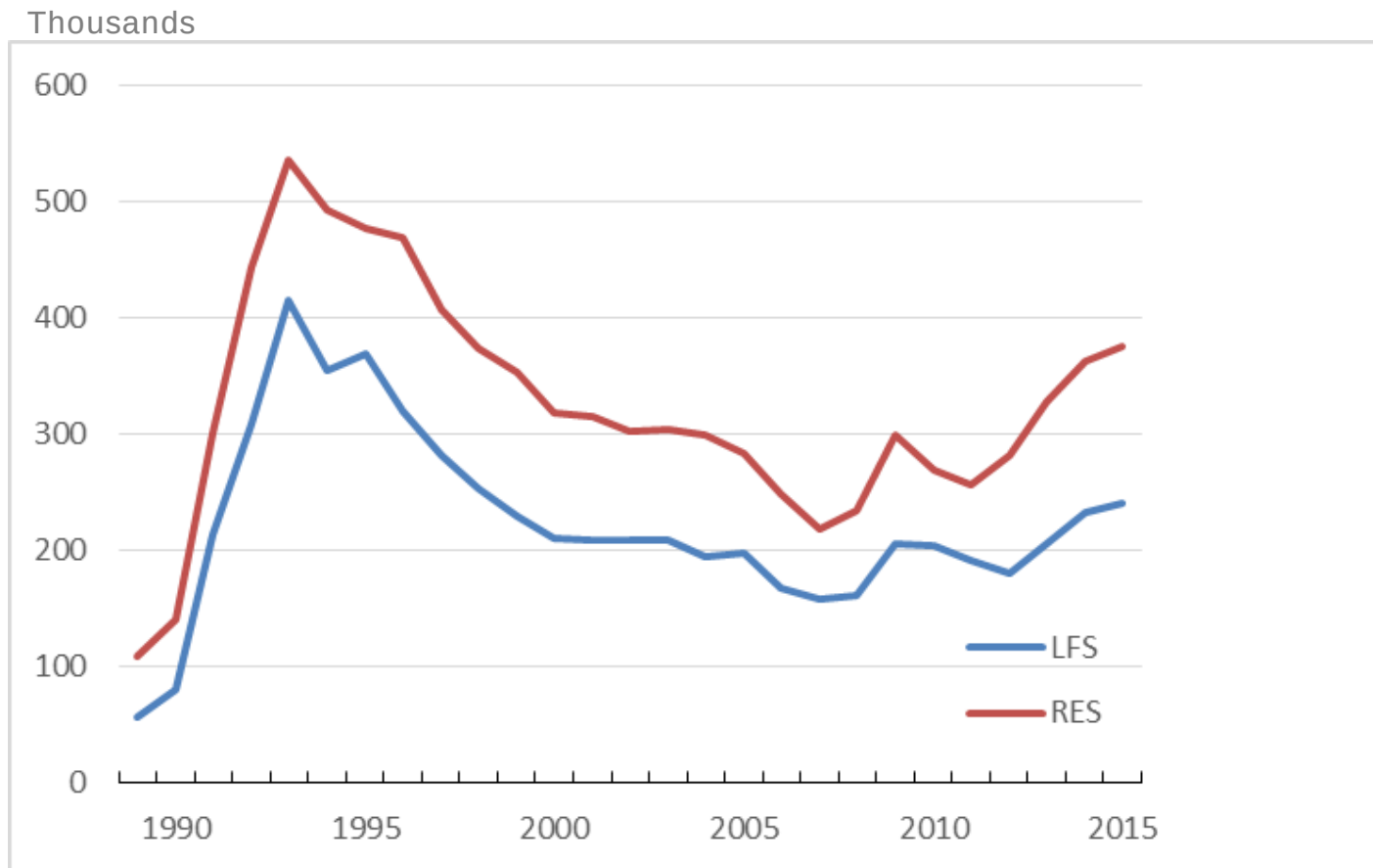
Activity status by questionnaire and by register 1990



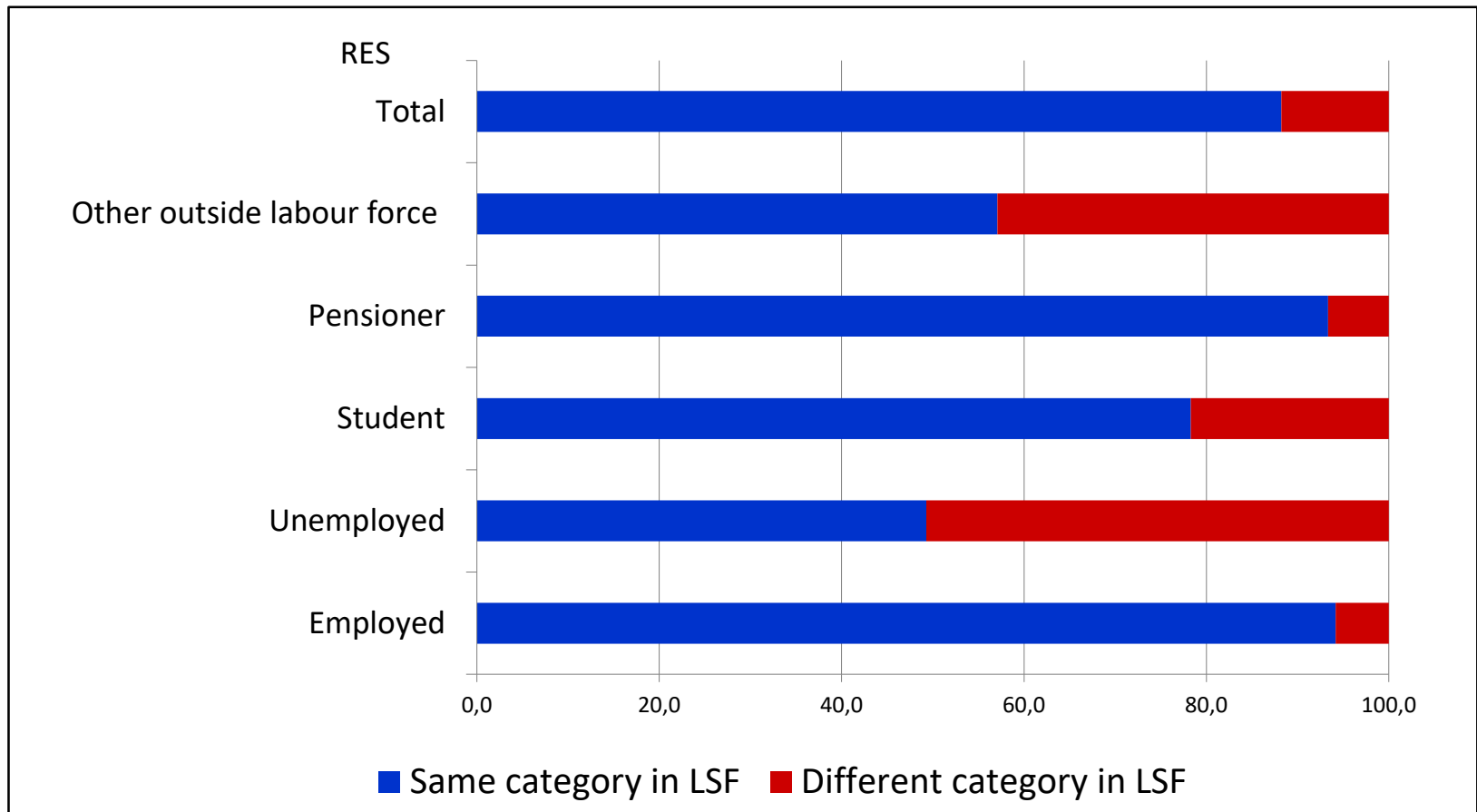
Employed according to the Labour Force Survey (LFS) and Register-based Employment Statistics (RES)



Unemployed in LFS and RES



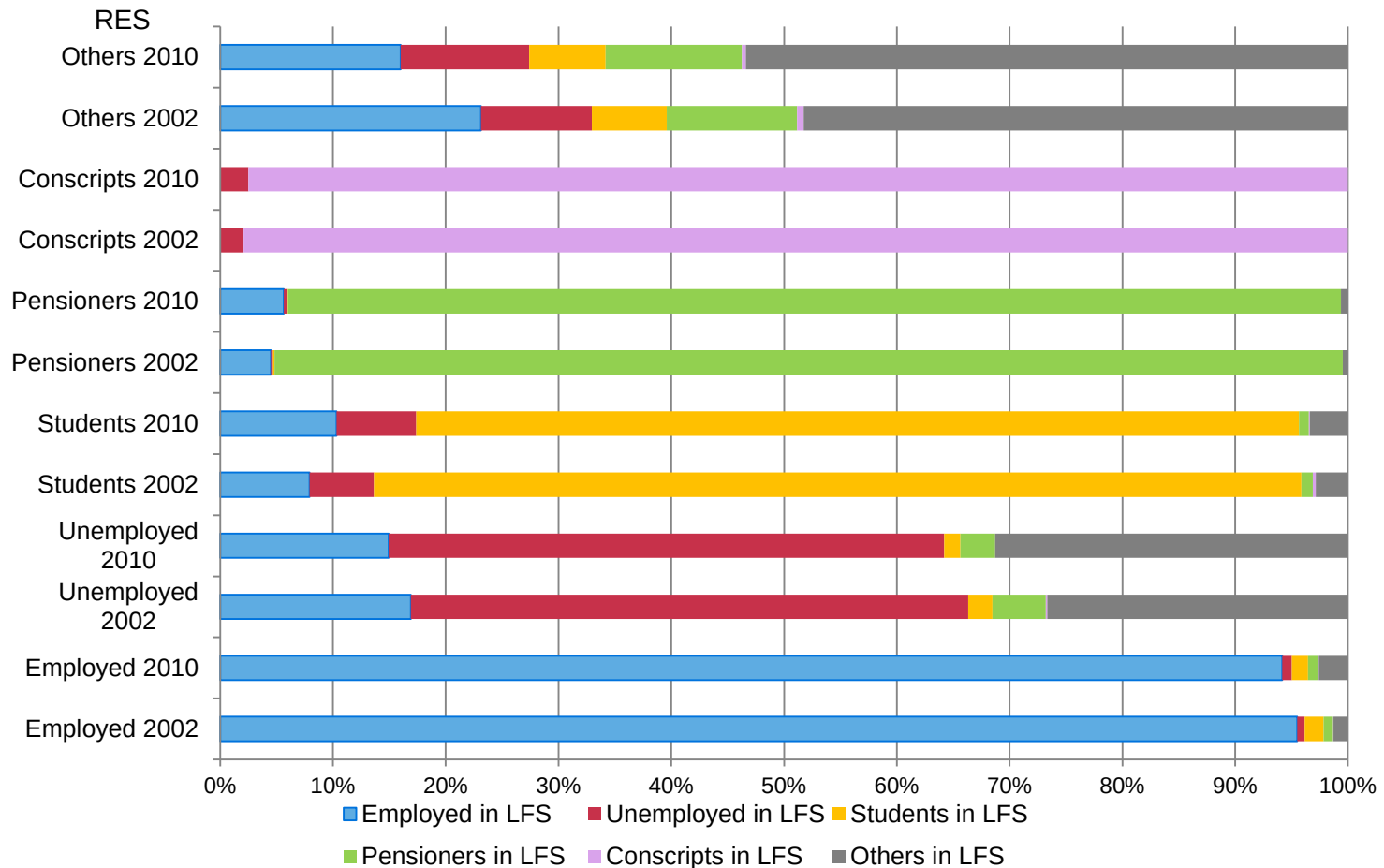
Main type of activity - Per cent classified in the same and different categories in the Labour Force Survey and Register-based Employment Statistics (without non-response)



Persons according to the Register-based Employment Statistics (RES) and Labour Force Survey (LFS) on December 2010 (persons)

RES	LFS							
	Total	Em- ployed	Unem- ployed	Stu- dents	Pensio- ners	Cons- cripts	Others	Non response
Total	9 295	5 346	432	915	2 051	42	509	2 848
Employed	5 303	4 994	45	76	50	1	137	1 474
Unemployed	556	83	274	8	17		174	252
Students	1 036	107	73	811	9	1	35	306
Pensioners	2 079	117	7	1	1 941		13	524
Conscripts	40		1			39		10
Others	281	45	32	19	34	1	150	282

Percentages classified in various categories in register-based employment statistics (RES) and the Labour Force Survey (LFS) in 2002 and 2010



Use of Survey Data (LFS) to Evaluate the Quality of Register-based Census in Finland

Presentations in Expert Group Meeting on Censuses Using Registers
22 - 23 May 2012 Geneva.

