

Use of Registers and Administrative Data Sources for Statistical Purposes

Best Practices of Statistics Finland



Use of Register and Administrative Data Sources for Statistical Purposes

Best Practices of Statistics Finland

Tiedustelut – Förfrågningar – Inquiries:

*Pekka Myrskylä
+358 9 17 341*

© 2004 Tilastokeskus – Statistikcentralen – Statistics Finland

*Tietoja lainattaessa lähteenä on mainittava Tilastokeskus.
Uppgifterna får lånas med uppgivande av Statistikcentralen som källa.
Quoting is encouraged provided Statistics Finland is acknowledged as the source.*

*ISSN 0355-2063
= Käsikirjoja
ISBN 952-467-333-9*

Valopaino, Helsinki 2004

Preface

Statistics Finland, like its sister agencies in the other Nordic countries, has for decades been working to develop the use of administrative registers for purposes of statistics production. In Finland the earliest uses of administrative data can be traced back to census collections in the eighteenth century, while modern statistical uses began in connection with the 1970 population and housing census. After two decades of systematic expansion, the 1990 census was collected exclusively from registers, without any direct data collection from the population. At the same time as census collection methods were developed, the statistical use of administrative data sources first expanded on the side of demographic and social statistics and eventually to business statistics. Today, in keeping with the requirements of the Statistics Act, registers and administrative records are the main source of data collection for Statistics Finland.

Population censuses represent a major undertaking for every statistical authority, both financially and operationally. For this reason, statistics offices in different parts of the world are seeking new, more cost-effective ways of producing the data traditionally provided by censuses to their users. Nowadays the need for census data, and regional data in particular, is more frequent than once every ten years. Many countries have therefore started to conduct so-called interim censuses. However in many cases even a five-year interval is not frequent enough, for instance for regional planning needs. From a budgetary perspective there is also the problem of costs peaking every five or ten years, which very much complicates the task of fund allocation.

Also in Finland, the above factors were the main reason why after the 1980 census a systematic search began, at the behest of the Ministry of Finance, for more economical approaches to producing census data. The obvious starting-point for a more cost-effective census approach in Finland was to use administrative data records.

Finland is among the world pioneers in the statistical use of administrative data sources, and Finnish experts therefore are very much in demand on the international seminar, training and conference circuit, particularly on the subject of register-based census systems. There exists by now quite an abundance of material on the use of register sources that might have wider interest and application. As various countries are continuing their efforts to facilitate the use of administrative sources in statistics production, especially with a view to the 2010 population and housing census, we thought it might be useful to compile our best practices into a manual for international distribution. Because of the contributors' background the main focus of this manual is on the register use in demographic and social statistics, but a brief discussion is also included on the use of administrative records in compiling business statistics.

We hope that this manual proves useful to experts in countries that are looking to promote the use of administrative records for statistical purposes. All comments about this manual are most welcome. We are well aware of the particular circumstances that make register-based statistics feasible in the Nordic

countries and that it may be assumed will never be reached in many other countries. However, we believe that many countries could make more extensive use of the administrative data that are produced in various administrative contexts – and that nowadays are usually available in electronic format as well.

This volume has been edited by Pekka Myrskylä, Senior Adviser at Statistics Finland, who has extensive experience in various aspects of population and housing census. Many others have also contributed with descriptions of the use of administrative data records and useful comments. I would like to thank all who have been involved in compiling and editing this material. Special thanks go to Pekka Ruotsalainen, Planning Officer (Chapter 7.2.1.), Salme Kiiski, Head of Statistics (Chapter 7.2.2.), Tuula Viitaharju, Head of Statistics (Chapter 7.2.3.), Kari Molnar, Director, Prices and Wages (Chapter 7.2.4.), and other Statistics Finland's employees mentioned in the list of references as well as to Director-General Heli Jeskanen-Sundström and Statistics Finland's other directors who have given their valuable comments to the texts. Thanks are also due to Translators David Kivinen and Mia Kilpiö.

Helsinki, September 2004

Riitta Harala
Director, Population Statistics

Contents

Preface	3
1. Register-based statistics production:	
General preconditions.	7
1.1. Legal basis	7
1.2. Public approval	7
1.3. Unified identification code systems	8
1.4. Comprehensive and reliable register systems developed for administrative needs	8
1.5. Cooperation among administrative authorities	9
2. Overviewing registers	10
2.1. What is a register?	10
2.2. Registers – cross-sectional data files	10
2.3. What are basic registers?	11
2.4. Updating register data	12
3. Major register sources used in statistics production at statistics Finland ...	15
4. Uses of administrative data.	16
4.1. Totally register-based statistics and the combined use of survey and register data	16
4.1.1. Direct use of register data	16
4.1.2. Register estimation	16
4.1.3. Survey studies, sampling frames, sampling updates, additional information	17
4.2. Register-based structural statistics, change statistics, combination statistics	19
5. Cooperation with administrative authorities	21
6. Register-based population census system	22
6.1. The development of a register-based population census system	22
6.1.1. A brief history of population censuses	22
6.1.2. Towards a totally register-based census method	23
6.1.3. Register-based population census system	25
6.2. The pros and cons of register-based population census statistics production	26
6.3. Reliability, comparability and timeliness of register-based census data ..	30
6.3.1. Reliability	30
6.3.2. Comparability with international recommendations	32
6.3.3. Timeliness	34
7. Register-based statistics systems	35
7.1. Statistics under the population census system	35
7.1.1. Register data on population structure and changes	35
7.1.2. Register-based building and dwelling data	36
7.1.3. Regional employment statistics	37
7.1.4. Occupational data	40
7.1.5. GIS and register-based census	41

7.2. Use of register data in some other statistics	43
7.2.1. Income distribution statistics	44
7.2.2. Labour Force Survey	44
7.2.3. Business Register	45
7.2.4. Register-based business statistics	47
8. Statistical legislation and data protection	49
8.1. General principles	49
8.2. Releasing data for research purposes	50
9. Register-based statistical systems: new opportunities for research and statistics production	52
9.1. Longitudinal data files	52
9.1.1. Longitudinal data file on population censuses 1970 – 2000	52
9.1.2. Longitudinal data file on employment statistics 1987 – 2002	52
9.1.3. Other longitudinal data files	53
9.2. Flow statistics	53
References	56
Appendix 1. How certain ID numbers and codes are created	58
Appendix 2. Use of register sources in Finnish population and housing censuses 1950–2000	61
Appendix 3. Data sources for regional employment statistics	62
Appendix 4. Data sources on occupational titles	64
Appendix 5. Business Register and administrative records	65

1. Register-based statistics production: general preconditions

Finland, as indeed all the Nordic countries, relies more extensively on administrative data sources for purposes of statistics production than many other countries in the world. It has long-standing experience in the area of register-based statistics production, which over the years has expanded to cover first the production of social statistics and more recently business statistics as well.

In purely volume terms, the vast majority or 96 per cent of Statistics Finland's current data reserves come from administrative registers, while only 4 per cent are collected directly from business enterprises, private individuals and other sources. Having said that, the role and significance of the data collected directly from data providers in agency statistics production clearly outstrips that four per cent. It is also important to note that administrative data records never can fully replace direct data collection, but these two methods are complementary to each other.

In the light of the Finnish experience there are certain key preconditions that facilitate the extensive use of register sources in statistics production. In Finland these have included the following:

1.1. Legal basis

Legislation provides a key foundation for the use of administrative data sources for statistical purposes. In Finland, national legislation reflects the broadly held view that it makes good sense to take advantage of existing administrative data sources rather than re-collect them for statistical purposes. The Finnish Statistics Act (2004) is based on the principle that whenever possible, statistics shall be compiled using administrative records. The Statistics Act also gives powers to Statistics Finland to access administrative data on unit level with identification data and to link them for statistical purposes. Furthermore, the Statistics Act provides a detailed definition of data protection (questions of data protection are discussed more closely under Chapter 8).

1.2. Public approval

It is also extremely important that the general public appreciates and understands the benefits of using register sources for statistical purposes and that there is broad public approval of the use of these administrative data for purposes of statistics production.

In Finland, this is indeed traditionally the case: people believe that government is rational and they also have strong faith in Statistics Finland. Open dis-

cussion and debate, explaining the rationale and benefits of register use has always been considered a key principle. It is also important that the national register legislation is up-to-date and the activity of register authorities is open and transparent.

Statistics Finland has succeeded in making a good case for the statistical use of existing administrative data by reference to cost efficiency, a reduced response burden on the population and the avoidance of overlap in data collection. No serious questions have been raised in the press with respect to data protection, for instance, even in connection with register-based population censuses. However it is important that the statistical agency always remains on guard in this respect because it is very easy to lose the confidence of the general public but a major effort indeed to rebuild that confidence.

1.3. Unified identification code systems

One major factor that facilitates the statistical use of administrative data records is the application of unified identification systems across different sources. In the absence of such unified systems it is extremely difficult and laborious, if not impossible, to link different registers, which is absolutely central to register-based statistics production. A bare minimum requirement is to have a unified identification system for key basic registers. The definition of basic registers is discussed in more detail in Chapter 2.1.

Finland adopted a unified system of personal identity codes in 1963. These codes are currently used in virtually all registers that are consulted for purposes of statistics production. Similar, almost equally unified identification systems are in use for business enterprises, buildings and dwellings: these are maintained in basic registers which contain unit data. The various identification systems are described in closer detail in Appendix 1.

In some cases it may be possible to link different registers even without unified identification codes, but this is certainly more laborious and time consuming. For example, the register data on private sector employment that is used in the Finnish register-based census system do not include enterprise codes: therefore each year the relevant codes have to be entered using alphabetical identification based on the name of the company and Statistics Finland's Business Register, for instance, and part of this work has to be done manually. This requires considerable labour input when compared to system maintenance overall.

1.4. Comprehensive and reliable register systems developed for administrative needs

In Finland the compilation of administrative data registers has initiated from the needs of the functioning of society and development of administration. It has

also been closely tied in with the development of social security and tax systems. Both of these are state-level systems and therefore it has also been necessary to have state-level registers. In recent years the payment of basic benefits has also been increasingly concentrated to the same authorities, which has further contributed to the concentration of register sources. Most social security benefits in Finland are taxable income, which is why the tax authorities need to be informed as well.

Both the people and the authorities in Finland have always shown high confidence in the accuracy of register sources, which after all are the basis for many individual basic rights and activities. For instance, the official domicile of every individual resident in the country is determined on the basis of register data. Likewise, an extract from the population register serves as a basic document that is needed when applying for a passport or an office, getting married or divorced, or when a funeral is held or an estate is distributed. Another factor that speaks for the reliability of register data for statistical purposes is that it is in the interest of each individual to make sure that all the data within register systems are indeed accurate.

1.5. Cooperation among administrative authorities

A concerted effort towards register-based statistics production also requires a firm and explicit commitment from the highest possible level as well as close collaboration among the relevant authorities.

In Finland, for instance, the Ministry of Finance made clear its position to Statistics Finland in connection with the 1980 census that this should be the last census based on extensive questionnaires, and that the agency should invest its efforts in developing a register-based census system. This ministerial intervention helped to enhance Statistics Finland's bargaining position vis-à-vis other administrative authorities.

Today collaboration is close between register authorities and Statistics Finland. It takes place not only in the scope of the register pool, the cooperation forum for national register-keepers, but also in the form of regular bilateral project and meeting work (for more details, see Chapter 5).

2. *Overviewing registers*

2.1. *What is a register?*

A **data register** is a **unit-level (total) data set that comprises a certain complete target population. Usually maintained in machine-readable format, data registers are regularly updated to keep track of any changes in the data describing the units and their attributes.**

The target population of a register may comprise the population resident in the country, housing dwellings, motor vehicles on the road, unemployed job applicants, taxable income earners, pensioners, etc. Ideally, for statistical purposes, the register should be as comprehensive and accurate as possible, but even incomplete registers can be used for compiling statistics.

Registers that are useful for statistical purposes are usually available in **machine-readable format**, which means that they can be processed using computers. Card index registers, such as certain company employee registers, association membership lists, etc., rarely have statistical application.

A key requirement with respect to data use and maintenance is that each unit in the register can always be **unambiguously identified**. This is easiest if the register has its own system of **identification codes** (e.g. personal identity code, real estate identifier, business identity code).

Identification is possible even without identification codes if sufficient information is available on each unit, such as name, date of birth, place of birth, address, etc. In this case, though, identification is more difficult and involves greater uncertainty – although in recent years even these techniques have developed rapidly.

2.2. *Registers – cross-sectional data files*

Other **machine-readable data files** may often be called ‘registers’ as well, even though de facto they simply are data sets collected to describe a certain cross-sectional situation, with no intention to update them as the situation changes. One example is provided by the so-called **Joint Selection Register**, which lists all applicants and entrants to vocational schools or colleges and upper secondary schools in a given year. The following year most names in this ‘register’ will be entirely different, which means it is in fact simply a machine-readable **data file**, even though these data are used in the same way as register sources to define and describe the student population in Finland. The same applies to many of Statistics Finland’s annual data sets, although they may be called registers.

The **Taxation Register** may be regarded as an intermediate form between a register and an electronic data file: this is a record of all persons who are subject to tax each year, listing their income and assets as well as their taxable benefits.

The personal data in this register are updated annually by reference to the Central Population Register in order to record deaths, removals, changes in family information, etc., but the taxation data proper are collected each year from the income earners themselves, their employers and pension institution – rather than being updated as such. However, data on assets are maintained and used for purposes determining any property tax and taxes on profit from sales. This information does not need to be separately reported each year unless it has changed.

There are some country differences in uses of the concept of register. In Sweden, for example, various statistical data sets are often called registers or statistical registers. Likewise, the process of storing or recording data is called registration. In the English language, on the other hand, register data are often described simply as administrative data.

2.3. What are basic registers?

There are some key registers in Finland that are known as **basic registers**. Regulated by Acts and Decrees, basic registers include information that is central to the function of society. They are highly reliable, comprehensive and have multiple uses. Part of the data are public, part confidential. The information contained in basic registers is carefully protected. If released, this must not contravene the requirements of privacy protection. All citizens have the right to check their own personal information and to decline the use of that information for direct marketing purposes, for example, but not for purposes of compiling statistics.

Basic registers include the

- population information system, which also includes data on buildings and dwellings;
- real estate information system; and
- business information system.

The population information system contains basic data on all Finnish citizens and foreign nationals permanently resident in Finland, as well as data on buildings, ongoing building projects and dwellings. The information system is maintained by the Population Register Centre and local register offices.

The real estate information system covers real properties and parcels of land. The real estate section includes details on total surface areas as well as on any rights and easements attached to properties. These data are maintained by the National Land Survey and municipalities. The registration section includes information on the owners of real properties and any mortgages and usufructs on the property. These data are maintained in the Register of Titles and Mortgages by the Ministry of Justice and district courts.

The business information system includes key information on business companies and organisations. It comprises the Trade Register, Register of Foundations and Register of Associations, which are all maintained by the National

Board of Patents and Registration; the business information system maintained by the Finnish Tax Administration; as well as the Statistics Finland Business Register.

This **basic register system** is unique in the whole world. It comprises constantly updated information on all persons resident in Finland as well as on Finland's building and dwelling stock and Finnish enterprises and entrepreneurs and self-employed persons. This information system is central to many register-based statistical systems, such as the population census.

Other major register systems include

taxation registers;
employment registers;
pension registers; and
the Register of Job Applicants.

Statistics Finland has also itself created various registers for purposes of statistics production. Most importantly, these include the **Register of Completed Education and Degrees**, which was formed in connection with the 1970 census, and the **Business Register** that was created in the 1980s. Furthermore, Statistics Finland was involved in collecting data for the **Register of Buildings and Dwellings** when it was set up in the Population Register Centre in connection with the 1980 census. A few years ago Statistics Finland established a **Student Register**, which covers all students at post-primary educational institutions. Statistics Finland also maintains regional databases that are very similar to registers.

The Business Register provides information on such details as branch of industry and legal form. This information is used in the population census system to establish the branch of industry for the employed population; more recently it has also been used in labour force surveys. This guarantees that these three systems will present fully harmonised information on branch of industry: regardless of the statistical system, all people working at the same establishment are always ascribed to the same branch of industry.

2.4. Updating register data

The basic idea of a register is that it is regularly updated in order to cover all relevant units and to keep the data describing these units fully up-to-date. New units entering the target population are therefore regularly added to the register (e.g. newborns and immigrants); while those exiting the target population are removed (or moved to history information) from the register (e.g. persons deceased or emigrants). The data describing the units may also change and therefore have to be updated as well. Changes may be recorded in such details as marital status, educational level, occupation, or number of children. Likewise, the floor area of a building may change, the owner of a motor vehicle may change, etc.

Example

Population information system (the Central Population Register) covers all persons who are permanently resident in the country, i.e. everyone with a permanent domicile in Finland and all buildings and dwellings. The population information system is maintained by the Population Register Centre and local register offices. Register data are used for purposes of information management on the scale of society as a whole: in public administration, in research, in organising elections, in compiling statistics. Some of the information in the Central Population Register is in the public domain and can therefore be made available for marketing purposes, for instance. However, citizens have the right to decline the release of their personal data for such purposes.

The Population Information System comprises basic identification data for natural persons, buildings and dwellings. The data registered for natural persons include name and personal identity code, address, nationality and mother tongue, marital status, and birth and death information. For buildings and dwellings, data items entered in the register include property identifier and location, owner, surface area, equipment and grid connections, usage of building, and year of construction.

Responsibility for updating the population information system rests mainly with the authorities, although change-of-address information is to be provided by the individuals concerned. Information on births and deaths is provided by hospitals; on marriages and namings by parishes and local register offices; on divorces and adoptions by courts of law; on changes of name by state provincial offices; on nationality by the Directorate of Immigration; on changes in dwelling and building data by local building authorities; and on changes in real estate information by land survey offices. All that remains for individual citizens is to notify of changes of address, which must be done within seven days of moving through one of the country's 37 local register offices. Personal information on people who have emigrated or deceased is also kept in the register: emigrants are moved to the category of absent population, and date of death is entered for the deceased. Personal data are always updated on the basis of the personal identity code, data on buildings and dwellings are accordingly updated on the basis of the building and dwelling code.

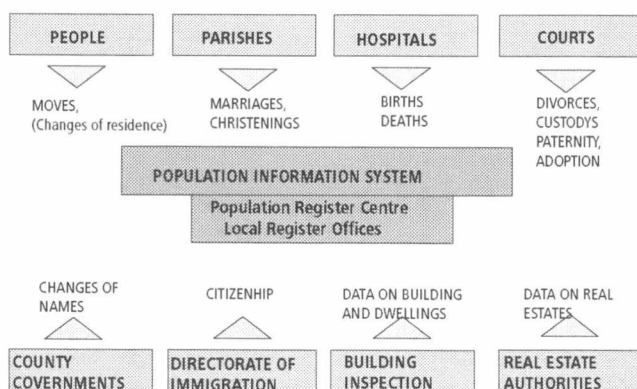


Figure 1.
Procedure for updating
the population information system

A key principle of keeping registers is that the relevant data are collected when they arise and on this one occasion only. When a newborn child is entered into a register, the registrar will record the child's place of birth, mother tongue, nationality, the mother's and father's personal identity codes, and mother's current residence. Change-of-residence information is entered in the Population Information System from which the new address information will then be transferred to other administrative files on the basis of the personal identity code. Similarly, all attribute data on newly built dwellings are entered into the relevant register and updated only when necessary. Information describing the country's entire housing stock will never be re-collected all over again.

Indeed the biggest advantage of good register system is that the need for processing is restricted to those units and those unit attribute data that have **changed**. Units and attribute data that remain unchanged are not processed at all. The more stable the data contained in a register, the less work involved in register maintenance.

3. *Major register sources used in statistics production at Statistics Finland*

The following lists the most important registers used by Statistics Finland in its own statistics production. These include:

- The Population Information System, containing information on the population, buildings, dwellings, business premises and summer cottages (maintained by the)
- The Customer Register of Taxation (from the customer database), Payment Control Register (contains monthly data on value added tax and employer contributions), taxation registers (e.g. Business Taxation Register), income and property data, employers' annual control data, the Trade Registration Code Register (which is being transferred to the business information system jointly maintained by the Finnish Tax Administration and the)
- The Trade Register (maintained by the National Board of Patents and Registration of Finland)
- Employment data in the employment pension systems (maintained by the ,, and other minor information producers)
- The Register of Job Applicants and Labour Market Training (maintained by the)
- People on old age, disability and unemployment pension, and information on housing allowances and illnesses (maintained by the Social Insurance Institution)
- The Register of Income Support Recipients (maintained by the)
- The Conscript Register (maintained by the)
- Student registers (maintained by and others)
- The Register of Completed Education and Degrees (maintained by Statistics Finland)
- The Vehicle Register (maintained by the Finnish Vehicle Administration)
- The Business Register (including central and local government units) (maintained by Statistics Finland).

4. *Uses of administrative data*

4.1. *Totally register-based statistics and the combined use of survey and register data*

There are many different ways in which administrative data can be used for purposes of statistics production. These include

1. Direct use of register data
2. Register estimation
3. Combined use of survey and register data
 - additional information from registers
 - use as sampling frame
 - non-response control
 - imputation
 - determining the structure of non-response

Statistics Finland uses all of these methods. Statistics are often produced using a combination of different methods.

4.1.1. *Direct use of register data*

Some statistics are produced using register sources only and without any linking of data or drawing of inferences. Examples include statistics on the population's sex and age structure, statistics on population change, various business statistics, income and wealth statistics, education statistics, crime statistics, and buildings and dwellings statistics. In these cases the use of register data is technically a rather simple and straightforward matter. Logical checks will be made in this connection to remove any obvious register errors: the surface area of a building may have been entered incorrectly, a decimal point may be missing from an income figure, a building may have electric heating but it is not connected to the electricity grid.

Another method is to produce new data on the basis of information contained in one register. For example, there is no specific register item on families, but these are formed on the basis of people living together using certain rules. A household-dwelling unit, then, consists of all people living in the same dwelling even if they are not members of the same family.

4.1.2. *Register estimation*

Often the statistical variable required cannot be obtained from one single register source, but it is necessary to consult several registers. The primary method used in the Finnish register-based census system and its sub-system of regional

employment statistics is that of register estimation. This involves using questionnaire data to construct a model or an optimal group of decision rules with which it is possible to derive from one or more items of administrative data a maximally good parameter value that conforms to statistical concepts to represent each unit of the population.

In other words, this method makes use of a number of register-based data sets at the same time. Deducing a person's main economic activity, for example, requires consulting more than 20 sets of register data. It follows that each register does not necessarily have to be fully comprehensive, but they can be used to complement one another. Information on the same person may thus be available from different data sets, allowing for the selection of the most reliable source from the point of view of the material as a whole.

This method is not particularly vulnerable to changes in register data, since data on individual phenomena can usually be derived from a number of sources. This means that personal employment data, for instance, are available via employment pension systems, taxation registers and reports submitted by the individuals themselves or their employers. If an employer has neglected to inform the taxation authorities, for example, the data will still be available through the pension system or the individual's own tax returns.

4.1.3. Survey studies, sampling frames, sampling updates, additional information

In Finland registers are used generally as **sampling frames**. The frame most commonly used in surveys on persons, households and families is the Central Population Register and smaller databases derived from it. Registers and databases contain the most important demographic data for units, such as age, sex, marital status, place of residence, number of children, nationality, etc., which can be utilised in defining the population and selecting the sample. If the sample unit is a household-dwelling unit, all persons living in the same dwelling can be selected to the sample. Finland also has comprehensive registers on buildings and dwellings, so the survey can also be directed to the owners of the building or the persons living in the building or dwelling. Samples may also be drawn from the Register of Enterprises and Establishments, the Register of Completed Education and Degrees or the Register of Job Applicants.

By using register data, the sample can be targeted so that adequate information is obtained to the survey on the desired sub-groups, e.g. women and men. This targeting can take place either by stratifying the population into different sub-groups, each of which are treated separately, or by clustering the units into clusters significant for the survey and usable in sampling. Register units also have exact location information, for which reason the units are relatively easy to reach for interviewing.

When the interviews have been conducted and the data collected, **additional information derived from the registers is used for improving the quality of the data obtained**. Any errors occurring in the data collection process, such as input errors, can be searched by comparing the data collected from the unit with regis-

ter data (*editing*). Non-response or missing information always present in surveys can be filled by utilising register data (*imputation*). In addition, data estimated from the sample can be specified by calibrating sample weights determined according to the method of sampling. In calibration new weights are estimated so that the population distributions derived from the registers, such as age distributions, are estimated correctly (*there are several of such new weights, from which the ones as close as possible to the original values are selected*). This change in weights lowers the error the estimates will have if the sample is skewed with respect to the population, and it can improve the reliability of the estimates considerably if there is a strong correlation between the used data and the survey variables.

The third way of using register data is to employ **register data to supplement the information content of the survey**, or to replace the collected data by register data. In Finland it has been customary to use demographic data as such in sample surveys, similarly as data on household-dwelling units and families and places of residence. Data on qualifications are usually derived from Statistics Finland's Register of Completed Education and Degrees for sample units as well and they are not collected again. For example, the data of the Register of Completed Education and Degrees are combined to the sample group by means of personal identity codes. Business Register data are also used when defining data on industry and location of workplace for the employed labour force. Thus it is possible to concentrate in the sample survey on the questions most essential for the survey, such as labour force participation, hours worked, consumption, and so on. Utilisation of register data often enhances the quality of the data as well. Interviewed persons may find it difficult or even impossible to specify their old qualifications, industry and legal form of their employer.

The fourth way of using register data is by the so-called **non-response studies**. Plenty of information is available from registers on persons or other survey units refusing to be interviewed or not reached for various reasons. The reliability of the sample survey can be assessed by means of such a study.

Totally register-based statistical systems and data sets combining survey and register data

Among the administrative sources used by Statistics Finland, a distinction can be made between totally register-based statistical systems and data sets that have been compiled or complemented by using administrative data. Examples of totally register-based statistical systems include:

- Population and vital statistics, families
- Population censuses
- Building and dwelling statistics
- Statistics on housing conditions
- Regional employment statistics
- Statistics on justice and crime
- Election statistics
- Income statistics

Examples of statistical systems that use register data for purposes of sampling and estimation and that are used as a source of additional and/or complementary information include:

- Income distribution statistics
- Household budget survey
- Labour force survey
- Wage and salary statistics
- Price statistics
- Business Register
- Structural business statistics
- Short term business statistics
- Dwelling prices statistics
- Statistics on accidents at work
- Education statistics

4.2. Register-based structural statistics, change statistics, combination statistics

Data contents of register data are also used in different ways in the actual compilation of statistics. The register databases are used in so-called structural statistics, such as those describing the number and age structure of the population in different regions, the number and attributes of business enterprises, or the size and age structure of the vehicle stock. These statistics are known as **structural statistics**.

Secondly, register data updates are used as a basis for compiling so-called **change statistics**. For instance, statistics on births, deaths, internal migration, marriages and divorces are based on register updates; the same applies to business start-ups and closures, registered vehicles, and buildings and dwellings com-

pleted. Most statistical systems are interested in both **structural data** and **change data**. There are two types of change data, i.e. **unit changes** and **attribute changes**.

The former, i.e. unit changes, include new units added to the register as well as units removed from the register, such as deceased or emigrated persons, closed-down businesses, demolished buildings, or scrapped cars. It may also be necessary to record changes to unit attribute data, such as marital status upon marriage or divorce, an enterprise's branch of industry or staff number, the floor area or equipment of a building, the owner of a motor vehicle, etc.

Structural data and change data are also combined to produce various **intensity figures**, such as fertility rates, mortality risks, propensities to move, etc., in different age groups.

The third type of statistics are those that are produced by **combining data from different register sources**. The linking of building and dwelling stock statistics with data on the occupants produces the housing conditions statistics, which describe the housing conditions of different household-dwelling units. The same method can be used to identify which families own a car or summer cottage, to determine the educational level or incomes of the gainfully employed population, or to describe the education and sex and age structure of a company' staff.

The fourth type of statistics is represented by so-called **flow statistics**, in which data for one and the same individual are chained together from consecutive years with a view to following statistical units over time. An example of typical flow statistics is the placement statistics describing transition from education to working life.

The Register of Buildings and Dwellings is used to compile the housing and building stock statistics as well as statistics on summer cottages. Statistics on housing conditions are compiled using data on housing dwellings. Statistics on housing development, the building of summer cottages, etc. are compiled on the basis of change data.

Educational data for any defined population can be obtained from the Register of Completed Education and Degrees. Register updates, i.e. data on the completion of new degrees, for their part, shed light on activities in the education sector during the past year.

The Business Register describes the existing stock of businesses as well as business activities during the past year. Change data in the register provide the basis for statistics on business start-ups and closures, changes in ownership, changes in branch of industry, etc. The Register of Job Applicants describes the number and structure of the unemployed population at a given point in time, register updates in turn provide information on the start and ending of periods of unemployment as well as periods of labour market training and placement. The Vehicle Register describes the existing stock of vehicles in the country, and change data provide information on all vehicles registered during the year. Reference data describing family relations can be used to produce intergenerational statistics (for instance on how children's characteristics and living conditions are influenced by parental background: children of well-educated parents tend to get a good education, some risk of unemployment is inherited, etc.).

5. *Cooperation with administrative authorities*

In the Finnish experience good and close cooperation with the relevant authorities is paramount to the effective use of administrative data sources. The use of these sources can be improved by working as closely as possible with the authorities, by exercising a real impact on the data content of registers, and by disseminating a better understanding of the use of administrative data for statistical purposes.

To these ends Statistics Finland has appointed for each register authority a contact person whose job it is to maintain open channels of communication with that authority, to monitor developments within the field concerned, and to work towards maintaining or improving the statistical applicability of register data. Each register authority has nominated a statistics contact person. In addition, Statistics Finland arranges annual meetings on the Directors General level with register authorities to discuss key issues and monitor progress in cooperation. This cooperation between Statistics Finland and register keepers is part of the coordination system of the Finnish official statistics.