In English:

[http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/
use_of_register/WP_11_Finland.pdf](http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_11_Finland.pdf)

In Russia:

[http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/
use_of_register/WP_11_Finland_rus.pdf](http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_11_Finland_rus.pdf)



**Thank you for your
attention!**

Developing a register-based statistical system for production

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- General context
- Overview
- Development
- Conditions
- Process
- Pros and cons
- Conclusions

General context

Costs

- NSIs have faced budget cuts/restrictions
- New and more detailed statistics demand
- Efficiency in production of statistics

Burden

- To lower administrative/statistical burden
- Non-response

General context

Administrative data

- For the NSIs almost no costs for data collection
- Allocate resources for improving existing data instead of collecting data
 - **Complement and correct existing data**
 - **Establish administrative - register statistical systems**
 - **Maintain the systems**

General context

- Register data is based on administrative definitions not on statistical definitions;
- Registers have good quality for administrative purposes;
- Combining data from different registers;
- Additional data collection might be necessary.

Register-based statistics is not free of charge but less expensive than sample surveys and especially than traditional censuses

Overview

Register-based statistical system refer to a system based primarily on administrative data that have been structured to be linked to statistical registers.

Statistical registers are seen as part of the same system and not as single registers.

Implications for all stages of statistical production: data collection, data processing, quality control and dissemination.

UNECE, Register-based statistics in the Nordic countries

Step-by-step approach

Identify



Development

Step 1: Identify the need and sources of information (Where to start)

- *Costs versus benefits: costly data collection (human and financial resources), resources for processing and producing data*
- *Basis for the use of register data for statistical purposes*
- Inventory of existing sources/registers and variables;
- Identification code;
- Investigation of variables: definitions, coverage, scope etc;
- Agreements with data owners - legal frame: data to be transferred, deadlines, data format, periodicity, confidentiality agreements, roles and responsibilities;

Development

- Access to forms used for collecting data
- Record descriptions
- Standards and classifications used

Development

Examples

- Population and dwellings/housing census
- Employment statistics

Development

Step 2: Test (what to do)

- Relevance
- Timeliness
- Exhaustiveness of requested statistical variables
- Linking and combining several registers
- Tools to process data

Development

Step 3: Compare (what to do)

- Definition of variables
- Level of details
- Time reference
- Data produced based on both sources, if possible

Development

Step 4: Utilize

- Redesigning and structuring the whole system
- Coordinating to minimize duplicate work and increase consistency between different registers.
 - ✓ Organizational coordination
 - ✓ Technical
 - ✓ Methodological coordination

Produce statistics

Development

Table. The year of establishing registers/introducing registers in census statistics by type of register and country

	Denmark		Finland		Norway		Sweden	
Type of register	<i>Estab- lished</i>	<i>First used in census</i>	<i>Estab- lished</i>	<i>First used in census</i>	<i>Estab- lished</i>	<i>First used in census</i>	<i>Estab- lished</i>	<i>First used in Census</i>
Central Population Register	1968	1981	1969	1970	1964	1970	1967	1975
Business Register	1975	1981	1975	1980	1965	1980	1963	1975
Dwellings	1977	1981	1980	1985	2001	2011	2008?	2011?
Housing conditions	1977	1981	1980	1985	2001	2011	2008?	2011?
Education	1971	1981	1970	1975	1970	1980	1985	1990
Employment	1979	1981	1987	1990	1978	2001	1985	1985
Family	1968	1981	1978	1980	1964	1980	1960	1975
Household ^a	1968	1981	1970	1975	2001	2011	2011?	2011?
Income	1970	1981	1969	1970	1967	1980	1968	1975
Totally register-based census		1981		1990		2011		2011?

^a Household-dwelling unit, i.e. all the persons living in one dwelling

UNECE = Register-based statistics in the Nordic Countries

Conditions

Legal base

NSI has the right to access administrative data on unit level with identification data and to link them with other administrative registers for statistical purposes.

Data protection/confidentiality

Provisions regarding processing of personal data to ensure that the use of registers containing personal information doesn't violate the legal rights of the individual citizen with regard to the protection and integrity of his/her data.

The existence of suitable administrative sources: Comprehensive administrative registers of target populations are essential

Conditions

Public acceptance

- the attitude of the general public to data linking and sharing within the government sector is a key factor in using such data for statistical purposes;
- balance between the efficiency of data sharing and concerns about the protection of data relating to individual units;

general public acceptance of the benefits and broad public approval of the use of administrative data for purposes of statistics production

- legislation is up-to-date and the work of the register authorities is open and transparent.

Conditions

Unified identification systems

- major factor that facilitates the statistical use of administrative data records across different sources
- minimum requirement is to have a unified identification system for base registers,
 - **unified personal identity codes (personal identification numbers);**
 - **unique identification code for enterprises/legal units**

Link different registers without unified identification codes more laborious and time consuming.

Conditions

Comprehensive and reliable register systems developed for administrative needs

- Comprehensive administrative registers of target populations are essential. The existence of large numbers of unregistered units, will make it extremely difficult to produce reliable register-based statistics.
- Mostly systems are ruled by the state therefore it has been necessary to establish registers on a state level;
- The purposes of administrative systems are connected and register information is exchanged between the institutions.

Conditions

Cooperation among administrative authorities

Moving to register-based statistics production needs strong and clear commitment of the highest level (government/legislators) and close collaboration among relevant authorities.

- the government should provide political support to the NSI efforts in developing a register-based statistical system;
- strong support to the NSI in their negotiations with administrative authorities on access to administrative data.

The administrative sources should be available to statisticians in a format that allows transfer of data.

Process

- The Population/Business Registers are the backbones of the system for persons/business
- Other files matched to the Population/Business Register, such that the true matches are maximised (aim: no missed matches) and the false matches (mismatches) are minimised

Matching variables:

- Identifier number
- Other personal identifiers: sex, date of birth and address
- Other business identifiers: name, activity and address

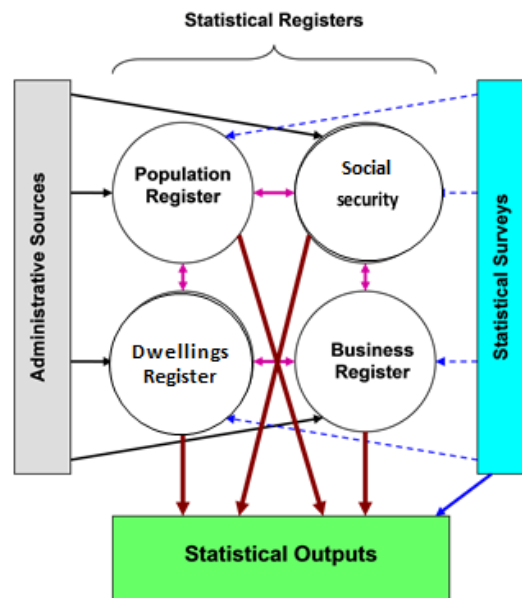
Process

- check the linked data and modify incorrect records;
- the results that are to be published are of higher quality than the original sources

an integrated process of: data editing, derivation of statistical variables and imputation to be executed

Rules for the same variable in two sources or relationship between two or more variables in one or more sources.

Process in Nordic countries



|

Pros and cons

Pros

- total coverage with relatively low cost of collection and processing;
- can produce more detailed statistics than by using sample surveys,
- offer a large potential because different registers can be linked together on the basis of clearly defined identifiers;
- administrative registers are usually consistent and of high quality,

Cons

- consistency of data
- definitions and administrative practice of the authorities responsible for the registers.
- administrative data from authorities' perspective
- not enough information about the precise data content, data processing, data quality

Conclusions

- Costs
- Response burden
- Relevance
- Accuracy
- Accessibility
- Timeliness

Thank you for your attention !

Linking register-based statistics and Geographical Information System

Kaija Ruotsalainen

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- Concepts and preconditions for grid based statistics
- Need and use of small area/grid statistics
- Some examples of small area/grid statistics
- Special characteristics of the GIS data
- Challenges and benefits
- European cooperation

Preconditions for small area / grid based statistics

- Data with location data
- Grids
- GIS software

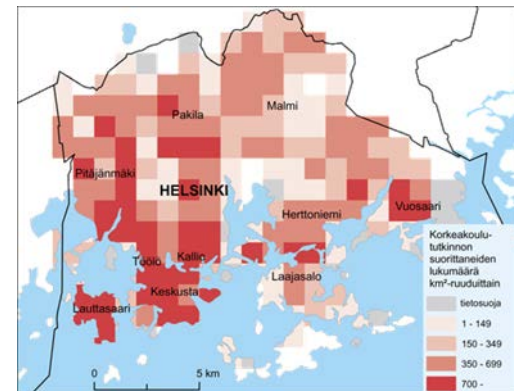
Concepts: Grid / Grid statistics

Grid

- Grid net
- It covers the whole geographical area

Grid statistics

- Aggregated statistical data
- It means summed up statistical
- Contains those grids in which there are calculated statistical data (“cases”)



Different area divisions

Administrative regions

- NUTS1, NUTS2, NUTS3
- LAU2 (Municipalities)
- Election districts
- ...

Areas not connected to the administrative boundaries

- grids (e.g. 250 m x 250 m or 1 km x 1 km)
- subareas of municipalities
- postal code areas
- Urban areas

- other(e.g. zones, sectors, distance...)

Why small area / grid statistics?

- A growing demand for small area statistics from all sectors
 - Central and local government bodies
 - Commercial and business
 - Research
 - Media
 - General public and citizens

How the small area / grid-based data is used?