Finnish authorities responsible of keeping basic registers also have a joint task force, the Register Pool. This group has the objective of promoting information exchange and cooperation among register authorities with a view to improving register usability and consistency, developing the contents, quality and accessibility of basic registers, facilitating the creation of effective information markets, and increasing cooperation among basic registers.

The following authorities are currently represented in the Register Pool:

- Finnish Vehicle Administration
- City of Helsinki
- National Land Survey
- Ministry of Justice
- National Board of Patents and Registration
- The Association of Finnish Local and Regional Authorities
- Statistics Finland
- National Board of Taxes
- Population Register Centre

The Register Pool is appointed by the Ministry of the Interior for two years at a time.

6. *Register-based population census system*

The population census provides the best example of the use of administrative records in Finnish statistics production. It is from this angle that the discussion below looks more closely at register-based statistics in Finland. Finland became the second country in the world after Denmark to succeed in conducting its population and housing census totally on the basis of register sources, without sending out questionnaires to the whole population. The first register-based census was completed in 1990. However, this important milestone was not achieved overnight: the long process of evolution began to unfold from 1970 onwards.

6.1. *The development of a register-based population census system*

6.1.1. *A brief history of population censuses*

In Finland, population censuses and population registration have been closely tied together for centuries. As early as the sixteenth century when Finland was still under Swedish rule, parishes used to keep a record of births and deaths for purposes of recruitment and taxation. In the seventeenth century these records were used by central government for its own survey and planning purposes.

Taulustolaitos (originally in Swedish Tabellverket), the predecessor of Finland's central statistical office, was established in 1748 to conduct the first ever population census in the country (Finland was still part of Sweden at this time). Parish and register office records were consulted to collect information on the number of population, marriages, deaths, sex, marital status and social class. Data collection for this first census took place in 1749.

Censuses were initially conducted at more or less irregular intervals, but from 1880 onwards they were taken once every ten years. Some of the biggest municipalities in the country began to organise their own complementary censuses from 1870. The Population Census Act was passed in 1938, laying down the requirement that the general census be taken every tenth year, complete with an enumeration of dwellings and buildings. The act remains in force to the present day.

Population and housing censuses based on direct data collection began in 1950, following the cancellation of the 1940 census because of the war. Prior to that there had been some smaller-scale inventories of parish registers. From the census in 1950, the range of data items covered has been wide indeed: in 1950 data were collected on persons, families, household-dwelling units, dwellings, business premises, buildings and summer cottages and even real estates. The 1950 and 1960 censuses were traditional questionnaire censuses where the data

were collected direct from the population. Since then, censuses have been taken once every ten years. In addition, interim censuses were completed in 1975 and 1985 on the basis of separate statutes, and in 1995 by virtue of the Statistics Act.

6.1.2. Towards a totally register-based census method

The development of a register-based population census system in Finland has been a long process, spanning a period of 20 years (Appendix 2). In population censuses based on the 1938 Population Census Act, administrative data sources were used for the first time in the **1970 population census**. In fact plans had long been in place to update and modernise the system of population registration. Automatic data processing was fast gaining ground at around this time, and the first personal registers were created in the early 1960s, the biggest of which was the register maintained by the Social Insurance Institution. The Population Register Centre was founded in 1969, and all persons permanently resident in Finland were entered in the **Central Population Register** (later called Population Information System).

In the 1970 population census, questionnaire data were linked with information drawn from the Central Population Register on religion and nationality, and with information drawn from taxation registers on income. In connection with this same census the statistical office also created a **Register of Completed Education and Degrees**, which was based on data on completed degrees collected in the census. Since 1970, this register has been annually updated on the basis of information provided by educational institutions. Working closely with the Population Register Centre, the statistical office also collected building coordinates and entered these data into the Central Population Register; these data, too, have since been regularly updated. This provided a sound foundation for the future development of regional statistics and for the use of geographical information systems.

The 1975 interim population census (also known as the 1975 Survey of Dwellings and Economic Activity) was less comprehensive in its coverage than a full-scale census based on direct data collection. The use of administrative data sources was increased. Data on age, sex, marital status, mother tongue and nationality were obtained from the Central Population Register. Data on degrees completed were drawn from the newly established register maintained by the statistical office.

The 1980 population census was based on a major questionnaire survey, but again it made increasing use of administrative records. In response to a growing recognition in the late 1970s of the need for an administrative source that would also facilitate more accurate statistics, a **Register of Buildings and Dwellings** was created in connection with the 1980 population and housing census. The data on buildings and dwellings were collected in connection with the population census by virtue of the now defunct Population Register Act using Population Register Centre forms. The Register of Buildings and Dwellings was created under the Population Register Centre, where it has since been maintained on the basis of updates from local building inspectors and local register offices. In subsequent

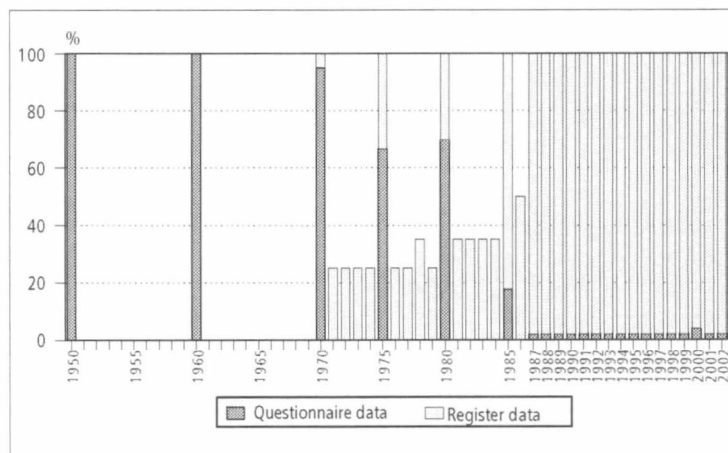
population censuses this register was consulted as a base source of information on buildings and dwellings, and from the mid-1980s, with improving reliability, it paved the way to annual register-based building and housing statistics. Statistics Finland's Register of Enterprises and Establishments was developed significantly in the 1980s both in its coverage and maintenance rate, one of the objectives being its utilisation for the register-based population census solution.

In its statement on the 1980 population census, the Ministry of Finance expressed the view that all future censuses in Finland should be conducted without any direct data collection among the population. In response, Statistics Finland immediately launched a project aimed at meeting this goal by the time of the 1985 census. As it turned out, however, the timetable was rather too ambitious, and data on person's place of work still had to be collected by questionnaires. All in all, by the time of the 1985 interim population census, the registers available had developed to such an extent that just one form was needed, mainly to collect information on the population's economic activity (main type of activity, place of work, branch of industry and location, occupation). However this form still had to be sent out to the entire population. As the Ministry of Finance refused to provide the necessary funding, local governments exceptionally undertook to cover around two-thirds of the costs of the census – concretely demonstrating the great importance of population census data to local municipalities in Finland.

The register project continued and eventually succeeded in reaching its target, totally register-based population census. As from 1987, Finland has been in the position to produce all its population census data entirely from administrative sources – a great leap indeed in Finnish statistics production.

The 1990 population and housing census was conducted entirely on the basis of register data, without collecting any questionnaire data at all. A total of some 30 administrative registers were consulted, most of which described the population's economic activity. Not a single questionnaire was sent out in the post, with the exception of one for a reliability study made in connection with the population census, which was used for collecting parallel data for comparing reg-

Figure 2.
Population census data 1950-2002 and the proportion of register data



ister and questionnaire data. The reliability study indicated that register-based data are at least as good as those used in previous questionnaire censuses. The costs of the register-based census were less than one-tenth of the costs of earlier censuses.

The same production model was used in the 1995 and 2000 population censuses as well.

6.1.3. Register-based population census system

The register-based population census system is built around a set of basic registers which contain comprehensive data on the units that are to be described in the population census. These registers include the data maintained by the Population Register Centre under the population information system and the Register of Buildings and Dwellings, as well as the data from the Business Register that is maintained by Statistics Finland. These registers cover all people resident in Finland, the buildings and dwellings in the country as well as all business companies and their establishments. All statistical units can be linked to one another by means of the identification systems: persons can be linked to families and household-dwelling units, to the dwelling and building in which they live, and to the employer for whom they are working. Similarly, all units can be located on the map using map coordinates.

In addition to the unit and attribute data contained in basic registers, the register-based population census system makes use of some 30 administrative registers which cover either the whole population or certain sub-populations. The most important among these sources are taxation register data (e.g. data on income, type of income, data on employer); register data from employment pension systems (e.g. data describing employment and employers); data from the Register of Job Applicants (e.g. data on unemployment); data from student registers; and register data from the Social Insurance Institution. Smaller registers include the data on persons in non-military service as well as employment data for persons in the employ of the Provincial Government of Åland. Appendix 3

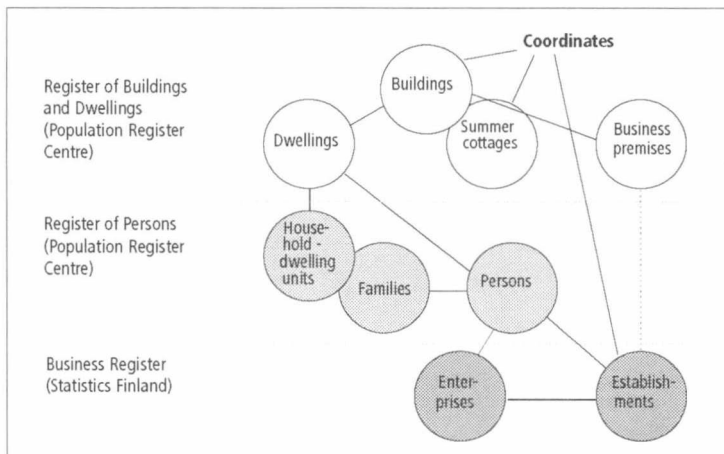
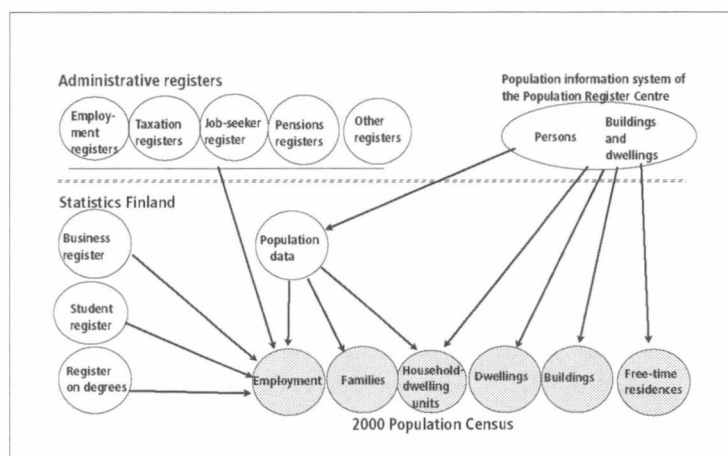


Figure 3.
Units covered by
the register-based
census system and
how they are
linked

Figure 4.
Data sources for the
2000 Population
and Housing Census



lists all the register sources that are used in the population census system. The key to the system lies in the identification systems of the different registers, which allow for accurate cross-linking of data. The registers may include overlapping data, but also contradictory data. Most importantly, the registers complement one another: the overlap that is built into the system means that if the required information for example on an employment relationship is not obtained through the Taxation Register, for instance, it is still available through the earnings-related pension system.

The population census data are produced using the method of register estimation, in which several register sources are used simultaneously to define for each statistical unit the value of the relevant variable. The decision rules are defined in such a way that the data they produce come as close as possible to the data collected by means of questionnaires. Data from earlier population censuses and register data from the same point of time are also consulted in constructing these rules. These include rules on prioritisation between different sources in the event of contradictory data.

6.2. *The pros and cons of register-based population census statistics production*

As was discussed earlier, work to develop a register-based population census system began after the 1980 census, in response to calls from the Ministry of Finance. The Ministry's main concern, of course, was with the recurring **high costs** of taking a census. Indeed, lowered costs are without question the biggest advantage of using administrative register sources. In 2003 money terms the 1980 population census cost 35 million euros, the latest census in 2000 cost less than one million euros. In addition, since the introduction of the register system, statistics on employment, buildings and dwellings and housing conditions have been compiled on an **annual basis**. Roughly, then, **the volume of statistics pro-**

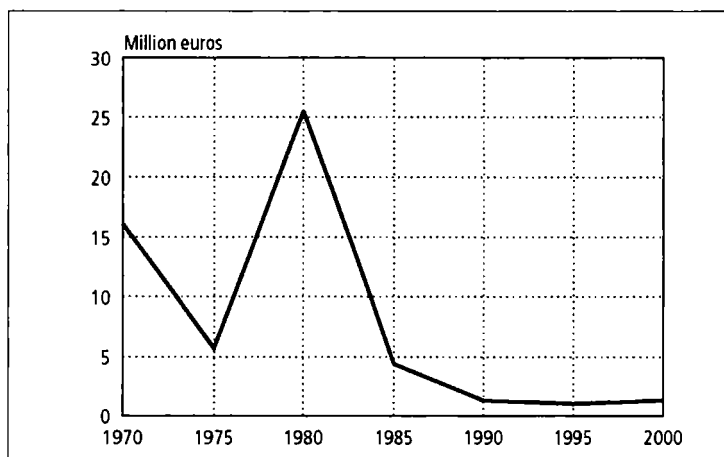


Figure 5.
Costs of censuses
1970–2000 in Fin-
land (million euros)

duced over a ten-year period has increased ten times over, while costs are down to one-third of the costs of one single major questionnaire survey.

A further key advantage of administrative sources is that **the need for processing is confined to those data items that have changed**. As a general rule it is cheaper to collect information just once and into one file and to process that information only if and when it changes. For instance, some 700,000 people in Finland change their address (i.e. their permanent place of residence) during any year, which means that for 4,500,000 people there is no change in place of residence. Nationality, religion and marital status change even less often; after a certain age it is very unusual to see changes in information on completed education and degrees; and in most dwellings the floor area and number of rooms never changes.

In most countries that do not use registers and administrative sources for census purposes, information on domicile, marital status, nationality, religion, degrees completed, number of rooms and dwelling floor area is re-collected and re-processed in connection with each census, regardless of whether there have been any changes. Some people may understandably be annoyed by this kind of 'unnecessary' data collection. Besides, sometimes the reliability of this information is affected by human forgetfulness. Over the decades people may well begin to have difficulty remembering which year it was that they completed their degree and especially shorter courses; the exact floor area of one's dwelling may seem a rather trivial matter especially to people living in the country; after years of unemployment, people may well lose track of the year they were first made redundant, etc.

Register systems also guarantee **unified data processing**. A student who works a few hours every now and then may indicate in a questionnaire he or she is a student, yet the definition of labour force is quite unequivocal: according to ILO definition anyone who works at least one hour a week shall be classified as being employed.

A register system also allows for the identification of **overlapping activities**. Thus it is possible to compile statistics on concurrent employment relationships, employment of students, etc. on a detailed regional level (though the Labour

Force Survey produces sample-based information on the same topic area), while on census questionnaire people usually classify themselves by one principal activity and their other activities remain outside compilation of statistics.