- Small area statistics support and facilitate
 - better informed decisions
 - evidence based policies
 - citizens' engagement and empowerment; democracy
 - better quality of life
 - better allocation and use of resources
 - better performance of the city and region
 - improved businesses
 - monitoring and measuring impact of territorial and place-based policies and programmes

Grid-based statistics

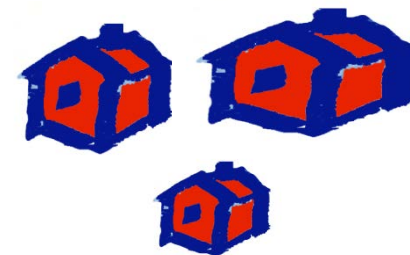
- regional statistics in which regional entities are defined by geographically referenced grid cells
- statistical variables are calculated and displayed on a regular grid net
- So called bottom-up method by direct aggregation with the help of point-based, detailed georeferenced data produces the best quality results

Detailed georeferenced data

Persons: PIN



Buildings:
building-id



building-id

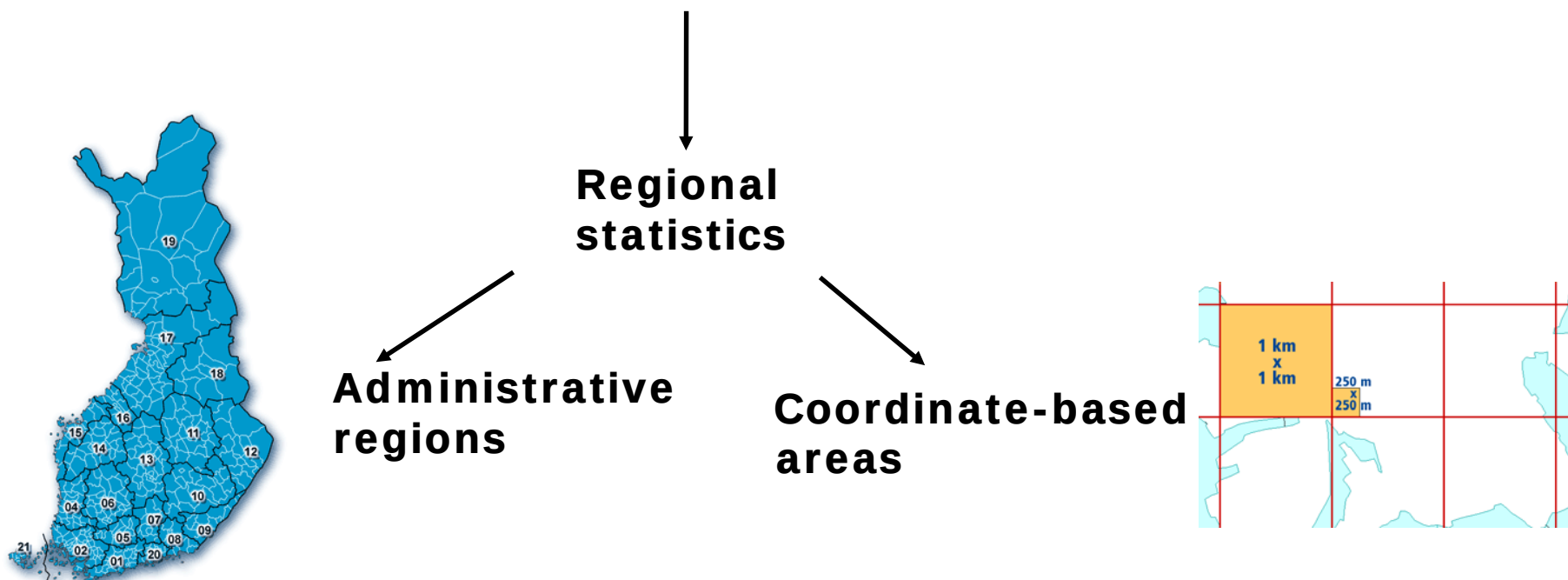
business-id

address



Enterprises
Establishments

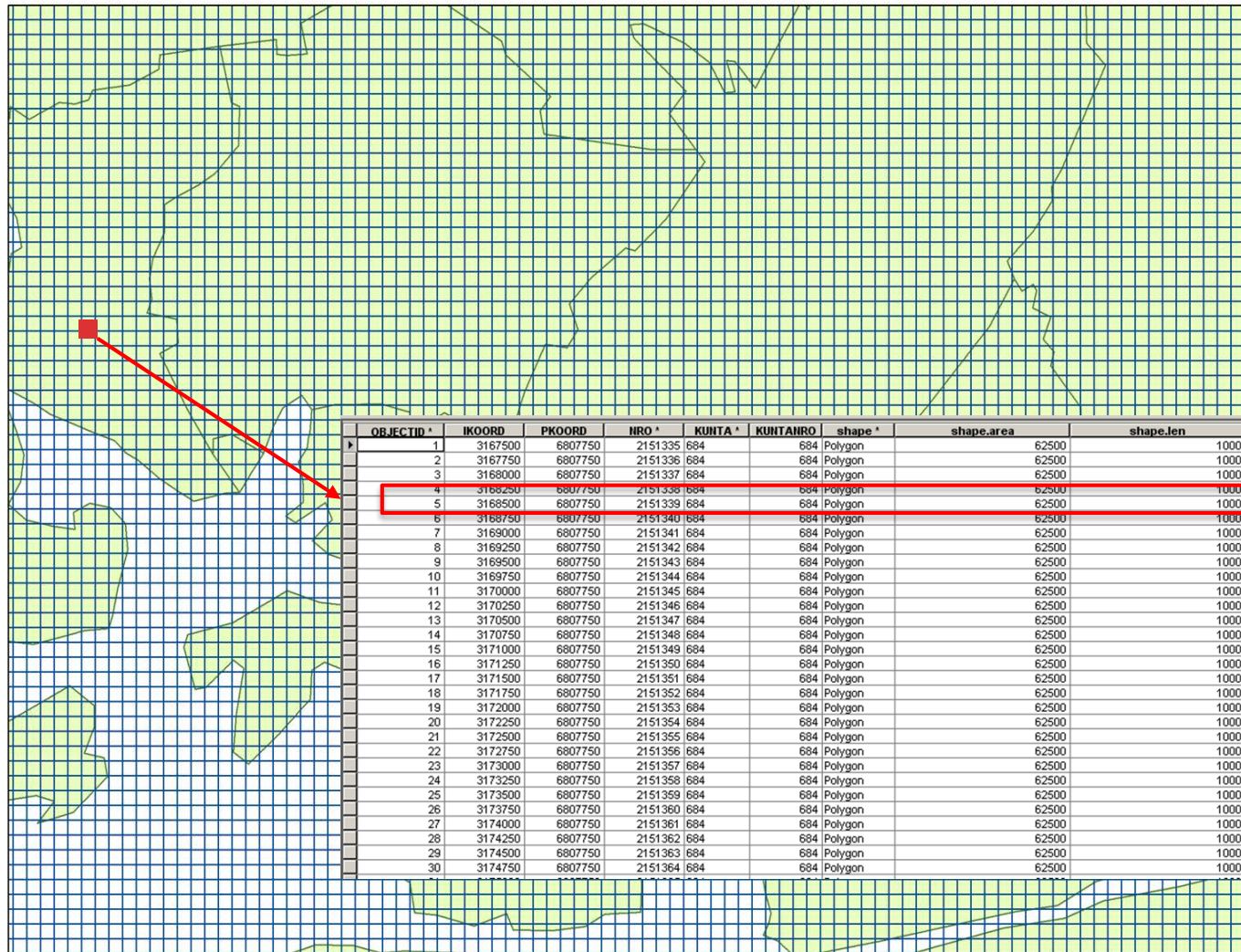
Registers **→** **Statistical data with georeferenced data**





European
Commission

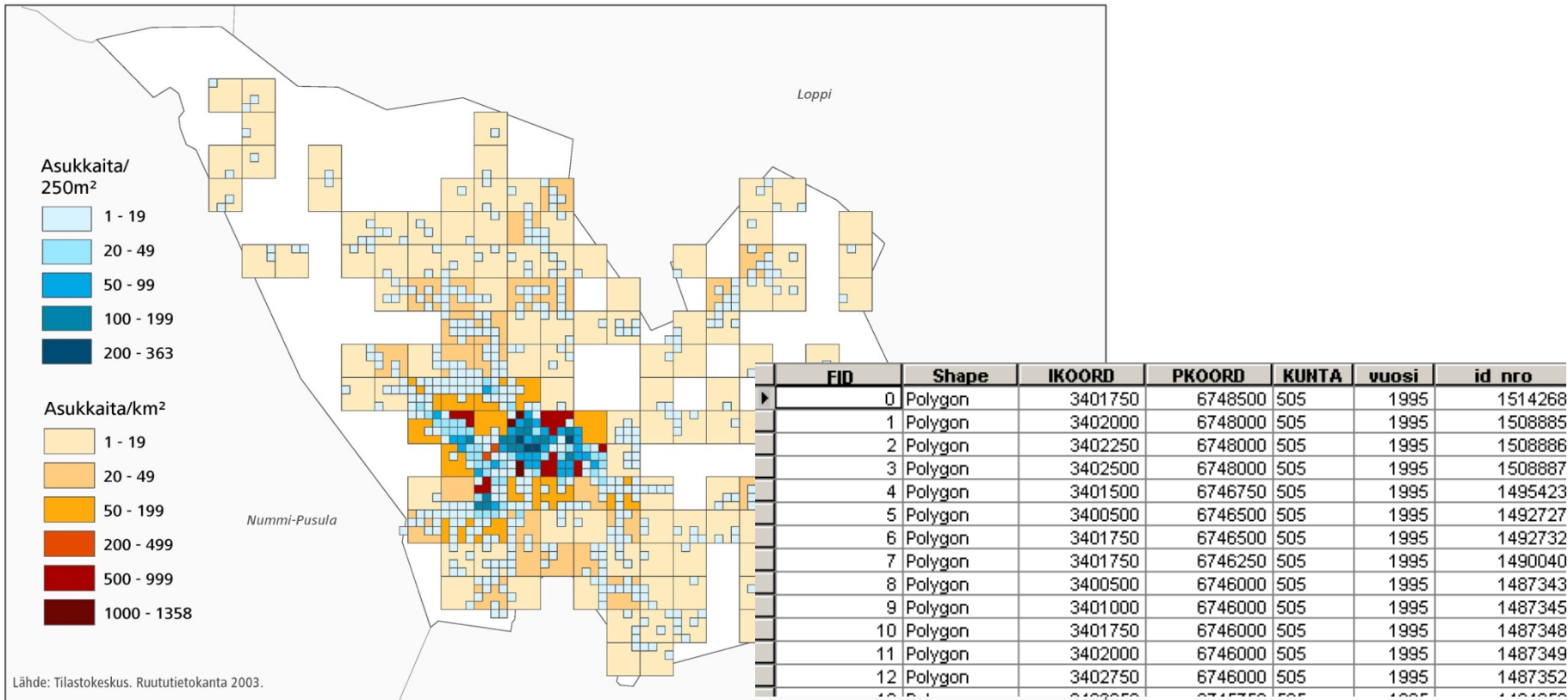
A grid for representing thematic information is a system of regular and georeferenced cells, with a specified shape and size, and an associated property”
(European Reference Grids, Workshop 27-29 October 2003, Ispra EC/JRC)



Example: population density

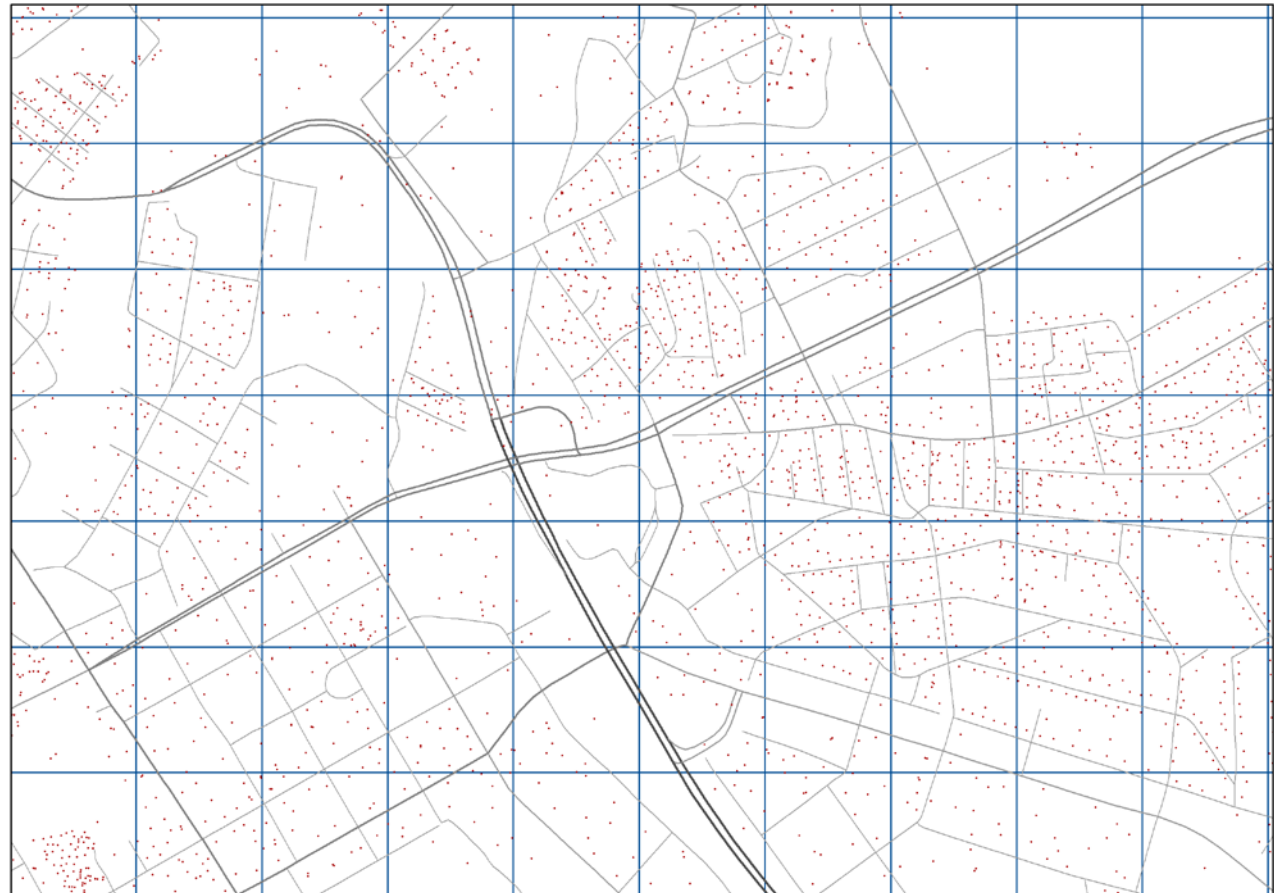
Väestötiheys Karkkilassa 250 m x 250 m ja 1 km x 1 km -ruuduittain

 Tilastokeskus



Statistics Finland: the data content of the Grid database

FID	1198	TK_KUTY	121
Shape	Polygon	TK_PL_TUL	65
XKOORD	3241750	TK_KE_TUL	28
YKOORD	6714000	TK_HY_TUL	28
KUNTA	853	TK_OV_KER	23797
ID_NRO	1142132	RA_RAKY	61
HE_VAKIY	232	RA_ASRAK	57
HE_NAISET	107	RA_ASUNN	146
HE_MIEHET	125	RA_AS_KPA	51,3
HE_KIKA	40	RA_AS_VALJ	29,7
HE_VAUVAT	3	RA_PT_AS	121
HE_LEIKK_I	8	RA_KT_AS	22
HE_ALLEK_I	11	RA_DMIS_AS	64
HE_ALAAS_I	12	RA_VUOK_AS	54
HE_TEINI_I	8	TP_TYOPY	90
HE_LAPSET	31	TP_ALKU_AB	0
HE_NUORET	25	TP_JALO_CF	12
HE_VAK_IKA	50	TP_PALV_GQ	78
HE_VAKIINT	31	TP_A_MAAT	0
HE_KESK_I	34	TP_B_KALAT	0
HE_VARTN	61	TP_C_MINER	0
HE_AKTV_I	179	TP_D_TEOLL	3
HE_ELAKK_I	22	TP_E_ENER	0
HE_VANH_I	13	TP_F_RAKEN	9
KO_IKA18Y	212	TP_G_KAUPP	8
KO_PERUS	71	TP_H_MAJD	4
KO_KOUL	141	TP_I_KULJ	1
KO_YLIO	31	TP_J_RAH	0
KO_AMMAT	72	TP_K_KIINT	7
KO_KORK	38	TP_L_JULK	0
HK_KTU	19012	TP_M_KOUL	0
HK_MTU	14865	TP_N_TERV	40
HK_TUY	209	TP_O_MUUT	18
HK_PL_TUL	96	TP_P_TYONA	0
HK_KE_TUL	56	TP_Q_KVJAR	0
HK_HY_TUL	57	TP_X_TUNT	0
HK_OV_KER	3015429	TV_TYOVY	132
TE_TALY	121	TV_TYOLL	110
TE_TAKKOKO	1,9	TV_TYOTT	22
TE_NUOR	35		
TE_EIL_NP	11		
TE_LAPS	14		
TE_PLAP	2		



Special characteristics of the GIS data

- Size and accuracy of the statistical area
- Scale of the visual presenting
- Size of the population
- Sensitivity of the variables, (for example income, religion, nationality)
- The number of the variables/other data (a cross tabulation)
- Absolute or relative figures

Disclosure control

- One of the major challenges
- A detailed map as background information for grid data may increase the need for data protection.
- In sparsely populated countries/areas grid-based statistics face confidentiality problems, especially in rural areas.

Disclosure control

- Possible solutions:
 - Increasing the size of the grid cell
 - Confidential data to be suppressed -> may cause crucial effects on the results of a spatial analysis
 - =>The user of the data must therefore be aware of how confidential data have been processed in order to understand the potential impacts on his/her analysis.
 - The data protection measures usually follow the general guidelines on disclosure control with the help of simple data suppressing.

Disclosure control

- Hardly any methods are used which take into account the special features of geographic information.
- For grid data: the Local Restricted Imputation method (LRI), where data protection is made locally so that the data will always be accurate at a hierarchically higher area level.
- Grant project (2016-2017) Harmonised protection of Census data
 - Guidance on the protection of the tables for the 2021 Census
 - recommend statistical disclosure control methods.
 - recommend how to handle confidential cells in grid squares and regional breakdowns.

Benefits of the co-ordinate based statistics

- Independent of the administrative boundaries
- To be connected to administrative areas - if needed
- Points/Grids do not change (space, time)
- Grids comparable inside a country and between countries
- Grid net for harmonising statistics by different kinds of territorial units – “Location the unifying factor”
- Grids as a basic unit for compiling statistics flexibly by small to larger areas, by natural boundaries, by distances...
- Grids for comparable functional areas (e.g. urban-rural)
- Grids for spatial analysis (accessibility, neighbourhood...)

European cooperation

- European Forum for Geostatistics
- ESSnet project Geostat
 - Harmonised population grid data
 - (1km x 1km grid cells) of 2006, 2011
 - Guidelines to produce population grid data
 - Examples of the use of grid data
 - European Forum for Geostatistics <http://www.efgs.info/>
- Inspire

INSPIRE (Infrastructure for Spatial Information in Europe)

- The Finnish Geoportal (CSW, WMS, WFS) by the National Land Survey (today > 500 map layers by 50 data providers <http://www.paikkatietoikkuna.fi/web/en/map-window>)
- SF's data available according to Inspire specifications since May 2013 (24 data layers)
 - Statistical units (Major regions, sub-regional units, municipalities, 1km x1km grids...)
 - Population by statistical units (total, sex, age) incl. grids!
 - Educational institutions
 - Production and Industrial Facilities<http://tilastokeskus.fi/tup/rajapintapalvelut/index.html>



**Thank you for your
attention!**

Developing a register-based statistical system-Romanian experience

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- Business demography statistics
- Employment statistics
- Building permits statistics

Business demography statistics

Business demography refers to those indicators which explain the characteristics of the business population.

- Events in the life cycle of an enterprise such as births and other creations of enterprises, deaths and other cessations of units, and their ratio to the business population;
- The follow-up of enterprises over time offering information on their survival or discontinuity;
- Development over time of certain characteristics like size;
- High-growth enterprises.

Business demography statistics

COMMISSION REGULATION (EC) No 251/2009 as regards the series of data to be produced for structural business statistics

- Annex 9 Business demography
- Annual demographic statistics broken down by legal form
- Annual demographic statistics broken down by employee size classes
- Annual preliminary results on enterprise deaths, broken down by legal form
- Annual preliminary results on enterprise deaths, broken down by employee size classes

Business demography statistics

- Population of active enterprises in t
- Number of births of enterprises in t
- Number of deaths of enterprises in t
- Number of persons employed in the population of active enterprises in t
- Number of employees in the population of active enterprises in t
- Number of persons employed in the population of births in t
- Number of employees in the population of births in t
- Number of persons employed in the population of deaths in t
- Number of employees in the population of deaths in t

Business demography statistics

- Number of enterprises newly born in $t-1$ having survived to t
- Number of enterprises newly born in $t-2$ having survived to t
- Number of enterprises newly born in $t-3$ having survived to t
- Number of enterprises newly born in $t-4$ having survived to t
- Number of enterprises newly born in $t-5$ having survived to t

Business demography statistics

- Number of persons employed in the population of enterprises newly born in $t-1$ having survived to t
- Number of persons employed in the population of enterprises newly born in $t-2$ having survived to t
- Number of persons employed in the population of enterprises newly born in $t-3$ having survived to t
- Number of persons employed in the population of enterprises newly born in $t-4$ having survived to t
- Number of persons employed in the population of enterprises newly born in $t-5$ having survived to t

Business demography statistics

- Number of persons employed in the year of birth in the population of enterprises newly born in $t-1$ having survived to t
- Number of persons employed in the year of birth in the population of enterprises newly born in $t-2$ having survived to t
- Number of persons employed in the year of birth in the population of enterprises newly born in $t-3$ having survived to t
- Number of persons employed in the year of birth in the population of enterprises newly born in $t-4$ having survived to t
- Number of persons employed in the year of birth in the population of enterprises newly born in $t-5$ having survived to t

Business demography statistics

Source of data: Statistical Business Register

Demographic events concerning enterprises in the BR

Birth

Death

Merger/Takeover

Break-up/Split-off

Business demography statistics

EUROSTAT – OECD MANUAL ON BUSINESS DEMOGRAPHY STATISTICS

Events	<u>Real, observable world</u>		<u>Business Register</u>	
	Number of enterprises	Number of enterprises	Number of creations	Number of deletions
	before the event	after the event		
Birth	0	1	1	0
Death	1	0	0	1
Change of ownership	1	1	0	0
Merger	n	1	1	n
Take-over	n	1	0	n-1
Break-up	1	n	n	1
Split-off	1	n	n-1	0
Creation of a joint venture	n	n+1	1	0
Cessation of a joint venture	n	n-1	0	1
Restructuring within an enterprise	1	1	0	0
Restructuring within an enterprise group	n	N	0 or more	0 or more
Change of group	1	1	0	0
Complex restructuring	n	N	0 or more	0 or more

Employment statistics

- Purpose: produce statistics on wages and salaries, earnings and hours worked according to ILO recommendations
- Reference period: October
- Statistical survey until 2012
- Since 2013 based on register and administrative sources
- One year using both approaches to test the consistency of data series

Employment statistics

- Source
- Process
- Statistical data
- Dissemination of data

Source

- Social security register owned by **National Agency for Fiscal Administration**
 - Form 112 The Statement of Payments of Social Contributions, Income Tax and Insured Persons registered
- **Who?**
 - Natural and legal persons who are employers
 - Entities similar to the employer who have the status of payers of income from dependent activities
 - Income payers for persons who make income of a professional nature, other than salary
 - Any payer of income assimilated to salary or wages
- **When?**
 - Monthly, exceptional cases quarterly
 - by 25th of the next month

Sources

- Submitted electronically to e-România portal
- Income tax on salaries;
- individual social security contribution retained from the insured employee;
- social security contribution due to persons for whom the payment of entitlements is covered by the unemployment insurance budget;
- insurance contribution for accidents at work and occupational diseases for the unemployed;
- social insurance contribution payable by the employer;
- contribution for health insurance from insured persons;
- the social security contribution for people on sick leave for work-related or occupational illness and other legal deductions;
- other
- number of employees
- total amount of wages

Sources

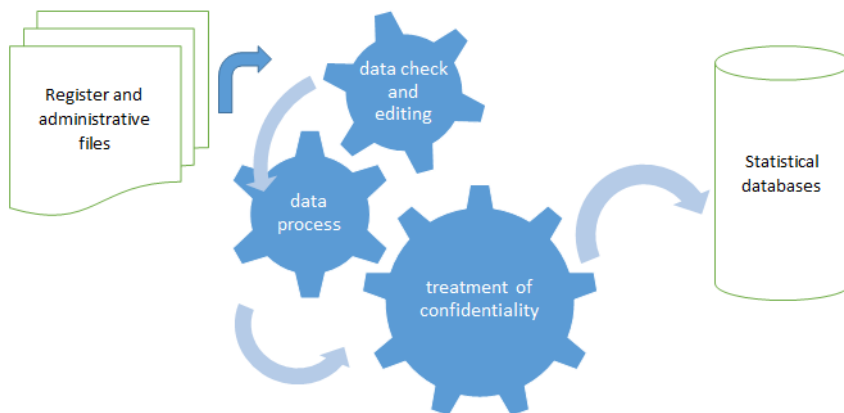
- **Register of Employees** exploited by Ministry of Labour
ReviSal (Labour Inspection)
- **Who?**
 - All employers, legal and physical persons, are required to complete ReviSal using the electronic tool developed and provided by the Labour Inspection or portal
- ReviSal includes all individual employment contracts in force. This database records the registration, changes or termination of employment contracts.