Register statistics are obtained from all geographical areas. Since registers cover the target population in its entirety, and since detailed geographical information can be obtained for all units, it is possible to produce all statistics for individual municipalities (of which there are currently 444 in Finland), and even for areas within municipalities (municipalities are divided into 12,000 sub-areas). Furthermore, statistics are produced for other area units, including post code areas, various freely defined areas and map grids of different sizes. It is this inclusiveness of its registers and the information on the exact location of register units that explains why Finland is in the position to produce geographical information at such a level of accuracy.

Register-based statistics are available every year. Growing information needs in the late 1960s and early 1970s created new pressures to step up the production of regional statistics: the regional data produced in connection with the ten-year population censuses were no longer enough to satisfy those needs. Again, this was a major asset of using register sources, allowing for more frequent statistics production. The dawn of register-based production also meant that many key statistics (including population and population trend statistics, family statistics, industry and employment statistics, building and housing statistics, statistics on educational structure) became available on an annual basis. It is hard to imagine how local governments, for instance, would cope today (in spring 2004) if the latest regional employment data (for example the structure and characteristics of the employed labour force: education, gender, age, place of residence, commuting, income etc.) available were for 2000 and if they had to wait until 2012 for the next set of statistics for the situation in 2010. Thanks to the register system now in place, municipalities currently have access to the data for 2002. Regional employment statistics have been produced annually since 1987. Demographic statistics have been produced annually in Finland ever since 1971. The production delay for data on workplaces is no more than 11–12 months, for demographic statistics just 3–4 months.

Register-based statistics production has also paved the way to a **broader scope and coverage of data contents** and to new methods of description, such as **flow statistics** at a high level of classification accuracy; examples include placement statistics on recent graduates and labour force flow statistics (for more details, see chapter 9.2.).

The reduced response burden on the population is a significant advantage indeed. In the 1980 population census people in Finland had to complete a total of some 8.5 million forms and answer some 100 million questions – even though existing administrative data were already being used quite extensively. If reading the instructions, filling in the form and mailing the questionnaire took each person 10 minutes, the total labour input for the population (and this is still a rather conservative estimate) amounts to more than two million hours. If we consider the price of one hour to be, say, 10 euros, then the overall cost of completing the census form would amount to 15 million euros.

The introduction of register-based statistics also has a dramatic impact on the statistical agency's job description of the personnel. There is no longer any need to design and test questionnaires, to send them out for printing, to pre-fill the forms, to mail them and send out reminders, to code and record the data. The agency no longer needs to recruit large numbers of people to sift through the returned forms. In connection with the 1980 census, for instance, some 2,500 employees were hired and trained. All this means that the statistical agency can concentrate its efforts on the most important job, which is to produce and analyse statistics. Indeed this is clearly seen in the rising qualification levels of Statistics Finland staff. Today, more than two thirds of staff members have completed at least a lower or lowest degree at tertiary level.

Furthermore, there is **improved data protection** as none of the answers of the population census ever have to be written down in plain language on paper. All the unit data are processed by computers. The number of people handling the data is also essentially reduced. All this improves the data confidentiality of the target population.

There is some lacking data. The use of administrative data sources also involves certain drawbacks that need to be taken into account. One such drawback is the fact that register-based descriptions have to rely exclusively on the information contents that can be formed on the basis of the registers available. This imposes some restrictions with respect to the phenomena that are available for description and may also undermine international comparability (more details on reliability and comparability in Chapter 6.3.). The use of registers also adds to the statistical agency's dependence on register authorities as well as on any changes in legislation and administrative practices. It is therefore crucially important to have close collaboration with the relevant authorities so that information on any changes reaches the office as soon as possible and ideally so that office can have its own say on the direction of change.

There are some data items that have had to be dropped from the register-based population census system because the relevant information is not available from any register: these include mode of transport to work, part-time work and mobile work. The 'crow flies' distance between place of residence and place of work can be measured using map coordinates. Samples can be used to collect such data that are not available through registers, for instance on weekly working hours, types of employment contract, working conditions, the household concept, income transfers between households, time use, adult education, prices, and wages and salaries.

Sometimes information may be collected with a survey on items on which register data are available, either because these data do not conform to international recommendations or because it would take too long to process the necessary data in time to meet international deadlines. For example, fast short-term business statistics are produced on the basis of separate data collection from enterprises and the Tax Administration's data are used to produce more comprehensive statistics by industry and area later on.

Since 2000 it has also become increasingly difficult to produce reliable register data on occupation. This is because of new procedures in the maintenance of

taxation registers, whereby data on occupation are no longer submitted for verification on each taxpayer's tax return form.

Data collection from registers will obviously shed no light for instance on grey labour or building work carried out without building permits, but on the other hand survey studies are unlikely to have much more success in this regard either.

Problems with reference periods and consistency. For reasons of statistical reliability it is important that change events are accurately recorded according to their true date. Information on dates of death and birth is usually accurate because it is recorded on the basis of certificates issued by the authorities: in most cases the reference time point is therefore right. Accurate information is also obtained on the dates of employment, unemployment and pension periods, whereas for periods of studying the dates are less accurate. In the event of a change of address the person who is moving may neglect to provide notification altogether, or be late in doing so.

The linking of a person's data on such variables as place of work, occupation and incomes from different register sources may sometimes give rise to consistency problems, i.e. it is not always clear that the information on occupation and branch of industry, for instance, describe the same period of employment.

Problems in linking data. There are some items in the Finnish register system where data linking has caused difficulty. The data on employment pension do not use the same business code as the taxation and business registers, and therefore extra work is needed to link individuals to the company where they are employed. Likewise, the linking of enterprises to the building where they are based is not always straightforward since the company address data are not necessarily fully accurate, or they may differ from the information in the buildings register.

There is no longer any collection tool for ad hoc needs. In many countries the population census system is an important tool of data collection that is used to meet emerging information needs. In Finland, additional information has been collected in previous traditional censuses on people working in Sweden, on fertility as well as on various aspects of the labour force. This flexibility is lost when data are no longer collected by means of questionnaires.

6.3. Reliability, comparability and timeliness of register-based census data

6.3.1. Reliability

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. Questionnaires were still used in the 1975–1985 censuses, and at that point the comparisons were made with other survey data. In 1980 the labour force survey was expanded to comprise some 70,000 sample persons in order to have a sufficiently reliable foundation for comparisons. It was concluded that the differences

between the data produced using registers and questionnaires were so small that the register-based census system could be given the go-ahead.

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 information. 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. Comparative studies have been published in Statistics Finland's series.

Annual quality checks are carried out to monitor the reliability of register-based employment statistics, for instance, while the accuracy of demographic data is ascertained by the Population Register Centre. Each year the Population Register Centre commissions a survey to establish the accuracy of address data recorded in the population information system. To this end Statistics Finland will send out inquiries to some 10,000 people. In 2003, 99.0 per cent of the respondents indicated that the information was indeed up-to-date. Annual quality controls have emerged as the most important means of checking the reliability of regional employment statistics: these controls involve comparisons with the results of the labour force survey. Assessments of the reliability of the 1995 and 2000 population censuses were also based on comparisons with the labour force survey.

These comparisons with the labour force survey work at two different levels. On the one hand, their aim is to monitor the quality of the results produced by the two methods, and on the other hand to establish how accurately the methods produce information that is classified in the same way at unit level.

Comparisons at unit level are done by cross-tabulating the register data and the interview data from the sample of respondents in the labour force survey with data on main economic activity and branch of industry at the same point in time. These comparisons have been carried out since 1987, when the deviations were analysed in some considerable detail. Any cases where the deviation increases on the previous year will be scrutinised by checking the source material at unit level. If it turns out that there has been an error in data processing, that error will be corrected; if, on the other hand, the explanation lies in changes made to the register source or to legislation, for instance, then the necessary changes will be made to the decision rules applied.

Deviations are often explained by differences between the register system and the interview method, and it is not always possible unambiguously to say which of the two methods gives the right result: even information obtained from interviews is not necessarily absolutely accurate. In many cases the respondent or

interviewer is confronted with the same kind of contradictory information for which there is a straightforward decision rule in the register method. The register method here has the advantage of logical consistency: presented with the same information, the computer will always make the same decision, whereas two different persons may well arrive at different conclusions.

The drawbacks and shortcomings of the register system are rather minor when compared to questionnaire surveys, which in Finland have reached response rates of around 97–98 per cent – after meticulous and expensive processes of sending out reminders. As some of the forms returned will furthermore have incomplete and/or lacking data, the quality of the information collected will rarely be of the same standard as that obtained from register sources.

6.3.2 *Comparability with international recommendations*

Register-based population census data largely correspond to the recommendations for the 2000 censuses of population and housing issued by the United Nations Economic Commission for Europe and the Statistical Office of the European Communities. According to the recommendations, there are three methods of collecting census data, namely:

- the traditional method of using census questionnaires;
- the method of using registers and other administrative sources;
- a combination of registers and other administrative sources and surveys (complete enumerations or sample surveys).

Finland's register-based census system is thus as a method in full compliance with the international recommendations.

According to the recommendations, data are to be collected in population and housing censuses on the following statistical units: (a) persons, (b) private households, (c) institutional households, (d) family nuclei, (e) living quarters, (f) buildings.

Data can be produced on all these units in the register census.

Place of enumeration. Living quarters and buildings have fixed locations and information on where they are is needed. Information on households, however, and on the persons in households, can be collected and entered on the census questionnaire at the place of their usual residence and/or at the place where they are found on the day of the census. In Finnish population register the concept of permanent residence is used according to the recommendations, and persons are counted in their permanent places of residence.

The family concept. A family nucleus is defined in the narrow sense as two or more persons within a private or institutional household who are related as husband and wife, as cohabiting partners, or as parent and child. Thus a family comprises a couple without children, or a couple with one or more children, or a lone parent with one or more children. Register data can be used to define family concepts fully in line with the recommendations.

A private household is either: a one-person household or a multi-person household as **the housekeeping unit concept**. Some countries use a concept of **the household-dwelling concept**. This is defined as the aggregate number of per-

sons occupying a housing unit. It is recommended that countries applying the household-dwelling concept give an estimate of the total number of housekeeping units in the census report.

In Finland the household-dwelling concept is used in population censuses and dwelling statistics and nowadays also as a background concept in family statistics. In household budget surveys and income distribution statistics the household concept is formed on the basis of the interview and the number of household-dwelling units in the whole country is estimated.

An **institutional household** comprises persons whose need for shelter and subsistence are being provided by an institution. The population register contains data about persons living in institutions and these are used in statistics.

Household-dwelling units are still classified according to the type of family living in them. All data on household-dwelling units can be produced on the basis of the population register.

Time of enumeration. Finland's accurate population information system makes it possible to calculate the population, dwellings and buildings on whatever day. The reference time used mostly in the enumeration is the last day of the year.

Demographic characteristics of persons. The population register contains all the usual demographic data, such as age, sex, legal marital status, de facto marital status, country/place of birth, country of citizenship, ethnic group, language, religion, total number of children born alive, date of marriage. These all are in the register except for ethnic group, on which data are not produced in Finland.

The classification of the population's main type of activity divides the population into the following groups: employed, unemployed, persons attending educational institutions, pension recipients, homemakers and others.

All these groups can be defined by means of registers, as can such data as: duration of unemployment, occupation, industry, type of sector, status of employment, number of persons working in the local unit, income, place of work, socio-economic group, location of school or university.

All employment relationships on which employment pension insurance is paid and taxes are deducted can be found from the register statistics. All so-called moonlighting work remains outside the statistics, similarly as unpaid work by family members of entrepreneurs.

Unemployed persons are determined on the basis of the Ministry of Labour's Register of Job Applicants, and the concept used is not fully compatible with the concept of the Labour Force Survey. The register thus includes such unemployed persons who have given up active search for work.

The following cannot be defined by means of register data:

Time usually worked

Mode of transport to work, and

Length and frequency of journey to work.

These data can be produced only by interview surveys.

Educational characteristics of persons. Statistics Finland's Register of Completed Education and Degrees maintained from 1970 includes for the whole

population the following data: educational attainments, educational qualifications, and field of study. The Student Register provides information about all persons currently studying (school attendance) and also about educational institutions.

Data about the characteristics of buildings and dwellings are now derived very exhaustively from the building and dwelling register. The tenure status can also be deducted from the registers.

The register does not contain information about the dwellings' durable consumer goods, for which reason data about them can only be had from interview surveys. The number of cars used by the household-dwelling unit is available from the motor vehicle register.

Each person is defined a domicile at the accuracy of dwelling and also by means of map coordinates. These data can be used to define any geographic regional division and data by square grid, for example.

6.3.3. Timeliness

During questionnaire censuses the most important preliminary data could be produced within about one year and the last final data around 2 to 2.5 years from the time of the census. The register system has speeded up the publication of census data covering the whole population. Demographic data are complete in about three months, data on household-dwelling units and families in around four months, data on buildings and dwellings in about six months, and educational data in eight months. Preliminary data on the economically active population are obtained in about one year and final data in around one year 10 months. The last to be finished are data on occupation and income. Income data take a long time because taxation lasts about one year in Finland and only after that the data are available for statistics. Occupational data require partly direct data collection and partly manual coding and therefore they are completed last of all.

In addition to the acceleration of data production, the main benefit gained is that all the data, except on occupations, can be produced every year.

7. Register-based statistics systems

7.1. Statistics under the population census system

The register-based population census system comprises several statistical register systems, including statistics on population structure, family and household-dwelling statistics, regional employment statistics, buildings and dwelling statistics, housing conditions statistics, as well as the separate register on completed education and degrees, data on occupations and related geographical information systems. With the exception of occupational data, all these data are produced regularly on an annual basis.

7.1.1. Register data on population structure and changes

A statistical register of persons was first created in Statistics Finland in connection with the 1970 population census: this was updated over the next four years using population change data collected by the Population Register Centre. From 1975 onwards population statistics have been produced entirely on the basis of the Population Information System (then the Central Population Register). Then the maintenance of Statistics Finland's own statistical register of persons ceased completely.

Nowadays register sources are used in Statistics Finland to produce the following annual statistics for municipalities and municipal sub-areas:

- Population structure statistics
- Household-dwelling unit statistics
- Family statistics.

Statistics describing population change, i.e. those on

- Births
- Deaths
- Migration between municipalities
- Migration within municipalities
- Immigration and emigration
- Marriages and divorces

are produced using register updates on a monthly basis for municipalities and on an annual basis for municipal sub-areas.

Concepts

A household-dwelling unit consists of the permanent occupants of a dwelling. Persons who according to the Population Register Centre's population information system are institutionalised, or are homeless, or are abroad, or are registered as unknown, do not constitute household-dwelling units. Additionally, persons living in buildings classified as residential homes do not form household-dwelling units if their living quarters do not meet the definition of a dwelling. The concept of household-dwelling unit was adopted in the 1980 census; prior to this the concept of household was used, referring to family members and other persons living together who made common provision for food. A subtenant providing for his or her own food constituted a separate household. Since 1980 subtenants have been classified in the same household-dwelling units with other occupants.

A family consists of a married or cohabiting couple and their children living together in the same dwelling; or a parent and his or her children living together; or a married or cohabiting couple without children.

Persons living in the household-dwelling unit who are not members of the nuclear family are not included in the family population, even if they are related, unless they form their own family. Brothers and sisters or cousins living together are not a family and do not belong to the family population. The same applies to people who live alone or with a person of the same sex. Families living in residential homes are included in the family population. By contrast, persons who live in institutions are not included.