Sources

- identification of all employees: name, personal identification number - CNP, citizenship and country of origin;
- date of employment;
- period of secondment and the name of the employer;
- occupation according to the Romanian Occupation Classification/ISCO;
- type of employment contract;
- duration of the working time;
- gross monthly gross salary and bonuses;
- period and causes of temporary leave, except for the cases for medical purpose;
- date of termination of the employment contract.

Process

- Data provided to NIS based on agreements with owner of these data
- Secure environment
- Strict access to data according to the roles



Statistical data

Main indicators

- Number of employees with FTE
- Wages and salaries

Breakdown

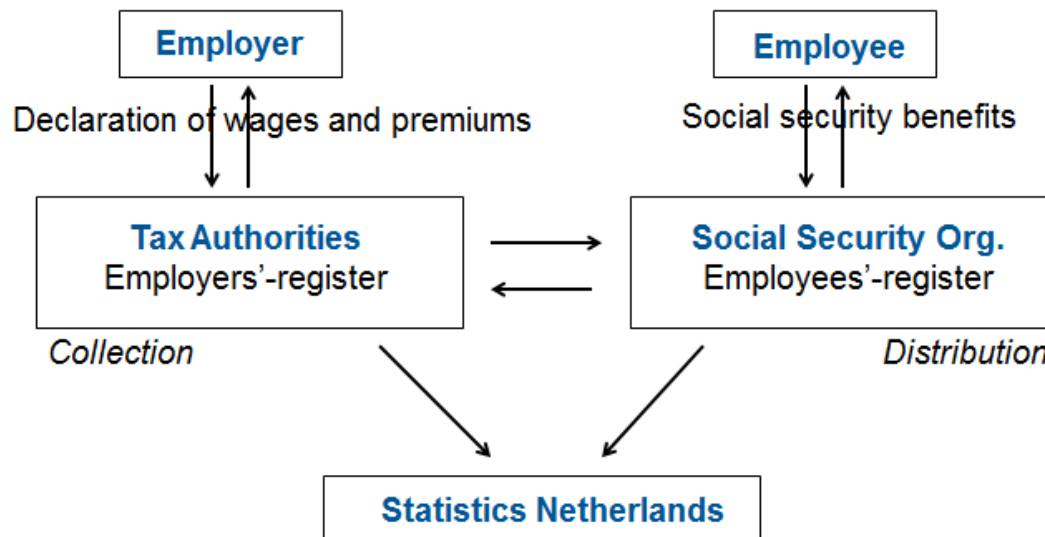
- NACE activity
- Gender
- Occupation

Dissemination of data

- Publication: Breakdown of employees by average earnings in October 20..., Women and men
- NIS online database
- ILO database

Employment statistics

Administrative data source (1)



Building permits

Council Regulation (EC) no 1165/98 concerning short term statistics

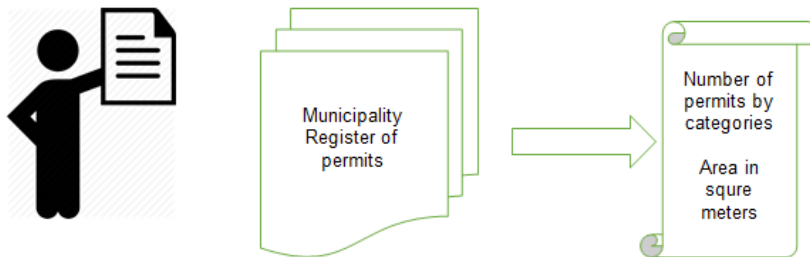
- **building permits** or **construction permits** is a leading indicator in the business cycle which provides some information about the workload of the construction industry in the near future.
- A building permit is the final authorisation to start work on a building project.

Two indices for building permits, representing different aspects:

- the number of dwellings;
- the square metres of useful floor area.

Building permits

- Flow of information



Conclusions

Despite the difficulties faced:

- identification code confidentiality issues
- different variables collected
- slightly different definition
- quality of data received
- data owners commitment and internal rules

The systems for data production are functional !!!

Thank you for your attention !

Innovative use of statistical data based on registers

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- Evidences
- Longitudinal approaches
- Micro-data labs
- Linking registers data

Evidences

- Statistical data demand increasing
- Costs to be maintained or reduced
- Burden on respondents to be kept at a “reasonable” level

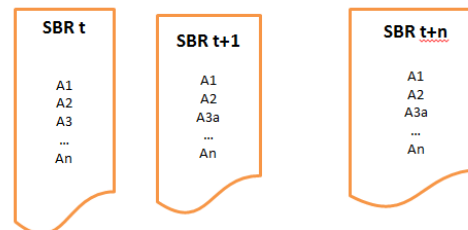


- develop statistical methods and software to analyse large micro-databases and to link the existing data: socio, economics, demographic and other administrative registers;
- new data sources to be exploited

Longitudinal approach

Statistical Business Register

- Sector analysis over time
- Who are the driving actors?



Longitudinal approach



Economic entities

Time

Activities

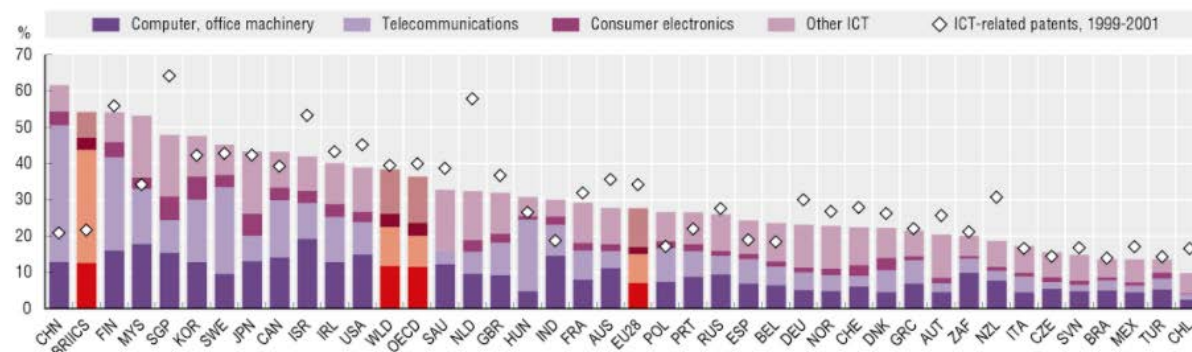
Micro-data labs

OECD Measuring the Digital Economy A New Perspective: A New Perspective

- OECD has developed a micro-data lab: compiles and links large scale data;
- Datasets on patents, trademarks, design rights
- Information about companies
- Analysis of emerging technologies and their links to companies performance.

Micro-data labs

Specialisation in ICT-related patents, 1999-2001 and 2009-11
ICT-related patents as a percentage of total PCT patent applications, by country

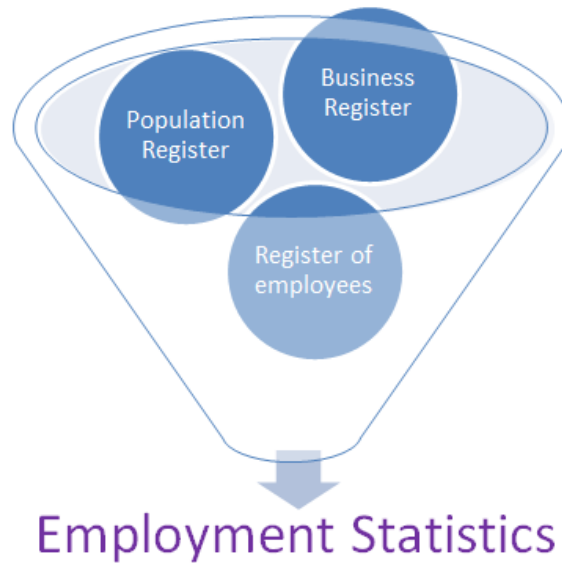


Micro-data labs

Umeå University created Stat4Reg Lab: STATistical research laboratory FOR the analysis of REGister data

<http://www.stat4reg.se/>

Linking Registers data



Linking registers data

- Changes in employment specialisation
- Dynamics of employment by sectors
- Regional flow of employment
- Employment migration





Innovative use of statistical data based on statistical registers

Kaija Ruotsalainen

ESTP - Moving towards register based statistical system
13 – 15 September 2017, Valencia, Spain

THE CONTRACTOR IS ACTING UNDER A FRAMEWORK CONTRACT CONCLUDED WITH THE COMMISSION

Content

- History of the administrative registers in Finland
- History of the Population Censuses in Finland
- From the traditional Census to the register-based model
- New possibilities for research and statistics production based on administrative sources
 - **follow up**
 - **regional statistics**
- Examples

History of Administrative Registers (1)

- The first exhaustive register of persons was established by Social Insurance Institution in 1960's
- The personal identity code was introduced 1964
- Later it was widely introduced in other administrative registers
- Earnings-related pension system began in 1960's, and in 1970 enlarged covering also self-employed
=> the data both on employees, self-employed and pension recipients were registered

History of Administrative Registers (2)

- Central Population Register was established 1969
=> Population Register as a basic register on population for all authorities
- Unemployment benefit developed strongly in 1970's. Also statistics on unemployed based on register started.
- Data collection for the Register of Buildings and Dwellings in the connection of 1980 Census
=> Data entered to the register which was owned by Population Register Centre

History of Finland's Population Censuses (1)

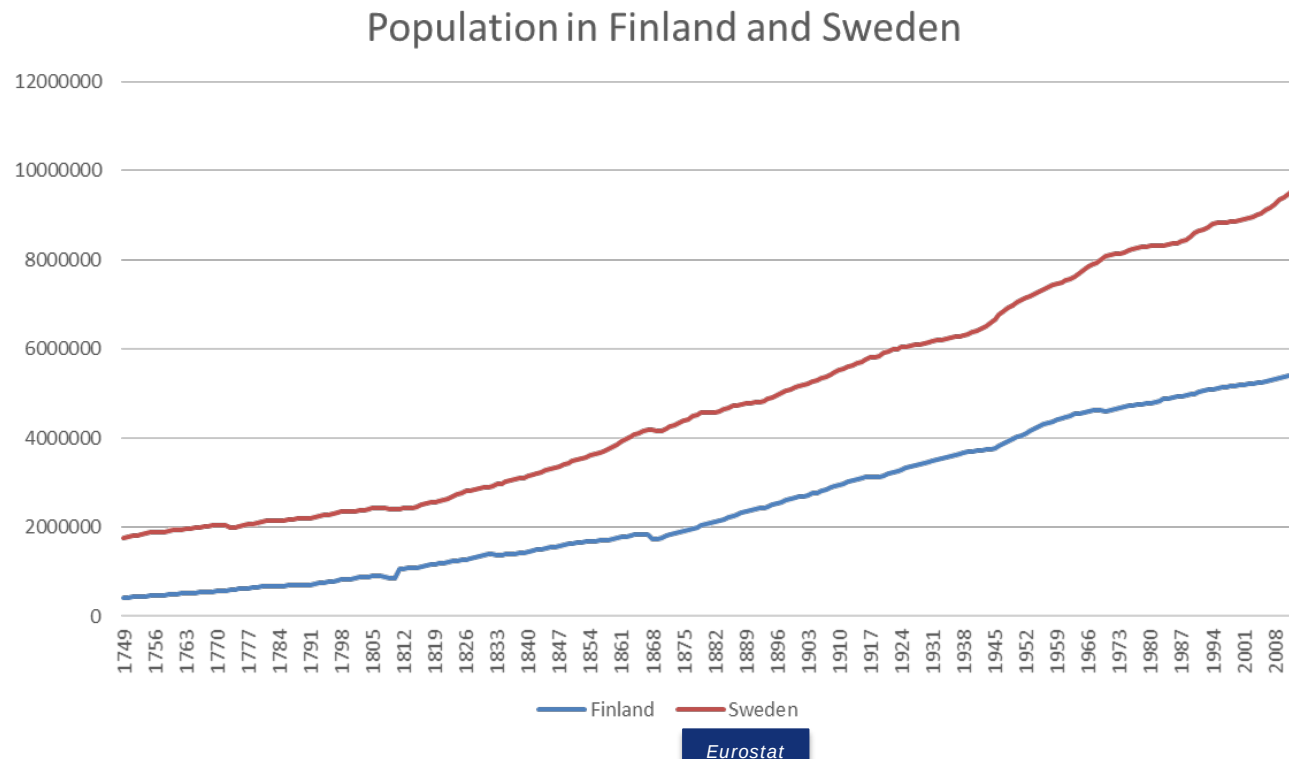
- In Finland, population censuses and population registration have been closely tied together for centuries.
- In 16th century in Sweden(-Finland) the first records on population were compiled for purpose of recruitment and taxation
- In 17th century parishes were obliged to keep records on births, deaths and marriages and also migrations between parishes
- In 17th century the idea of continuously production of statistics on population