A cohabiting couple is defined as two spouseless adults of different sexes aged 18 or over and occupying the same dwelling on a permanent basis, provided their age difference is less than 16 years and they are not siblings.

In other words, then, these family and household-dwelling unit statistics are based entirely on the central population register and its update information. Furthermore, data on people living in the same dwelling unit are used for purposes of forming household-dwelling units and families. Other register data are also linked to individuals, such as those on employment, education and incomes.

7.1.2. Register-based building and dwelling data

Statistics Finland uses the building and dwelling data in the Population Information System to compile statistics on building and housing production, the building and dwelling stock and housing conditions. In addition, the register data on buildings and dwellings are used as the sample population for various surveys concerning dwellings or households, such as the rental inquiry and the household budget survey.

The building and dwelling register has been used in compiling **statistics on building and housing production as well as on building permits** since 1982. Information on new building permits, building starts and building completions is updated in the Population Information System by local building inspectors. Since the statistics on building production are compiled from register sources that are used for other administrative purposes as well, the quality requirements for this register are more stringent than they would be for a data file used for statistics production.

The data files on the **building and housing stock** and **housing conditions** are produced once a year using the relevant data from the Population Information System. Statistics Finland uses the domicile code to form household-dwelling units of all persons permanently resident in the same dwelling. Summer cottages are drawn from the building stock by reference to the classification of the building's intended use.

Person and dwelling are linked using the domicile code, which is composed of the real estate code, building number and dwelling code. **Household-dwelling units** formed at Statistics Finland are linked with register dwelling data through the domicile code. Taxation data for the household-dwelling unit, most commonly unit incomes, are linked with the household-dwelling data set.

Dwelling occupancy status is updated on the basis of the resident's change-of-address notification and also on the basis of the stamp duty register for purchases of shares in housing corporations.

The sample population for the **housing conditions statistics** consists of all persons permanently resident in dwellings proper. Persons living in institutions and boarding houses are not counted among this population. People who move to a boarding house or an institution or who become homeless are given a domicile code based on their municipality of domicile, which includes separate codes for village and town district. These people are removed from the household-dwelling files.

Statistics on the building and housing stock, summer cottages and housing conditions are produced each year for all municipalities and sub-areas within municipalities.

7.1.3. Regional employment statistics

Register-based employment statistics are among the most important statistical systems in Finnish population statistics. They also have an important role in register-based population and housing censuses. The main purpose of these statistics is to provide annual regional data on the population's economic activity: these include such items as main type of activity, industrial status, branch of industry, location of workplace, and income.

Planning for a register-based system of employment statistics started in 1981, and the first set of employment data based on this system was produced in 1987. Since then this information has been produced annually. Before 1987 the system was tested twice (in 1980 and 1985) by comparing register-based data with data obtained from questionnaires. In addition, the register data have been compared annually with the data from labour force surveys.

Some 30 different registers are used in compiling the annual employment statistics. Most of them are needed for defining main type of activity at year-end, which is indeed the primary task of employment statistics.

The Population Information System is a basic register that defines the population covered by the system of register-based employment statistics system. This population comprises all persons whose official domicile on December 31st, according to the Population Information System, is Finland. The Population Infor-

mation System is administered by the Population Register Centre. The data on domicile are updated on the basis of change-of-address notifications.

Data on employment relationships are obtained from several different sources. There are three main types of employment relationship, i.e. those in the private sector (including the self-employed), state sector, and local government sector (including municipal federations).

All private sector employers in Finland are under obligation to take out pension insurance for their employees through an insurance company. Pension arrangements for employees in central and local government are slightly different. Entrepreneurs provide for their own pension insurance.

In connection with data collection from various administrative sources, a separate register is compiled on all employment relationships in the current year. In addition, the Business Register and other registers on public sector establishments are consulted to obtain information on data on branch of industry, ownership type and legal form of employment relationships. For people employed in enterprises with several establishments, the data on branch of industry are only available for the last week of the year on the basis of a separate inquiry.

Data on location of workplace. Data from the Central Pension Security Institute (employment relationships in the private sector) only include information on the enterprise in which people are currently employed. For enterprises with a single place of business, data are assigned to individuals according to the business data (branch of industry, type of ownership, institutional sector and location). If an enterprise has more than one place of business, individuals are allocated to the unit specified by the enterprise in the inquiry.

Statistics Finland conducts annual inquiries to collect information on enterprises with several places of business: these inquiries cover all enterprises with more than one establishment and at least ten employees (approx. 4,000 enterprises). In this connection enterprises are requested to provide information on each employee's place of work within the company at the end of the previous year. Pre-filled questionnaires based on data from the previous year are mailed to each enterprise for updating. The enterprises may also provide the relevant information on disk or in the form of computer printouts.

In the case of some **entrepreneurs** and **self-employed persons**, such as farmers, branch of industry is deduced from pension insurance data and type of income. The establishment's location is the entrepreneur's home address.

In addition, an annual questionnaire is sent out to the offices of local governments and municipal federations that have more than one establishment (approx. 1,000 units).

At the end of the 1960s, the income tax data collected by local tax authorities were gathered into a **Tax Register covering all income recipients**. A Tax Register is compiled each year by the National Board of Taxes on the basis of figures reported by individuals in their tax returns. This register provides a wide range of data on income and benefits received as well as on the taxes, assets and liabilities of the persons subject to taxes. Data on wages/salaries for all employment relationships are obtained from the employer's statement of earnings and tax deductions to the National Board of Taxes. These data are needed for purposes of

identifying the employment relationship valid at year-end. Every year, as a control measure, Finnish employers provide the tax authorities with figures on how much employees have earned and how much tax has been withheld on employee earnings. These data are given under the employers' ID number. The resulting data are then combined into a national register, which makes it possible to determine the organisation (identification) number of the employer of every employee.

The Tax Register was used for the first time in the 1970 census. The data obtained from these sources include income subject to state taxation, entrepreneurial income, wage and salary income, assets subject to taxation, deductions, debts, capital taxes, capital revenues, etc.

The Ministry of Labour's **Register of Job Applicants** contains information on all persons who are seeking work through the network of regional employment offices as well as on all persons currently unemployed. The data used in employment statistics refer primarily to unemployment and persons in subsidised employment programmes (since 1993).

Information on conscripts and conscientious objectors is obtained from **the Defence Staff Register of Conscripts and the Ministry of Labour Register of Conscientious Objectors**.

Registers maintained by the Social Insurance Institution and the Central Pension Security Institute provide information on pensioners. The data items consulted for purposes of employment statistics are initiation date and type of pension.

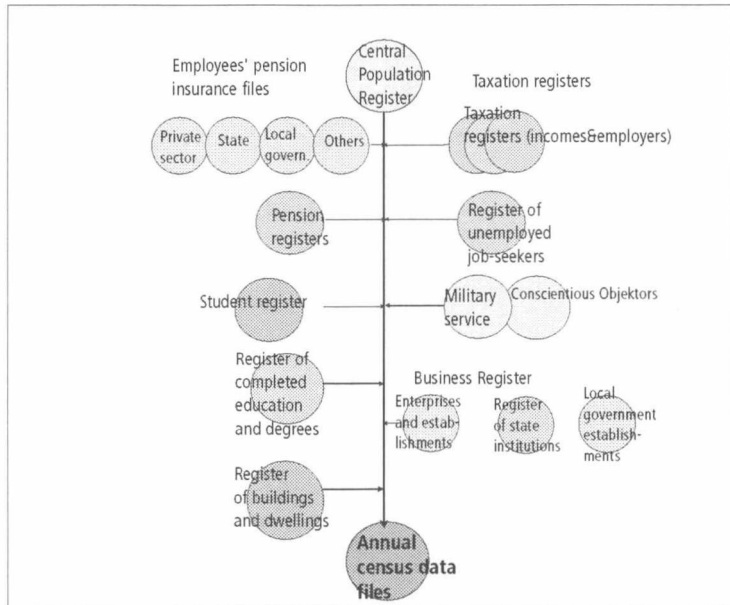
Regional employment statistics also use the registers created by Statistics Finland. Information on studying is obtained from the Student Register and on degrees from the Register of Completed Education and Degrees.

Statistics Finland compiles a **Student Register** from a variety of sources: Statistics Finland's Register of University Students provides information on persons who are studying at a university during the autumn term or who are registered at a university, as well as on the type of university studies. The Register of Study Aid to Students provides information on study aid granted during the year (during the autumn and/or spring term). The Joint Selection Register lists all applicants and entrants to vocational schools or colleges and upper secondary schools in a given year. From 1999 onwards data have been collected at individual level on students in post lower secondary education, which has helped to improve the quality of student data.

Statistics Finland's Register of Completed Education and Degrees lists all the degrees taken in Finland as well as most degrees completed by Finnish citizens abroad. This register was started in connection with the 1970 population census. Since then the register has been compiled annually with information collected from different educational institutions. Information on degrees completed by Finnish students abroad is obtained from the State Study Aid Centre and the National Research and Development Centre of Welfare and Health. The register does not include the degrees completed by foreign nationals before moving to Finland.

For the purposes of regional employment statistics the Register of Completed Education and Degrees is consulted to obtain information on the highest degree

Figure 6.
Use of registers and administrative records in register-based employment statistics



completed. However the register also contains information on all degrees a person has taken: the degree itself, date when taken, location of educational institution, type of educational institution, and code of educational institution.

7.1.4. Occupational data

In the 1990 and 1995 population censuses, occupational data were produced entirely on the basis of registers. Occupational title data were gathered from diverse register sources and machine-coded using an occupational dictionary. Where machine coding was not possible, manual coding was done using computer terminals.

The sources for occupational data are as follows:

- Occupational data on persons employed by central government
- Occupational data on persons employed by local government
- Occupational data on persons employed by the state church
- Occupational data on persons employed by the Social Insurance Institution
- Employer organisations' data on the occupations of persons employed by organised enterprises
- Occupational data collected with entrepreneurs' tax returns
- Population register system data on the occupations of persons who have moved
- Occupational data on unemployed persons in the Register of Job Applicants

In the population and housing census in 2000, some occupational data had to be inquired from employers using questionnaires. This was due to the adoption of a new tax proposal system in which data on occupation are no longer submitted each year for verification to each taxpayer.

Appendix 4 provides a more detailed description of the register sources used in producing occupational data.

7.1.5. GIS and register-based census

Traditionally, the most important regional unit in statistics has been the administrative area. The area code system for administrative areas has been the cornerstone of regional statistics. However, administration is inherently dynamic and keeps changing, which means that the boundaries of administrative areas are also constantly changing. It is not easy to keep up with these changes if the smallest statistical unit is an area whose boundaries continue to change with time.

Buildings were used as the basic unit of regional statistics for the first time in the 1970 population census. This meant that the locations of buildings had to be specified using map coordinates: these gave the exact location of each statistical unit.

The Finnish register-based statistical system is therefore **point-based rather than area-based**. This 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 paved the way to more flexible determination of statistical areas. The use of building coordinates and the phenomena occurring within them, i.e. in an area specified on the map, may be described as a system of 'regional statistics'. With most generic GIS packages, this point-in-polygon analysis can be performed with data located to points (buildings) and polygons (statistical areas).

The most commonly used non-administrative areas in statistics production are as follows:

- a statistical classification between urban, semi-urban and rural areas
- localities (urban settlements)
- municipal sub-areas
- post code areas
- 1 km x 1 km grid squares.

Figure 7 illustrates the data content of the register system in the area of one municipality. The register includes all buildings in the municipality and the dwellings in the buildings. The person register contains all persons living in the dwellings and people have exact information on which dwelling they are living (address and domicile code). If the person moves, the domicile code is changed into that of the new dwelling.

Enterprises and their establishments operating in the municipality are combined by means of their addressed to their location buildings. When all buildings have been assigned map coordinates, this means that all the people living and the enterprises located in the buildings have up-to-date coordinate data. By means of coordinate data it is possible to place the municipality's population, buildings,

Small area concepts

Localities and rural agglomerations. In the very first population census that covered the whole of Finland in 1950, localities were used as statistical areas alongside administrative areas. At that time population centres with at least 500 inhabitants were marked with a free-form boundary line. However, it was not until 1960 when a joint Nordic definition was adopted and applied for the first time that work began to compile statistics on localities and Statistics Finland geographical information.

According to this **Nordic definition**, all clusters of buildings at a mutual distance of less than 200 metres and with a total of at least 200 inhabitants, are classed as localities. In the first few instances the aim was to roughly locate such population centres on as large-scale maps as possible and to calculate how much of the area they covered. However it was not yet possible to produce statistics proper. It was not until 1970 that it became possible to sum up of the data, once the centroids of the buildings had been determined with coordinates.

Although the definition of locality has remained unchanged since 1960, the methods used in drawing up the boundaries of localities has changed in step with the development of data technology. Geographical information software made it possible automatically to demarcate localities in 1990. This produced a free-form line which followed the edge of the building cluster. The emergence of the polygon created by the line could be directly influenced by a function which calculated the distances between the buildings and formed population agglomerations according to the given specification. In 1998, mainly in response to the needs of rural studies, the method for delimiting localities was further developed to cover rural agglomerations with fewer than 200 inhabitants as well. The delimitation in 1998 focused on rural agglomerations with 30 to 500 inhabitants; the production of statistics on these agglomerations is currently in progress.

Municipal sub-areas. Since the 1970 population census, municipalities have been able to define the sub-areas by which register-based statistics are produced for them. However, the introduction of a sub-municipal division has been voluntary rather than aimed at creating standard-sized areas. The service was born out of individual municipalities' own needs, and indeed the sub-area system has been adopted by nearly all Finnish municipalities. Some municipalities apply several different sub-area divisions for the statistical needs of different administrative authorities.

Post code areas. In addition to their code and map coordinates, all buildings in Statistics Finland's files also have an address, which includes a five-digit post code. This code has facilitated the aggregation of statistics by post code areas since the 1980s. Demand for statistics by post code areas did not start to pick up until the 1990s. The main customers are enterprises involved in marketing and especially those producing market analyses. Gradually it also became necessary to create post code boundaries for geographical information systems. Post code boundaries are updated annually.

Grids and other flexible areas. Grid data generally refer to those statistics on population, labour force and employment in which a regular grid square constitutes the statistical area. The location of the squares is pinpointed with map coordinates. The most commonly used square size is 1km x 1 km.

Data by grid square have been available from Statistics Finland since the 1970 population census. Grid square data are regarded as the most flexible statistical geographical data. The square is a spatial stable: it does not move from one year to the next in the way that administrative areas do. It is therefore independent of any regional changes. Data by small squares can easily be summed up to form larger areas.

As with municipal sub-areas, it has also been possible to produce statistics on other areas specified by the customer.

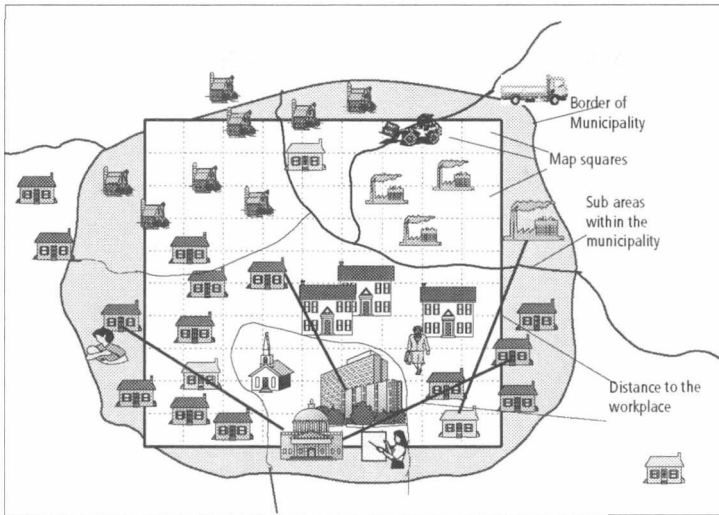


Figure 7.
Units of regional
statistics and links
between them

dwellings and workplaces within the borders of the municipality, in municipal sub-areas or into square grids of desired size, provided that the boundaries of the areas are digitised.

The register system is used to find out each year the employer enterprise of each employed in the municipality and the establishment of the enterprise where he or she is working.

The link connection to be kept up-to-date

person -> dwelling -> building -> map coordinate
and the link connection

employed -> enterprise -> establishment -> building -> map coordinate
make it possible to place people into correct dwellings, buildings and sub-areas
and employed persons into correct enterprises, establishments and small areas.

Regional statistics are produced by compiling statistical units of any area – persons, employed, dwellings or buildings. It is also possible to calculate the distance of an employed person from his or her workplace or the length of a pupil's journey to school and find out how many people commute to work to the area of the municipality from other municipalities.

7.2. Use of register data in some other statistics

A couple of other sets of register-based statistics are discussed below by way of an illustration: these are income distribution statistics, Labour Force Survey, Business Register and structural and short-term business statistics.

7.2.1. *Income distribution statistics*

Income distribution statistics are an annually conducted survey describing income of households and income differences between households in Finland. In the income distribution survey households are interviewed either by personal interview or by telephone. Households are formed and their structure is determined on the basis of interview data. Interviews are also used to find out the activity of the persons belonging to the household in the calendar year, occupational and workplace data on employed household members and information about households' dwelling. In the interviews data are collected about income transfers between households but all other income data derive from registers.

Various administrative registers are utilised widely in different production stages of the income distribution statistics. For example, as much as over 90 per cent of households' income is nowadays obtained from registers. Register sources of the income distribution statistics are such as the Population Register Centre's Central Population Register, the Tax Administration's tax database, the Social Security Institution's benefit recipient registers, the Finnish Centre for Pensions' pension register and Statistics Finland's Register of Completed Education and Degrees.

Total income distribution data covering the whole population are based on registers and serve the income distribution statistics and survey. The total income distribution data have as the income use unit the **household-dwelling unit** derived from register data, instead of the household based on the interview.

When the survey data based on registers cover the whole population, many uncertainties related to sample surveys are avoided, such as problems related to the sampling design, non-response, standard errors, etc. In addition, total data enable production of reliable income and income distribution analyses on considerably smaller areas than by sample data. For example, in a regional survey on income differences the total data based on registers are the only reliable survey data. The total data covering the whole population make it possible to conduct reliable longitudinal analyses, that is, they can be used for following the income development of people and population groups.

7.2.2. *Labour Force Survey*

Selection and updating of the sample. The sample of the Labour Force Survey is drawn twice a year from Statistics Finland's database based on the Central Population Register. Working-age people are selected to the sample (aged 15 to 74). The sample is updated by the Population Register Centre's population change data (addresses updated, deaths removed).

Data collection. Address data derived from the register are used for the target's contact information and when dividing the targets between interviewers. Sample data are also used as response data. Address and other data are also updated by the data obtained in connection with interviews.

For data collection, the latest data concerning workplace and occupation are retrieved from the **register-based employment statistics**, used as additional infor-

mation for reaching interviewees and also in interview situations. Employment statistics data are pre-filled in the paper questionnaire, used as an additional questionnaire in the interview. The actual response data are collected with computer-aided Blaise questionnaires.

Processing of the data. **Statistics Finland's Register of Enterprises and Establishments** is used in coding the industry and employer sector for employed interviewees. By means of the establishment's name, industry data and address obtained by the interview, the establishment is identified from the register. The qualification attained is searched for the sample from **Statistics Finland's Register of Completed Education and Degrees** (code and plain-language text) to be used as additional information in coding the occupation.

Response data. Some of the data on the sample person or the household are derived from registers in connection with sampling and they are taken updated to the printout file. Information on qualifications attained come from **Statistics Finland's Register of Completed Education and Degrees**. When calculating the results the qualification data are updated afterwards so that they correspond to the survey period. Data connected with regional classifications are from **Statistics Finland's register of municipalities**. Income data are used in customer data files and they are selected from **Statistics Finland's income distribution statistics data**, to which the data derive from different registers (**National Board of Taxes, National Research and Development Centre for Welfare and Health, Social Security Institution, Finnish Centre for Pensions, etc.**).

Estimation. Estimation uses data on the respondent's sex, age, area and job applicant status and also preliminary population figures relating to the whole population by address. The data are derived from **Statistics Finland's population database**. The data on job applicant status are obtained monthly from the **Ministry of Labour's Register of Job Applicants**.

Use of register data in the Job vacancy survey. The sample is selected from **Statistics Finland's Business Register**. Addresses of establishments, industry codes and size of personnel are also obtained at the same time. Register data are also used for contacting, as response data and in estimation.

7.2.3. Business Register

The most important registers for Finnish business statistics are the registers maintained by the tax authorities. The first registers intended for purposes of business taxation were developed in the 1970s to include companies liable to turnover tax, for example. In the 1980s the tax authorities revised their information systems of business taxation, converted them into electronic format and introduced the Business Code. After these revisions **Statistics Finland** was in the position to construct an almost exhaustive Business Register, where information on business start-ups and closures could be maintained almost in real time. Until the 1990s, the use of business taxation registers in official statistics was centred on the **Statistics Finland's Business Register**.

The Business Register is a basic statistical register that covers all business companies and organisations in the country, their establishments and operation.

The register has been regularly maintained since 1970. It also comprises central government and local government units, the former since 1985 and the latter since 1990.

The basic units of the register are the **legal unit** (enterprise), its local units and places of business; nowadays the group level is included as well. Each unit has its own unique identifier. The identifier for the legal unit is the Business Identity Code. In Finland all business companies and organisations are registered under a Business Information System (BIS) jointly administered by the tax authorities and the National Board of Patents and Registration, where they receive this code. Other unit codes are assigned at Statistics Finland.

The Business Register is nowadays primarily maintained and updated on the basis of administrative records and other sources. This is greatly facilitated by the system of unified codes across different registers.

The most important administrative sources are:

- various records maintained by the tax authorities (customer database, annual notifications, data on business taxation, data on payment control)
- Bank of Finland records
- Finnish Customs records
- National Board of Patents and Registration records
- State Treasury records
- Local Government Pensions Institution records
- Finland Post records

Data from administrative registers are used as a source for the Business Register in two different ways: either as such, or as a basis for estimating data that are not directly available from the registers.

Not only administrative sources but also materials produced by private information providers are nowadays consulted in compiling the Business Register. Examples include data from Suomen Asiakastieto Oy on consolidated financial statements as well as collaboration with Opasmedia Oy (who check certain basic information on companies by phone).

The Business Register is updated by reference to administrative records as well as data collected through business surveys. Inquiries are sent out to all companies with more than one place of business, to a sample of single-location companies, to recent business start-ups and to government agencies. The survey among multi-location companies and government agencies is carried out in conjunction with the data collection for regional employment statistics. These Business Register surveys produce complementary information that is essential for register maintenance and that is unavailable through administrative sources (location data for multi-location enterprises, data on business transfer), or where data quality is not good enough, or where it is considered necessary to carry out checks at regular intervals (branch of industry and addresses).

Enterprise units as well as data on the enterprise, such as name, address, municipality, legal form, preliminary branch of industry and dates of start-up and closure, are obtained from the tax authorities' customer database. The informa-

tion is updated four times a year. **Local units** (establishments) and data on them are collected by Business Register surveys as described above.

Number of wage earners in the company is estimated on the basis of the tax authorities' annual notification data and other records. In addition, a cruder monthly estimate is produced using employers' payment control data.

Number of entrepreneurs is estimated on the basis of register data from the Central Pension Security Institute and the Finnish Tax Administration.

Data on turnover and payroll costs, as entered in company balance sheets, are obtained from the business taxation records maintained by the Tax Administration. Balance sheets data are updated as and when they are received, which is currently four times a year.

Data on **branch of industry** and **sector** are entered for each unit as soon as it is added to the register. These data are constantly updated on a running basis, but the final data for each year cannot be deduced until all the information for the statistical year has been completed. For instance, the main activity of a multi-location enterprise cannot be established until all the location data have been completed for that enterprise. **Data on ownership** and **data on foreign ownership** are obtained from the Bank of Finland and Invest Byrå in Finland Oy, and they are updated from consolidated financial statements.

The Business Register is designed with a view to maintaining an up-to-date basic register so that data on business structures, start-ups, closures and classifications are as real-time as possible. The final statistical files for each year and the statistics describing Finnish enterprises and establishments are completed once a year.

7.2.4. Register-based business statistics

The 1990s saw much progress in the use of administrative records for purposes of compiling business statistics. First, from 1995 to 1996, the Business Taxation Register was adopted as the primary source for structural business statistics. The production of monthly statistics on enterprises' turnover and wage bills was started in 1998. These are based on the tax authorities' payment control register, which contains information on value added taxation and wages and salaries paid by registered employers. **Structural business statistics** are compiled annually by combining questionnaire data with information from the Business Taxation Register and the Business Register. Direct data collection involves only a few per cent of all enterprises. **Short term business statistics**, which describe monthly turnover and wage sum indicators, are compiled by combining questionnaire data (collected from less than one per cent of all enterprises) with data from the VAT and employer payment register and the Business Register.

Although the architecture of data collection for structural statistics and monthly indicators is similar, the methodological challenges in these two branches of statistics are quite different. In structural statistics, deficiencies in administrative records are largely corrected by means of automatic editing and imputation. In the case of monthly indicators, by contrast, the problems of timeliness and undercoverage are primarily addressed by means of panel methods.

Data on business taxation may be lacking altogether for some enterprises, and in quite often these data will contain some errors. The presence of errors can be tested by checking whether the accounts have closed: once they have, invalid accounts can be searched to detect common types of errors. When these errors are understood, they can be automatically corrected. This work will be rewarded if it is done persistently and assiduously. If financial statement data on business taxation are either missing or cannot be closed, they have to be imputed in relation to turnover and number of personnel, which can be derived for all enterprises from the Business Register. With perseverance, the accuracy of imputation can be improved by using increasingly sophisticated methods which take into account such factors as branch of industry and enterprise size.

Since administrative records are originally collected for purposes other than statistics production, the statistical authorities do not have very much say on problems related to timeliness. However it is important that these problems can be tackled more effectively, particularly in statistics on business companies' monthly turnover and payroll costs. When these data first arrive from the tax authorities for a given month, they cover no more than some 70 per cent of the final number of VAT returns, for example. It is difficult to know whether the return for an individual enterprise is missing because it is delayed or because the enterprise has closed down. Therefore Statistics Finland has to request and re-request the use of all data for a given month for six months until they are exhaustive. This does not solve the problem of timeliness, however. In the case of monthly indicators, a first step towards solving this problem was the adoption of the panel method. The quality of these statistics has been further improved by hard work and perseverance in the development of methodologies.

8. *Statistical legislation and data protection*

8.1. *General principles*

National legislation on personal data protection in Finland states that data originally collected for administrative purposes may be released to third parties for purposes of conducting scientific research and compiling official statistics. This principle is also contained in the EU data protection directive, which allows for the linking of data in administrative records for statistical purposes on the grounds that this is not considered to threaten the individual's privacy or rights.

The Finnish Statistics Act (2004) requires that wherever possible, official statistics shall be compiled using existing data sources that have been collected earlier. Among the basic principles of Nordic democracy is a commitment to openness and transparency in government, which includes the right of citizens to know about the actions of government authorities. It follows that much of the information held by the authorities is in the public domain. However, as soon as this information is released to the statistical authority for purposes of compiling statistics, the principle of public access no longer applies. Confidentiality is absolutely paramount to the operation of statistical authorities and essential for obtaining reliable information from citizens and enterprises. Data providers always judge the reliability of the statistical authority as a whole, without being able to make a distinction between information received from public and non-public sources: application of the principle of openness in this connection would therefore have disastrous consequences for the statistical authority. Indeed the Statistics Act says that data received from administrative records for statistical purposes are confidential; the only exception is information describing the operation of central and local government authorities as well as some of the Business Register data. Likewise, the right of citizens to check the personal data held on them in administrative registers does not apply to the statistical files kept by the statistical authority.

The flow of information runs in one direction only, i.e. **from the administrative authorities to the statistical office – never the other way round**. The administrative authorities would certainly benefit from much of the information collected and linked for statistical purposes, but the statistical office must never release the outcomes back into administrative systems. The Statistics Act requires that whenever the statistics can be compiled without identifier data, those data shall not be collected. If it is necessary to link different data sets, that is considered a sufficient reason to collect the data complete with identification codes.

Although legislative obstacles to the statistical use of administrative sources are sometimes justified by reasons of data protection, register-based statistics also offer various **advantages in terms of data protection**. In register-based statistics production the number of people who have access to the statistical data is

much smaller than in traditional data collection. For example, the plain-language questionnaire forms collected in the 1980 population census were spread out for several months in 350 regional offices. Some 2,200 short-term employees were involved in processing these forms, and any information on their neighbours might well have tickled their interest. It is clear that in this situation it is much harder to maintain high standards of data protection than it is with register data, which come to Statistics Finland in machine-readable format and are shown to no one else than the computer. It is much easier to keep account of a magnetic tape, a cassette or a hard disk data file than it is of miles of archives spread out across the country.

Another definite data-protection advantage of register-based census collection is that it means there is no further need for subcontractors. Subcontractors were needed and used in virtually all questionnaire-based population censuses in Finland: after all these involved mailing and recording data from millions of forms within a short space of time.

Inevitably, of course, large-scale register linkages and massive databases do involve certain data-protection risks as well. In Finland, where personal identity codes have been used in population censuses since the 1970s, the changeover to register-based censuses was not really all that dramatic a change in data protection terms. In the register-based data collection system, interview or questionnaire data are simply replaced by register data. Indeed, it is not so much the **contents of the data as the method of data collection** that changes. Having said that, the register system may be somewhat more inclusive in its data contents because it is so easy to draw relevant information from registers. However the change in the method of data collection makes the data system no more sensitive than the questionnaire method. In fact in most cases the opposite is true, for the data are all the time in machine-readable format.

For reasons of improved data security all personal identity data in census files as well as in annual registers are encrypted in connection with archiving. These files also have the highest possible Statistics Finland security classification.

Work to maintain the highest possible level of data security and public confidence is an ongoing effort for every statistical authority. At Statistics Finland, these efforts are most clearly evident in the work of the statistical ethics committee and the data protection working group.