History of Finland's Population Censuses (2)

- In 1748 the Statistical Office of Sweden(-Finland) was established
- In 1749 conducted the first ever population census in the country
- Parishes and register office records were consulted to collect information: births, deaths and marriages by sex, number of population, later also social class

History of Finland's Population Censuses (3)

- All information was not collected every year, but anyway since that the number of population have been available yearly in Finland (and of course also in Sweden)



History of Finland's Population Censuses (4)

- Administrative sources (Church Book / records) were used already then for statistical purposes!
- 1865 Central Statistical Office was founded in Finland.
- Years 1870-1930 the traditional censuses in the biggest cities
- 1938 Census Law for the conducting the first census in the whole country 1940
- Cancelled because of the war
- Traditional censuses 1950, 1960, 1970, 1975, 1980 and 1985
- First time register were used in 1970 Census
- Totally the register-based Census from 1990

Census data by the type of the data collection method 1950 -2010

q = census questionnaire	qr = for non-respondents from registers										
R = registers	Rq = register data supplemented with questionnaires										
- = not included in Census											
Data Item	Year									core 2010	non core 2010
	1950	1960	1970	1975	1980	1985	1990	2000	2010		
Demographic data											
Age (Date of Birth)	q	q	q	R	R	R	R	R	R	x	
Sex	q	q	q	R	R	R	R	R	R	x	
Marital Status	q	q	q	R	R	R	R	R	R	x	
Mother Tongue	q	q	q	R	R	R	R	R	R		x
Citizenship	q	q	-	R	R	R	R	R	R	x	
Country/place of birth										x	
Religion	q	q	R	-	R	R	R	R	R		x
Usual Place of Residence	q	q	q	q	q	R	R	R	R	x	9

Census data by the type of the data collection method 1950 -2010

Data Item	Year										core 2010	non core 2010
	1950	1960	1970	1975	1980	1985	1990	2000	2010			
Economic Data												
Main Type of Activity	q	q	q	q	q	qr	R	R	R	x		
Status in Employment	q	q	q	q	q	qr	R	R	R	x		
Location of Workplace	q	q	q	q	q	qr	R	R	R	x		
Industry	q	q	q	q	q	qr	R	R	R	x		
Occupation	q	q	q	q	q	qr	R	Rq	Rq	x		
Socio-economic Group	q	q	q	q	q	qr	Rq	Rq	Rq			x
Income data	-	-	R	R	R	R	R	R	R			x
Education Data												
Completed Education	q	q	q	R	R	R	R	R	R	x		
School Attendance							R	R	R			x

Census data by the type of the data collection method 1950 -2010

Data Item	Year									core 2010	non core 2010
	1950	1960	1970	1975	1980	1985	1990	2000	2010		
Household and Family Data											
Type of Household	q	q	q	q	R	R	R	R	R	x	
Size of Household	q	q	q	q	R	R	R	R	R	x	
Type of Family	q	q	q	q	R	R	R	R	R	x	
Size of Family	q	q	q	q	R	R	R	R	R	x	
Dwelling Data											
Number of Rooms	q	q	q	q	R	R	R	R	R	x	
Kitchen	q	q	q	q	R	R	R	R	R		x
Water, sewage, toilet	q	q	q	q	R	R	R	R	R	x	
Heating System	q	q	q	q	R	R	R	R	R	x	
Tenure Status	q	q	q	q	R	R	R	R	R	x	

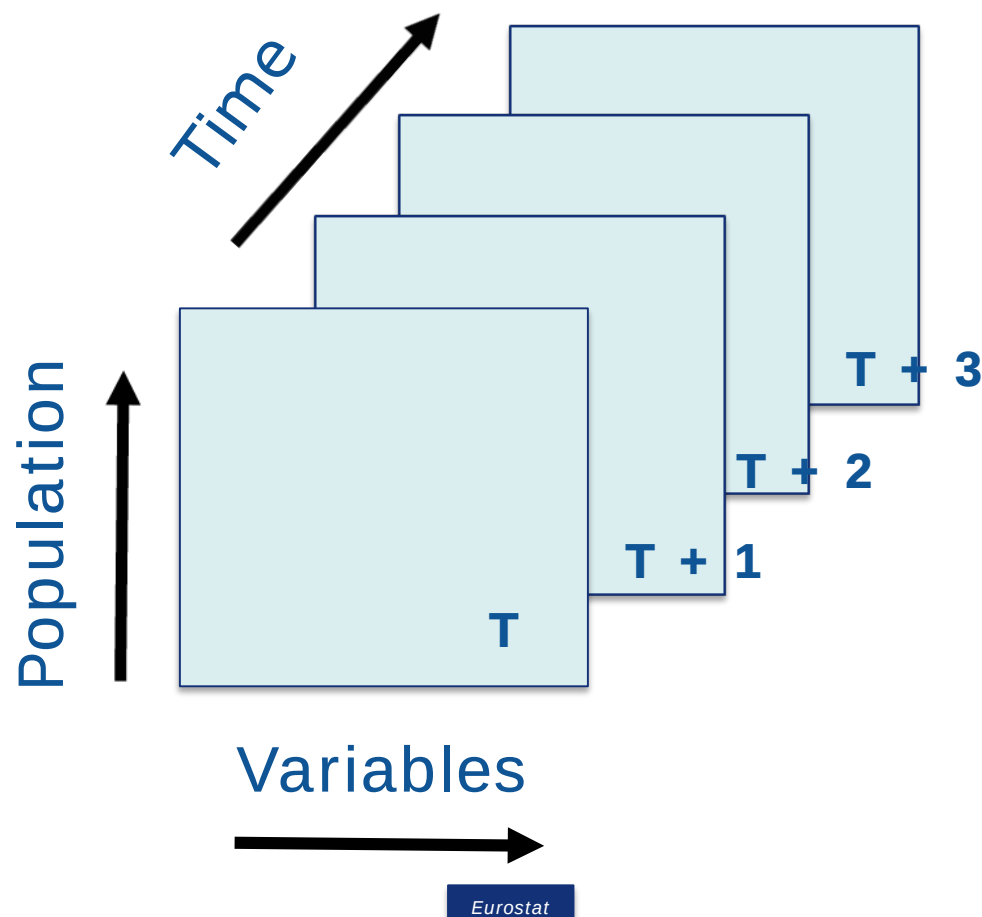
Census data by the type of the data collection method 1950 -2010

Data Item	Year										core 2010	non core 2010
	1950	1960	1970	1975	1980	1985	1990	2000	2010			
Building data												
Type of Building	q	q	q	q	q	R	R	R	R	x		
Year of Construction	q	q	q	-	q	R	R	R	R			
Construction Material	q	q	q	-	q	R	R	R	R			
Main Use of Building	q	q	q	-	q	R	R	R	R			
Number of Dwelling Units	q	q	q	-	q	R	R	R	R			
Capacity (m3)	-	-	q	-	-	-	-	-	-			
Heating System	-	q	q	-	q	R	R	R	R			
Number of Storeys	q	q	q	-	q	R	R	R	R			
Coordinates of Building	-	q	R	R	R	R	R	R	R			

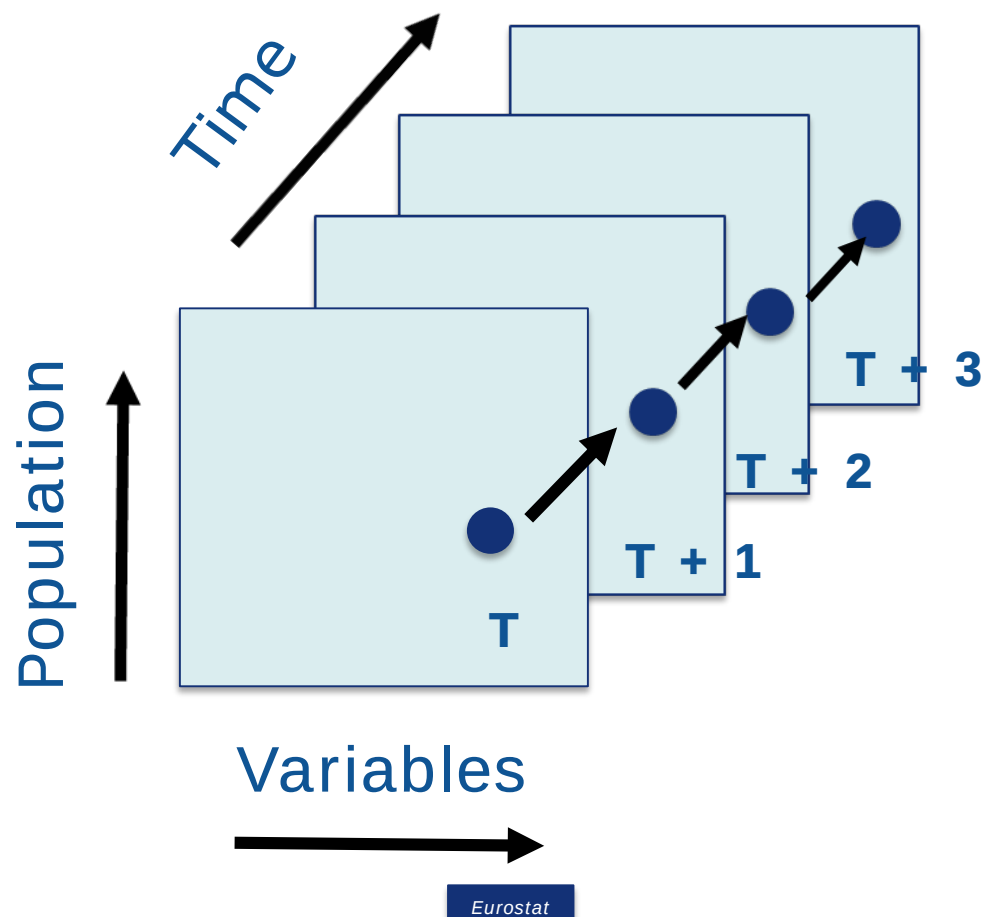
New possibilities for research and statistics production

- Statistics usually provide cross-sectional information on a variable at a given point in time, such as population number or the number of people in gainful employment.
- On this basis we can see to what extent these figures have changed.
- The register system offers the added advantage of allowing us to identify the individuals behind these changes: who has got a job, who has completed a degree.
- Changes can be monitored by linking unit data from consecutive years.