8.2. Releasing data for research purposes

Major register-based statistical files have also facilitated the development of databases for research uses both within Statistics Finland and outside the agency. Each year Statistics Finland releases some 200 research data sets for outside use; in addition a few dozen researchers make use of Statistics Finland's research laboratory to study materials that for reasons of data protection may not be released outside the agency.

The Statistics Act says that all data collected for statistical purposes are confidential. However, confidential information may be released for purposes of scientific research and statistical surveys concerning societal conditions. The data may not be released for administrative, control or other corresponding purposes. Under the main rule, the data may only be released in forms that do not allow direct or indirect identification of persons or the direct identification of other data subjects. In exceptional cases, personal data on age, sex, education and occupation may be released together with identification information.

Decisions on the release of statistical data sets for research purposes are made by the respective Director at Statistics Finland. Difficult cases are submitted to Statistics Finland's statistical ethics committee, which consists of representatives of various departments. Decisions on the release of research materials to foreign countries are made by the Director-General upon consultation of the ethics committee.

9. *Register-based statistical systems: new opportunities for research and statistics production*

9.1. *Longitudinal data files*

9.1.1. *Longitudinal data file on population censuses 1970–2000*

During the period from 1970 to 2000, population censuses have been conducted once every five years. Comparisons of the data from different years has been rather difficult because they are scattered across different publications, and because they have used different classifications. For these reasons the decision was made in the late 1980s to compile all key data in one and the same file and to harmonise classifications and decision rules. This file is known as the **longitudinal data file on population censuses**, which currently covers the data for every person who lived in Finland during the period from 1970 to 2000 from all those censuses in which they were included. Population censuses have been taken on seven occasions, viz. in 1970, 1975, 1980, 1985, 1990, 1995 and 2000.

In addition, the longitudinal file includes data on a 10 per cent sample drawn from the 1950 census. In other words the longitudinal data file comprises all key census data for 30 (or even 50) years, i.e. data on domicile, demographic data, family data, data on workplace, education, incomes, etc. This is a unique data set that can be used to compile both time series and indicators describing population mobility

9.1.2. *Longitudinal data file on employment statistics 1987–2002*

Another longitudinal data file is based on information from register-based employment statistics, covering the period from 1987 to 2002. These data are even more detailed than in the census file and can therefore be used to produce highly accurate analyses and time series. Different groups can by now be followed for periods of up to 15 years. This is an extremely useful source for purposes of studying changes in the labour market, for instance, or changes in employment and housing conditions in connection with moving, as well as exploring the impact of different background variables. Employment statistics have been compiled in exactly the same way since 1987, which greatly facilitates comparisons between different years. The only source of some difficulty in this regard is the new industrial classification that was adopted in 1995.

Statistics Finland publishes reports based on the longitudinal data file. In addition, samples can be drawn from the file for research purposes. It is required that the samples are collected in such a way that individual citizens cannot be identified. The release of data files for research purposes is discussed in more detail under chapter 9.2. above.

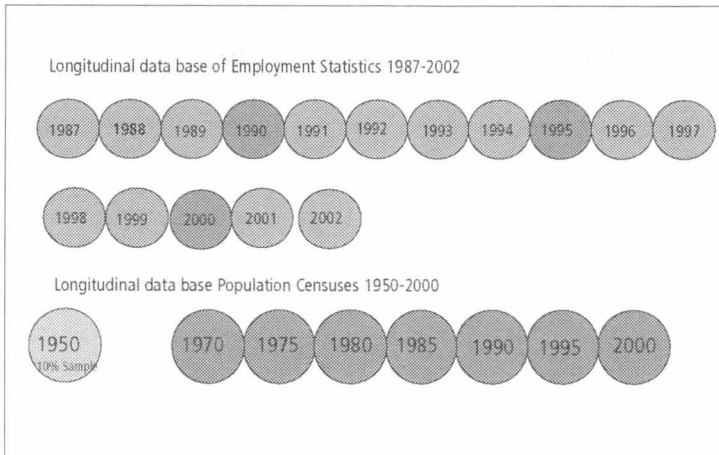


Figure 8.
Longitudinal data
files in Finland

9.1.3. Other longitudinal data files

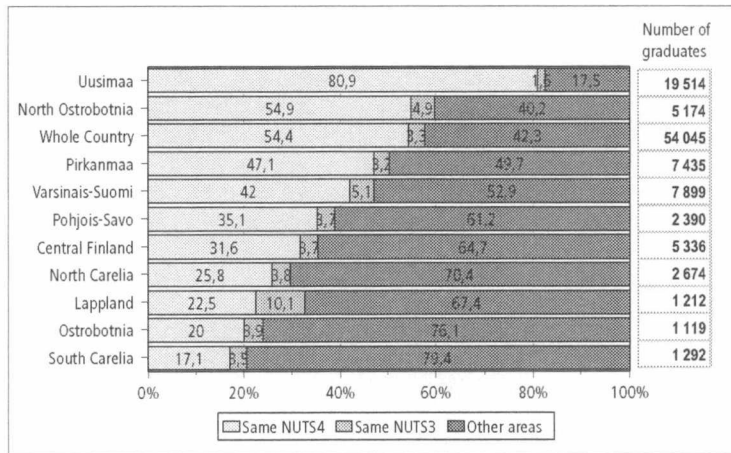
Drawing on these two longitudinal data files, Statistics Finland has in recent years also produced other longitudinal materials for different research purposes. These include databases describing family formation, the impacts of labour policy measures, various panel materials on business enterprises, the combined employee and employer enterprise database, as well as a database on living conditions and causes of death. These are used in the same way as data files from which samples are made available to researchers either outside Statistics Finland or, if this is necessary for reasons of data protection (this applies particularly to business data), at Statistics Finland's research laboratory that has been specially established for this purpose.

9.2. Flow statistics

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.

The statistics that are created using this method are known as **flow statistics**. They describe how people's activities and conditions change over the year: the entry of a graduate into the labour market, the employment of people out of work, the rise in social status, changes of jobs, the educational structure of company staff, staff turnover in companies, the employment of immigrants, etc. In the space of a few years, longitudinal data files have become a mine of information for various lines of inquiry. Flow tables can be used to describe all movements during the year from outside the labour force, for instance.

Figure 9.
Students graduating from university between 1 Jan 1998 and 31 July 2001 by region of study and region of employment as at 31 Dec 2001

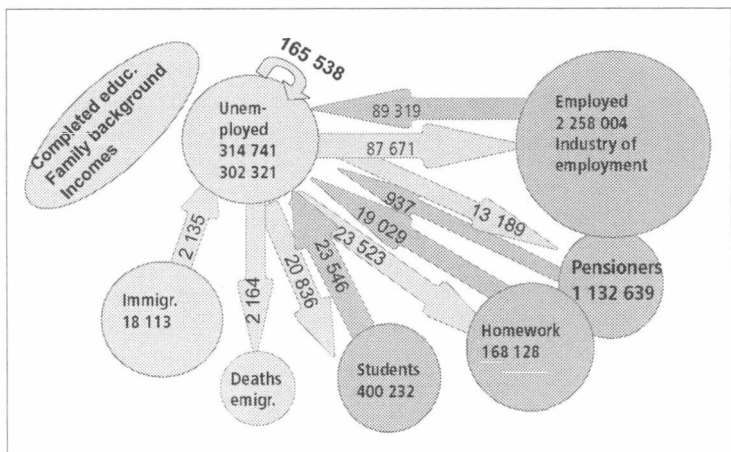


Monitoring statistics on students have expanded and diversified all the time. Every educational institution now receives information on graduate placement and on how different institutions and regions compare with one another. These **placement statistics** provide useful information on employment prospects and future pay levels, as well as on the geographical regions where the prospects of finding a job are the best

Figure 9 shows how large a proportion of university graduates in different regions get a job in that same region. The major universities in the country are in the Helsinki area in the region of Uusimaa, and here 80.9 per cent of graduates found a job in this area; 1.6 per cent were employed elsewhere in the region of Uusimaa; and 17.5 per cent in other regions. In the region of South Karelia, only 17.1 per cent of university graduates found a job in this same region.

The **Student Register** provides a useful tool for monitoring the movement of newly admitted students within the educational system: how they progress from lower to higher levels, how many drop out, how many go to work before graduating, set up a family, etc.

Figure 10.
Flows between different activity groups, unemployed 2001–2002



Longitudinal data files also provide a source of explanatory variables in studies of **mortality differences** or **fertility differences** between population groups. The data on employment history contained in longitudinal files are used in explaining mortality differences between different socio-economic groups.

The register system also includes linkage data between children and parents, allowing researchers to study the impacts of childhood home and parents' living conditions on children's living conditions later in life. Parental social status and educational level strongly predict the future social status of children.

Figure 10 provides an illustration of flows from unemployment to other activities and vice versa between 2001 and 2002.

During 2002, 87,671 unemployed persons found employment, while 89,319 employed persons lost their jobs. The number of immigrants was 18,113, of whom 2,135 were unemployed at the end of the year. In all, 20,836 unemployed persons had started studying and 23,546 had become unemployed job seekers. In addition, the Register of Completed Education and Degrees provides information about the educational level of the unemployed who got employed and how the educational level influences employment possibilities. With the help of such flow statistics decision-makers are able to assess the efficiency of labour market training, for example.

References

- Harala, R. (2004) Future Challenges for Population Censuses in the Light of Finnish Experiences. International Symposium 'The Russians in the Mirror of Statistics: All-Russia Population Census 2002' Moscow 2004 (available at Statistics Finland).
- Harala, R. and Tammilehto-Luode, M. (1999) 'GIS and Register-based Population Census', Statistics, Registries and Science, Experiences from Finland, Published by Statistics Finland 1999.
- Harala, R. and Reinikainen, A-L. (1996) Confidentiality in the use of administrative data sources, *Statistical Journal of the United Nations ECE* (1996) 361–368 IOS Press.
- Harala, R. and Reinikainen, A-L. (1994) 'Statistical confidentiality and the use of statistical data for research purposes – Finnish aspects', Second International Seminar on Statistical Confidentiality, Luxembourg 1994 (available at Statistics Finland).
- Harala, R. (1995) 'Continuous quality check of register-based employment statistics', European Workshop on Using Administrative Data in Population and Housing Censuses, Statistics Finland and Eurostat, Helsinki 1995.
- Harala, R. (1995) 'Evaluation of the results of the 1990 register-based population and housing census in Finland', *Statistical Journal of the United Nations ECE* (1995) 63–72, IOS Press.
- Harala R. (1998) 'Statistical Properties and Quality of Register-based Census Statistics in Finland', Symposium '97, New Directions in Surveys and Censuses, Proceedings, Statistics Canada, Ottawa 1998.
- Jensen, P. (1983) Toward a register-based statistical system – some Danish experience, *Statistical Journal of the United Nations ECE*, 341–365.
- Johansson, S. (1987) Statistics based on administrative records as a substitute or a valid alternative to a population census. International Statistical Institute, Invited papers, Tokyo.
- Korpi, H. (1987) Main Type of Activity and Occupational Status in the 1985 Census: Register-based Parallel Data, CSO of Finland, Studies 152, Helsinki.
- Kotzamanis, B., Cantisani, G., Dekker, A., Logiadu-Didika, D., Duquenne, M-N., Castori, A. (2003) Documentation of The 2000 Round of Population and Housing Censuses in the EU, Efta and Candidate Countries, Eurostat, Luxembourg 2003.
- Laihonen, A. and Myrskylä, P. (1987) Use of Registers and Administrative Records in Population Censuses in Finland, Paper prepared for the European Population Conference, June 11–16, 1987, Jyväskylä, Finland (available at Statistics Finland).
- Laihonen, A. and Myrskylä, P. (1989) The Population and Housing Census: General Plan, The CSO of Finland Reports No.1989:17, Helsinki.
- Myrskylä, P. (1991) Census by Questionnaires – Census by Registers and Administrative Records: Experience of Finland. *Journal of Official Statistics (JOS)*. Vol 7. No 4, 1991 pp. 457–474, Statistics Sweden.
- Myrskylä, P. (1994) Changes in the Labour Force, Special Publication of Population and Housing Census of 1990, Statistics Finland 1994.
- Myrskylä, P. (1995) The New World of Statistical Possibilities by Combining Administrative Data with Census Type of Data Files. Presented at the International Statistical Institute (ISI) conference in Beijing, China 28.08.1995 (available at Statistics Finland).
- Myrskylä, P. (1995) Registers and Administrative Records Used in Production of Census Statistics. Euskas Estatistiska-Erakundea. Instituto Vasco de Estadística. Vitoria-Gasteiz, Spain 1995.

- Myrskylä, P., Taeuber, C. and Knott, J. (1995) Uses of Administrative Records for Statistical Purposes: Finland and United States. Myrskylä, P (Statistics Finland), Cynthia Taeuber (United States Bureau of the Census), Joseph Knott (United States Bureau of the Census). Presented in Arlington, USA, 20–23.3.1995 at Annual Research Conference (available at Statistics Finland).
- Myrskylä, P. and Tarkoma, J. (1996) Employment Flows between Private Sector Enterprises and Establishments in the Helsinki Metropolitan Area in 1989–1993. The 20th Conference on Regional and Urban Statistics ISI/IAOS/SCORUS, Madrid 1996 (available at Statistics Finland).
- Myrskylä, P. & Ruotsalainen, K. (1997) Annual System of Small Area Statistics Based on Administrative Records and Registers. Presented at the International Statistical Institute (ISI) conference in Istanbul, Turkey 1997 (available at Statistics Finland).
- Nordisk statistisk sekretariat (1984) Folk- och bostadsräkningar i Norden 1980/81, Tekniska rapporter 34, Köpenhamn.
- Putkonen, C. (1984) Measurement and Processing Errors, CSO Studies No 99, Helsinki.
- Recommendations for the 2000 Censuses of Population and Housing in the ECE Region, Statistical Standards and Studies – No. 49, jointly prepared by the United Nations Economic Commission for Europe and the Statistical Office of the European Communities, United Nations Publication, Sales No. 98.II.E.5, ISBN 92–1–116685–3, 03500P
- Redfern, P. (1987a) European Experience of Using Administrative Data for Censuses of Population: The policy issues that must be addressed, Paper for the Symposium on Statistical Administrative Data, November 23–25, Ottawa, Canada.
- Redfern, P. (1987b). A study of the future of the Census of population: alternative approaches. Eurostat Theme 3, Series C. European Communities, Luxembourg.
- Redfern, P. (1986). Which Countries Will Follow the Scandinavian Lead in Taking a Register-Based Census of Population? *Journal of Official Statistics*, pp. 415–441, Stockholm.
- Schulte Nordholt E. & Chamberlain J. A New Approach by Combining Different Sources, Statistics Netherlands, 2004.
- Starck, C. (1989). The 1985 Population and Housing Census, Reliability Study, CSO Studies No 157, Helsinki.
- Statistics Finland: Evaluation Study of the 1990 Census, Population Census 1990: Volume 9B Helsinki 1994.
- Statistics Finland (2002). Guidelines on Professional Ethics, Handbooks 30b, Helsinki 2002.
- Statistics Finland (2002). Quality Guidelines for Official Statistics, Handbooks 43b, Helsinki 2002.
- Statistics Finland: Population Census 2000, Handbooks 35c, Helsinki 2001.
- Thygesen, L. (1983) Methodological Problems Connected with a Socio-Demographic Statistical System Based on Administrative Records. International Statistical Institute, Invited Papers, Madrid.
- Turner I. R: Over-arching census study 2002. Report for the Statistics Commission, Version 1:4 Cap Gemini, Ernst&Young.
- Valkonen, T. (1993) Socio-economic Mortality Differences in Finland 1981–1990, Statistics Finland, Population 1993:1
- Wroe D: Beyond 2001. Alternative to the Census, Study for Office for National Statistics, Volume 1, Report, London 1998.