New possibilities- follow up



New possibilities – follow up



New possibilities regional statistics (1)

- Traditionally, the most important regional unit in statistics has been the administrative area.
- However, administration is dynamic and keeps changing
=> may occur difficulties to keep up these changes
- The building-based code system with its coordinates has provided a solid foundation for reliable and flexible statistical areas.
- Despite major changes in administrative areas, it is still possible to produce time series for different regions.
- The adoption of map coordinates for buildings has also make it possible to define more flexible statistical areas.

New possibilities regional statistics (2)

- Calculation of accessibility (workplaces, services)
 - Distance to work, school, voting place
- Flow statistics
 - employment flows
 - student flows
- Longitudinal researches
 - Data from 1970, 1975, 1980 and 1985 Censuses and annually data from year 1987 in the Census Data Warehouse
 - Data over 7 million persons (Population of Finland 5,5 million in 2016)

Examples

- Commuting distance and time for employed
- Flow statistics
- Cohorts
- Combining historical data to persons

Commuting distance and time for employed

- Commuting Distance
 - General annual update for the Census Statistics Data Warehouse
- Commuting time
 - Enriching with traffic sensor data
- Digiroad, National Road Database of Finnish Transport Agency
 - Accurate data on location of all roads and streets in Finland

Commuting distance and time for employed

- Census Data Warehouse
 - Dwelling coordinates and work place coordinates
 - Coordinate coverage
 - for the place of living of population around 99 %
 - for the workplace of all employed around 91 %
- Traffic sensor data of FTA
 - Currently 437 stations (vehicle detection loops) giving information for speed, direction, length and class of a passing vehicle.
 - Open data services available as well (Digitraffic)

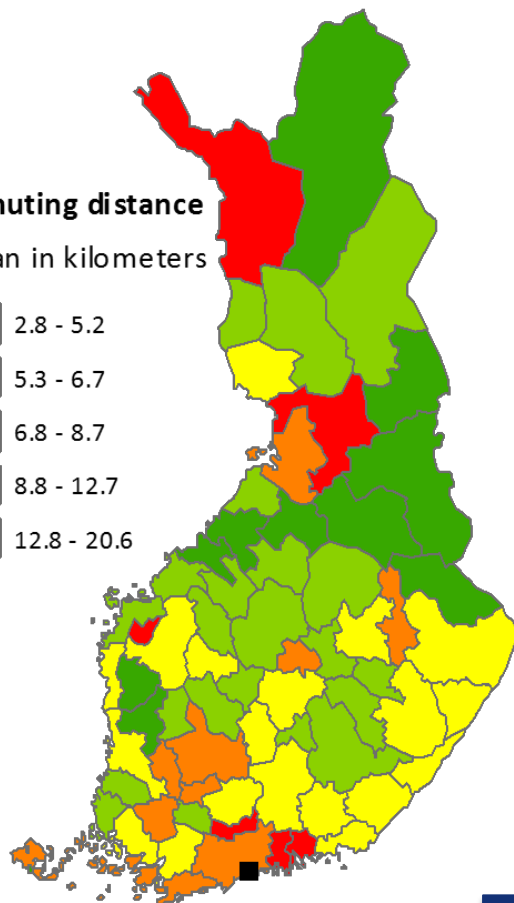
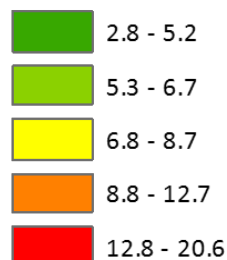
Commuting distance and time for employed

- Read more: [Pasi Piela. Commuting time for every employed: combining traffic sensors and many other data sources for population statistics](#) in European forum for geography and statistics, Krakow 2014

Commuting distance and time for employed in subregions (LAU1) in 2012

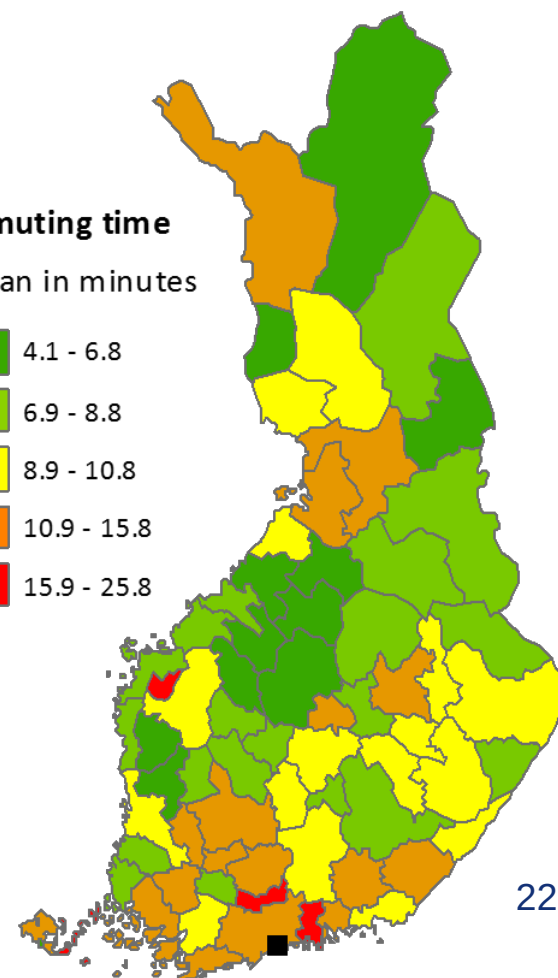
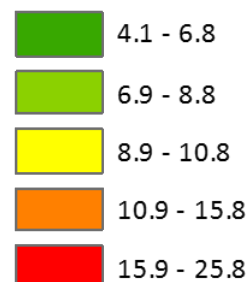
Commuting distance