Appendix 1. How certain ID numbers and codes are created

Personal identity code. The personal identity code of a boy born on 17th April 1975 is made up of his date of birth, month of birth, and the last two digits of his year of birth; in addition, all girls and boys born on the same day are numbered consecutively, girls with even and boys with odd numbers.

170475-029F

A check character is added at the end to ensure error-free processing: this is based on the remainder that is obtained on division of the code digit sequence by 31.

If the remainder is a number between 0 and 9, the check character is this number.

If the remainder is 10 or greater, letters are used.

10= A,

11= B etc.

The remainder of the personal identity code above is 15 and the check character therefore F.

Some letters (I, O, Å, Ä, Ö) are not used as check characters to avoid possible confusion.

In addition, a sign is inserted between the date of birth and the number code as follows:

- a plus (+) sign for people born in the 1800s
- a minus (-) sign for people born in the 1900s
- the letter A for people born in the 2000s

The real estate identifier is composed of the code number for the municipality of location (AAA), code for village/town district (BBB), code for estate/quarter (CCCC) and check character (T).

When the numbers of the buildings (DDD) located on the estate are added to the real estate identifier, the resulting chain of characters is called the **building code**.

Furthermore, when the entrance/staircase codes (E) and flat numbers (FFF) and any flat divider codes (F) are added, this gives the **dwelling code**.

The flat number describing each individual dwelling is a consecutive number within the range of 001–799.

All buildings on the same real estate share the same real estate identifier in their code, and all flats on the same real estate share the same real estate and building identifier in their dwelling code.

The dwelling code identifies every **dwelling unit** in Finland.

The code for **business premises** has the same structure, but the consecutive number identifying the premises is within a separate range, i.e. between 800 and 999.

The code structure for summer cottages is identical to that of buildings, but the data on intended use are different than for residential buildings.

The dwelling code is entered not only in the Register of Buildings and Dwellings, but also in the population register for each person living in the dwelling unit: here it is called this person's **domicile code**. For some special groups, including *people living in institutions* and *diplomats*, the flat number in the dwelling code is replaced by a special number. For homeless people the domicile code only indicates their place of residence; no data are included on building or dwelling.

The domicile code thus has the following structure:

049-413-1400-739-M-004-D-008.

In houses that have no more than one dwelling unit, the dwelling code is 000. In terraced houses the staircase code is used to distinguish between different buildings. This facilitates mail delivery, for instance, because the building number is not used in the postal address. The postal address identifies the real estate and the dwelling.

For example, at the Espoo address of **Juhaninkuja 3 D 8, 02800 Espoo**, the numbers in the domicile code correspond to the following information:

Street address/ real estate	Building	Stair-case	Flat	Post code	Muni-cipality
Juhaninkuja 3	–	D	8	2800	Espoo
413-1400-739-M	4	D	8	–	49

Household-dwelling units (which are formed at Statistics Finland) are linked to the register dwelling data by means of the domicile code.

The code system establishes links between building and real estate, between dwelling and building, and between occupant and dwelling; at the same time a chain is formed between occupant, dwelling, residential building and real estate.

For the **population permanently resident in a dwelling**, the code indicates the permanent domicile of each registered individual down to the accuracy of dwelling unit. All persons who share the same domicile code (i.e. who live in the same dwelling) constitute a **household-dwelling unit**. Married or cohabiting couples and single parents living in the same dwelling are formed into **families**.

The Population Register Centre is responsible for issuing and maintaining the codes mentioned above.

Business Identity Code. According to the Business Information Act, a newly established business shall be assigned a Business Identity Code on the basis of its notification of start-up. The code is assigned through the business information system that is jointly maintained by the National Board of Patents and Registration and the Finnish Tax Administration.

The Business Identity Code is formed of a serial number and a control mark in that order. The serial number has at most seven digits.

The control mark is calculated by multiplying the first digit of the serial number from right to left by 2, the second digit by 4, the third digit by 8, the fourth digit by 5, the fifth by 10, the sixth by 9 and the seventh digit by 7 and the products obtained are summed up and the sum is divided by 11. The control mark is determined on the basis of the remainder of the division so that if the remainder is zero, the control mark is zero, and if the remainder is higher than one, the check mark is the difference obtained by deducting the remainder from 11, and if the remainder is one, the serial number corresponding to the control mark is left unused.

Appendix 2.

Use of register sources in Finnish population and housing censuses 1950–2000

Population and housing censuses 1950–2000 in Finland									
	1950	1960	1970	1975	1980	1985	1990	1995	2000
Demographic Data									
Age (Date and Birth)	q	g	q	R	R	R	R	R	R
Sex	q	g	q	R	R	R	R	R	R
Marital Status	q	g	q	R	R	R	R	R	R
Mother Tongue	q	g	q	R	R	R	R	R	R
Citizenship	q	g	—	R	R	R	R	R	R
Religion	q	g	R	—	R	R	R	R	R
Usual Place of Residence	q	g	q	q	q	R	R	R	R
Economic Data									
Main Type of Activity	q	g	q	q	q	qr	R	R	R
Status in Employment	q	g	q	q	q	qr	R	R	R
Industry	q	g	q	q	q	qr	R	R	R
Occupation	q	g	q	q	q	qr	R	R	Rq
Workplace	q	g	q	q	q	qr	R	R	R
Location of workplace	q	g	q	q	q	qr	Rq	Rq	Rq
Socio-economic Group	q	g	q	q	q	qr	R	R	R
Income	—	—	R	R	R	R	R	R	R
Completed Education									
Degree	q	g	q	R	R	R	R	R	R
Subject	q	g	q	R	R	R	R	R	R
Household and Family Data									
Household Type	q	g	q	q	R	R	R	R	R
Household Size	q	g	q	q	R	R	R	R	R
Family Type	q	g	q	q	R	R	R	R	R
Family Size	q	g	q	q	R	R	R	R	R
Dwelling Data									
Size of Dwelling Unit	—	g	q	q	q	R	R	R	R
Number of Rooms	q	g	q	q	q	R	R	R	R
Kitchen	q	g	q	q	q	R	R	R	R
Water, Sewage, Toilet	q	g	q	q	q	R	R	R	R
Heating System	q	g	q	q	q	R	R	R	R
Tenure Status	q	g	q	q	q	R	R	R	R
Business Premises Data									
Floor Area	—	g	q	—	R	—	R	R	—
Tenure Basis	—	g	q	—	R	—	R	R	—
Use of Floor Area	—	g	q	—	R	—	R	R	—
Intended Use	—	g	q	—	R	—	R	R	—
Building Data									
Type of Building	q	g	q	q	q	R	R	R	R
Year of Construction	q	g	q	—	q	R	R	R	R
Construction Material	q	g	q	—	q	R	R	R	R
Main Use of Building	q	g	q	—	q	R	R	R	R
Number of Dwelling Units	q	g	q	—	q	R	R	R	R
Capacity m ³	—	—	q	—	—	—	—	—	—
Heating system	—	g	q	—	q	R	R	R	R
Number of Storeys	q	g	q	—	q	R	R	R	R
Summer Cottage Data									
Owner/Rented	—	—	q	—	q	—	—	—	—
Joint Ownership	—	—	q	—	q	—	—	—	—
Year of Construction	q	—	q	—	q	—	R	R	R
Floor Area	—	—	q	—	q	—	R	R	R
Suitability for Year-Round Use	—	—	q	—	q	—	—	—	—
Fuel	—	—	—	—	q	—	—	—	—
Stove	—	—	q	—	q	—	—	—	—
Original Intended Use	—	—	—	—	q	—	—	—	—
Map Coordinates of Building	—	q	R	R	R	R	R	R	R

q = data obtained from census questionnaires

qr = data for non-respondents obtained from registers

R = data obtained from registers of administrative records

Rq = register data supplemented with questionnaires

— = item not included in census

Appendix 3.

Data sources for regional employment statistics

Register	Data items used	Keeper/Owner of register
Population information system	Population of statistics, population permanently resident in the country on the last day of the year For natural persons: – domicile and address – demographic data – family reference data For buildings, dwellings and summer cottages: – location data (codes and coordinates) – owner and occupant data – attribute data	Population Register Centre
Register of Job Applicants	Information on job seekers, people out of work, persons in subsidised employment programmes and in employment training – periods seeking work – periods out of work – periods in subsidised employment – periods in employment training	Ministry of Labour
Pension registers	Employment and national old age pension recipients – type of pension – pension start date	Central Pension Security Institute and Social Insurance Institution
Study Aid Register	Recipients of study aid – information on study aid granted during each term	Social Security Institution
Joint Selection Register	Entrants to educational institutions through the joint selection procedure – information on applicants and entrants	National Board of Education
Private sector employment register	Private sector wage/salary earners and entrepreneurs – employer's name and pension insurance number – start and end dates of employment (or months of validity) – National Pension Act	Central Pension Security Institute
Central government sector employment register	Wage/salary earners in central government sector – office/unit code – start and end dates of employment – wage/salary data – occupation	State Treasury
Local government sector employment register	Wage/salary earners in local government sector – member organisation and unit code – start and end dates of employment – wage/salary data – occupation	Local Government Pensions Institution
Social Insurance Institution's employment register	Social Insurance Institution – enterprise and location code – start and end dates of employment – enterprise and location code – start and end dates of employment	Social Insurance Institution

Register	Data items used	Keeper/Owner of register
Evangelical Lutheran Church employment register	Evangelical Lutheran parish employees – enterprise and location code – start and end dates of employment	Church Council
Bank of Finland employment register	Bank of Finland employees – enterprise and location code – start and end dates of employment	Bank of Finland
Provincial Government of Åland employment register	Wage/salary earners in the employ of the Provincial Government of Åland – enterprise and location code – start and end dates of employment	Provincial Government of Åland
University of Helsinki employment register	University of Helsinki wage/salary earners – enterprise and location code – start and end dates of employment	University of Helsinki
Employment registers of certain major cities	Wage/salary earners in the employ of certain major cities – office and unit code – start and end dates of employment	Major cities
Register of Conscientious Objectors	Persons in non-military service – start and end dates of service	Ministry of Labour/Educational and Civilian Service Centre
Register of Conscripts	Persons in conscript service – start and end dates of service	Defence Staff
Personal Tax Register	Personal tax files – income subject to state tax itemised by type of income – taxes and other tax-like payments	National Board of Taxes
Customer database	Used in regional employment statistics for the deduction of employer sector for some wage/salary earners	National Board of Taxes
Employers' annual notifications	Wage/salary paid to individual employees by employer – employer's Business Identity Code – wage/salary paid	National Board of Taxes
Wage statistics	Teachers of upper secondary schools and comprehensive schools – member organisation and unit code – validity of employment contract in October – monthly salary – occupation	Statistics Finland
Register of Completed Education and Degrees	Persons completing qualifications – matriculation examination – highest degree completed – date when degree completed and – educational institution	Statistics Finland
Student Register	Students at educational institutions – type of educational institution – educational institution and its location – target degree	Statistics Finland
Register of Enterprises and Establishments	Employer's (enterprise/agency) – data on enterprise and places of business – branch of industry – sector – address/location data	Statistics Finland

Appendix 4.

Data sources on occupational titles

Source	Number of titles drawn from source	Date of occupation	Administrative use
1. Central government employee register	150,000	Year 2000	Registration of employment relationships for pensions, etc.
2. Local government employee register maintained by Local Government Pensions Institution	420,000	2000	Registration of employment relationships for pensions, etc.
3. Social Insurance Institution's employee register	5,000	2000	Registration of employment relationships for pensions, etc.
4. Church Council employee register	17,000	2000	Registration of employment relationships for pensions, etc.
5. Employer organisations' registers	700,000	Last quarter of 2000	Information on employees covered by wage negotiations
6. Provincial Government of Åland and Bank of Finland registers	2,000	2000	Information on employees
	1,100	2000	Information on employees
7. Data inquiry	330,000	2000	Information on occupational titles
8. Ministry of Labour Register of Job Applicants	300,000	2000, registration date	Information on job seekers finding work
9. Taxation and employment pension systems for entrepreneurs and farmers	290,000	2000	Occupational titles
10. PRC population information system (change-of-address notification)	30,000	2000, last removal of the year	Occupational titles

Appendix 5. Business Register and administrative records

Register	Data items used	Keeper/Owner of register
Customer database	Enterprises, new enterprises and organisations, updates for existing enterprises – Business ID, name – postal address, location – municipality of location, language – legal form – branch of industry – accounting period – start-up and closure dates	National Board of Taxes
Business taxation database	Income statement and balance sheet data, including – turnover, wage bill – balance sheet total	National Board of Taxes
Employers' annual notifications	Enterprises and organisations – wage bill by employer/enterprise	National Board of Taxes
Payment Control Register for VAT and employer contributions	Enterprises and organisations (subject to VAT and employers) – sales subject to VAT – monthly wage bill	National Board of Taxes
Trade Register	Business closures, mergers – consolidated financial statements, subsidiaries	National Board of Patents and Registration
Bank of Finland records	Businesses in foreign ownership – country, share of ownership	Bank of Finland
Foreign trade records	Businesses engaging in foreign trade – import or export companies – covered by Intrastat register	Finnish Customs
Register of agencies and offices	Agencies and offices in central government sector – agency/office code – agency/office name	State Treasury
Central government sector employment register	Wage earners in central government sector – personnel numbers in local units	State Treasury
Local government sector unit register	Units in local government sector – member organisation code – local unit code – local unit name – local unit location data	Local Government Pensions institution
Local government sector employment register	Wage earners in local government sector – personnel numbers in local units	Local Government Pensions Institution
Register of Buildings and Dwellings	Buildings (Business premises) – coordinates	Population Register Centre
Address databases	Addresses – basic address file – business addresses	Finland Post
Consolidated financial statements	Consolidated financial statements – consolidated groups	Suomen Asiakastieto Oy
Business database	Business enterprises – telephone number, fax – address	Opasmedia Oy

Tilastokeskus, myyntipalvelu
PL 4C
00022 TILASTOKESKUS
puh. (09) 1734 2011
faksi (09) 1734 2500
myynti@tilastokeskus.fi
www.tilastokeskus.fi

Statistikcentralen, försäljning
PB 4C
00022 STATISTIKCENTRALEN
tfn. (09) 1734 2011
fax (09) 1734 2500
myynti@stat.fi
www.stat.fi

Statistics Finland, Sales Services
P.O.Box 4C
FI-00022 STATISTICS FINLAND
Tel. +358 9 1734 2011
Fax +358 9 1734 2500
myynti@stat.fi
www.stat.fi

ISSN 0355-2063
= Käsikirjoja
ISBN 952-467-333-9