Median in kilometers

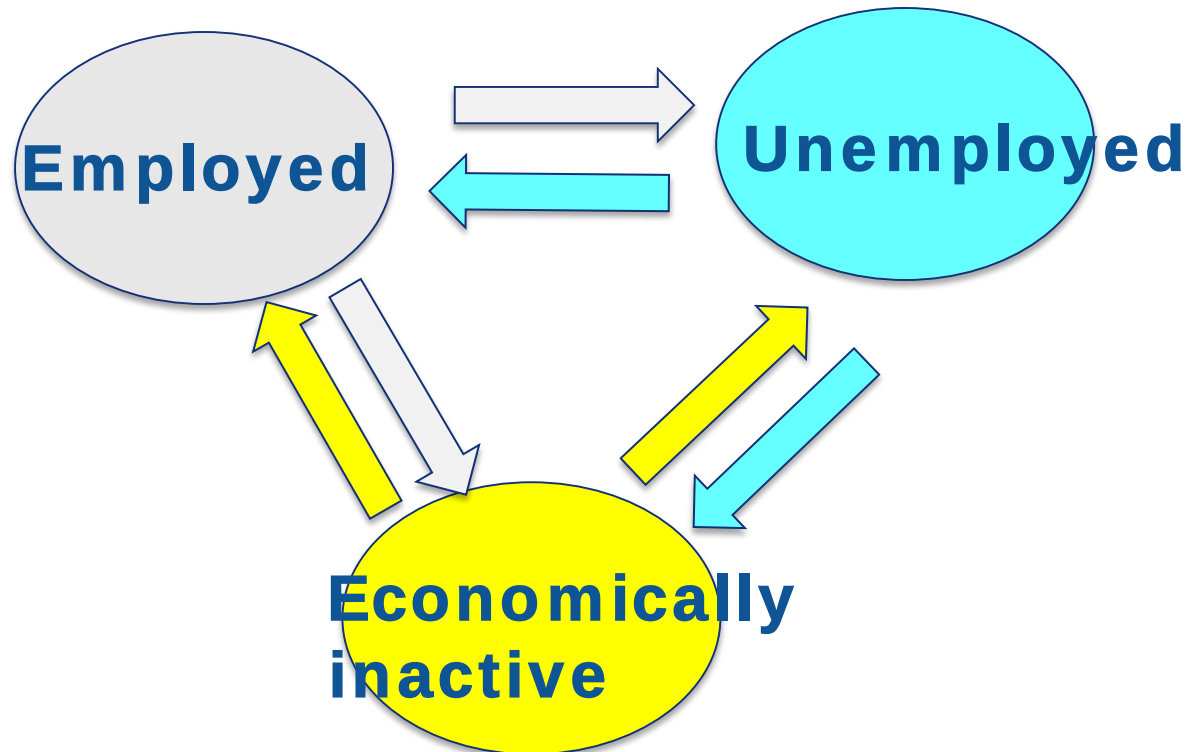


Commuting time

Median in minutes



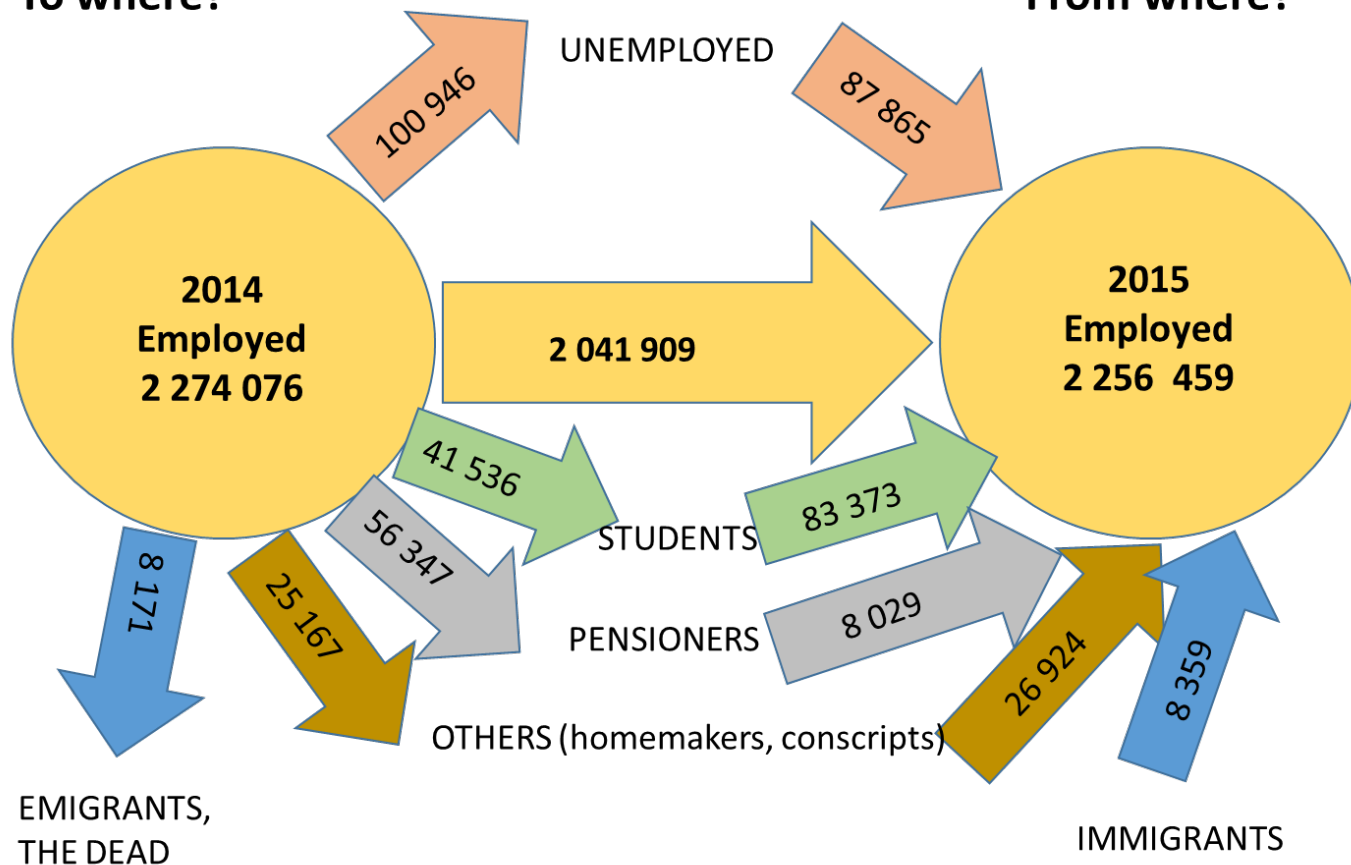
Flows between different activity groups



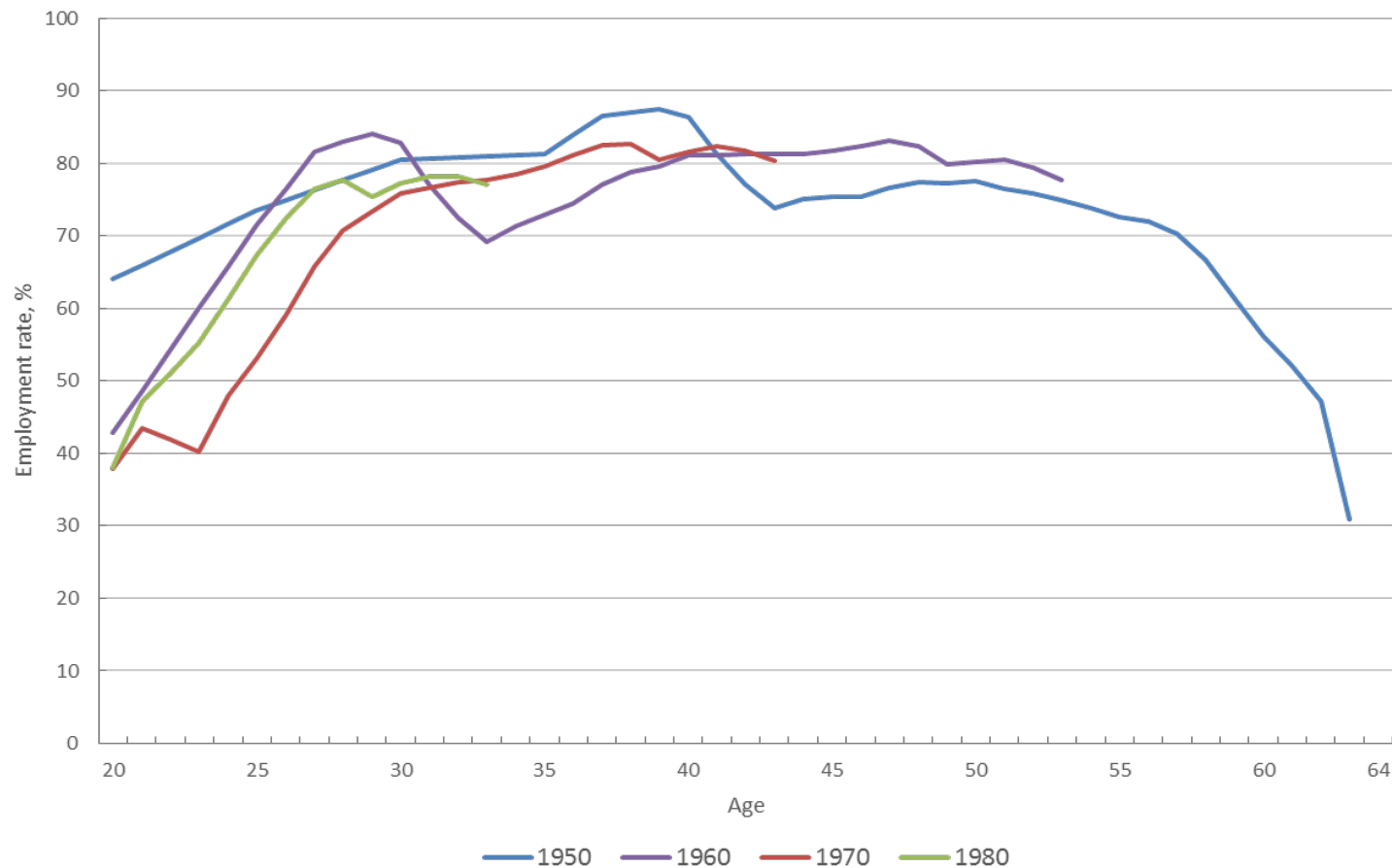
Flows between different activity groups: Employed 2014-2015

To where?

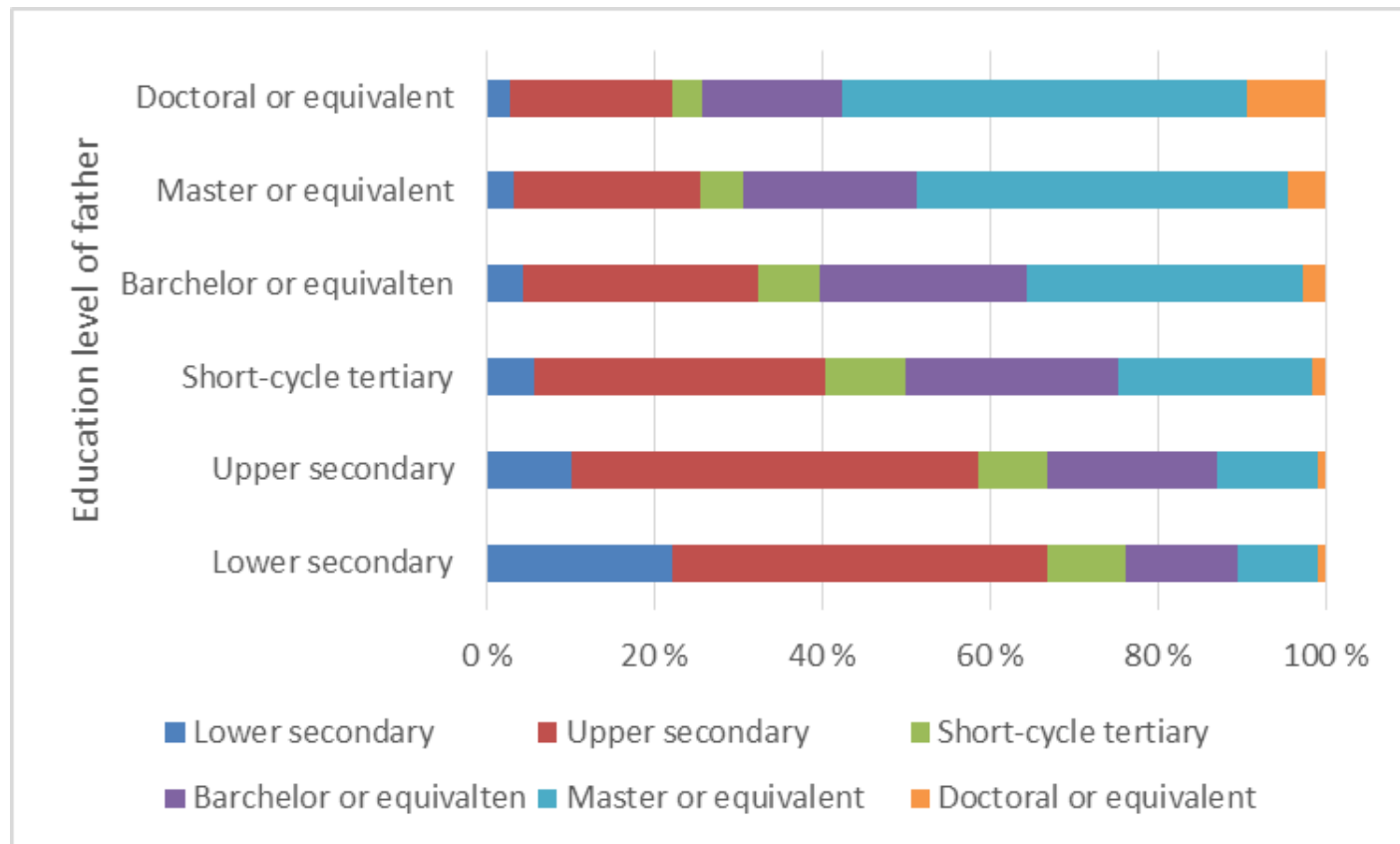
From where?



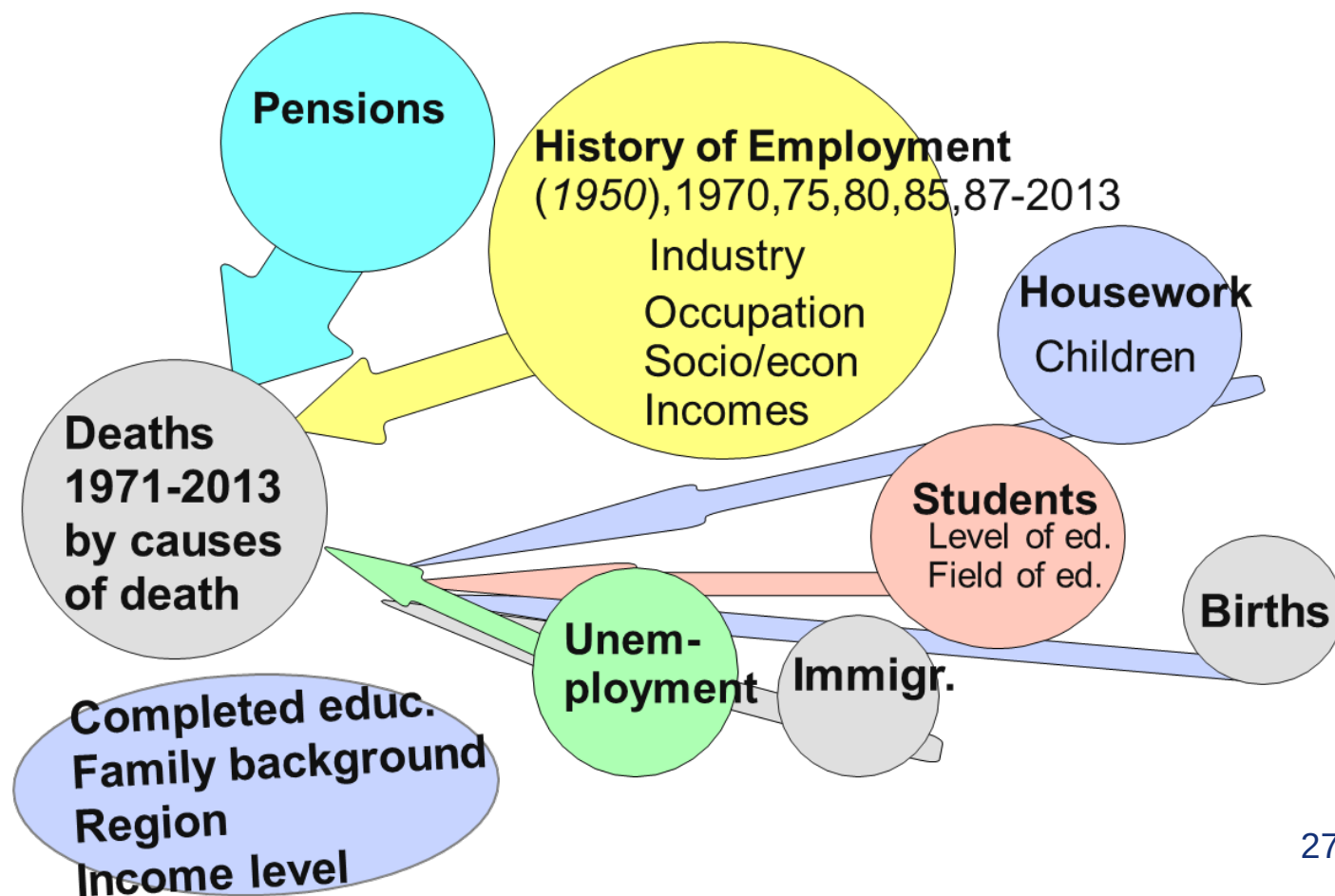
Employment rate by birth cohorts



Heritability of the education: effect of the father's education level on the children's education level, 2015



Deaths and the history of different activities





**Thank you for your
attention!**