United Nations Economic Commission for Europe

Register-based statistics in the Nordic countries

Review of best practices with focus on population and social statistics

UNITED NATIONS

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

Register-based statistics in the Nordic countries

Review of best practices with focus on population and social statistics

UNITED NATIONS
NEW YORK AND GENEVA, 2007

NOTE

Symbols of United Nations documents are composed of capital letters combined with figures. Mention of such symbol indicates a reference to a United Nations document.

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

UNITED NATIONS PUBLICATION

Sales No. E.07.II.E.11

ISSN 0069-8458

ISBN 978-92-1-116963-8

Printed at United Nations, Geneva (Switzerland)

CONTENTS

Preface

The Nordic countries have a long tradition in using administrative registers in the production of official statistics. They also have a long tradition in cooperating in the field of statistics. One of their common experiences is that the use of administrative records in censuses is the last step in a process that begins with producing statistics on different subject areas, depending on the type of registers available. By producing statistics on population or employment based on administrative data, they have learned about the influence the actual registers have on the quality of the statistics. After a period of testing and improvement, they realized that the quality of the administrative data was compatible with the quality recommended for censuses, and decided to also use the registers for census purposes. The time it takes from the establishment of an administrative register to having the quality data needed for census purposes may differ from one subject area to another. Nevertheless, the process is similar from subject area to subject area and from country to country. This also seems to be the case for establishing a statistical system based on register information.

In recent years, an increasing number of countries in the UNECE region are considering the possibility of producing statistics based on administrative registers. Therefore, the National Statistical Institutes (NSIs) of the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) decided to share their experience and knowledge with the international statistical community, by producing comprehensive documentation of their best practices. The objective of this volume is to give strategic and planning officers in the NSIs an understanding of what register-based statistics are, covering also the necessary technical and administrative capacity, and the possible applications of these methods to produce official statistics. The emphasis of the publication is on the use of administrative registers to produce demographic and social statistics.

In publishing the present volume, the United Nations Economic Commission for Europe (UNECE) would like to support the Nordic countries in sharing their experience in this field with the international statistical community at large. The volume represents a valuable tool for all NSIs (both within the UNECE region and outside it) that are planning to produce official statistics based on administrative registers. It also supports the implementation of the 2010 round of population and housing censuses since an increasing number of countries outside the Nordic group are planning to use administrative sources to supplement or replace the traditional data collection in the field. Guidelines on how to use registers for population and housing censuses are also included in the new set of "Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing"[1], prepared by UNECE in cooperation with the Statistical Office of the European Communities (EUROSTAT). The new Conference of European Statisticians Recommendations (CES) Census Recommendations contain, for the first time, a detailed description of methodological approaches to censuses including those based on administrative registers, exclusively or in combination with other sources. For each of these approaches, the necessary conditions are highlighted, as well as the relative advantages and disadvantages.

The present report was prepared by a working group established by the Chief statisticians of the Nordic countries. The group included the following members: Johan-Kristian Tønder (Head of the group, Statistics Norway), Finn Spieker (Statistics Denmark), Pekka Myrskylä (Statistics Finland) and Claus-Göran Hjelm (Statistics Sweden). Harald Utne (Statistics Norway) acted as the secretary of the group. Statistics Iceland could not take part in this project.

[1] Available on the UNECE Statistical Division website at: http://www.unece.org/stats/publ.htm

The UNECE would like to thank the National Statistical Institutes of Denmark, Finland, Norway and Sweden that promoted the preparation of the report, and the above-mentioned experts who conducted the work. The UNECE would also like to thank the persons who made useful comments to early drafts of the report, especially experts in the NSIs of Latvia, Netherlands, Slovenia, Switzerland and United Kingdom.

Geneva, 2007

Marek Belka
Executive Secretary
UNECE

About the report

1. Over the years, the Nordic NSIs have presented a number of reports on producing statistics based on administrative sources. The purpose of the present report is to collect all main experiences in *one* document. The experiences in the Nordic countries are very similar, so it is possible to describe some "best practices" common to all countries. However, some examples from single countries are presented, and comparisons between countries are given when relevant.

2. What is true for the Nordic countries is not necessarily true for other countries. The decision on the role of register-based statistics in a country should always be based on national circumstances. We do however hope that a description of the Nordic experiences could be useful in the decision process.

3. The authors have tried to keep the report relatively short. This of course means that a lot of details have been omitted. A list of reports containing supplementary information is given at the end of this report.

4. The present report is based on several reports and papers, published as well as unpublished. However, the following reports may be considered as main sources:

- Statistics on Persons in Denmark – a register-based statistical system (Statistics Denmark, 1995)
- Use of Registers and Administrative Data Sources for Statistical Purposes – Best practices in Statistics Finland (Statistics Finland, 2004)
- Register-based statistics – administrative data for statistical purposes (Wallgren & Wallgren, Statistics Sweden, published in Swedish)
- Reducing Costs of Censuses in Norway Through Use of Administrative Registers (Longva, Thomsen and Severeide, Statistics Norway 1998)

5. The structure of the report is initially based on the Finnish report, but has been modified in accordance with the scope of the present report.

Chapter 1. WHY REGISTER-BASED STATISTICS?

6. All national statistical institutes (NSIs) have a duty to produce official statistics with the highest possible quality with reference to:

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

7. When planning a census or a sample survey, the NSIs have some ideas of quality when defining units and variables. We want to use methods for compiling and editing data and for the presentation of statistics, based on the purposes of the statistics and on the ideals that the statisticians have learned at university or from their colleagues in the NSI. In practice, we have to adapt. We may not be able to interview all persons in a household, and have to accept the answers from one of them (proxy interview). The variable we really want to measure may be so complicated to explain that we have to simplify it, and in spite of all simplifications it may be difficult for some people to answer the questions. In addition, budget restrictions and restrictions on the response burden may reduce a big census to a small one, or a census to a sample survey (or a combination of the two methods). Thorough manual editing may be reduced to a rougher automatic editing. We may be able to estimate the effect of the "short cuts" and compensate for this by imputation at unit level or by corrections in the tables. We must at all times strike a balance between the quality wanted and the practical and economical realities.

8. Administrative data are produced on the basis of some administrative processes, and units and variables are defined out of administrative rules and demands. The definitions may differ from the needs of the official statistics, but the data are usually of good quality for their administrative purposes (more about quality in chapter 9).

9. Statisticians always have to make compromises with their ideals in order to get practical results from the data collection. In a way, the use of administrative data has rearranged the situation: We have the product of a data collection process, and have to compare that product with our quality requirements for the statistics to see if the difference is acceptable. The administrative definitions of the target population may not correspond to our needs (employees, but not self-employed, in employment registers), the variables are not always defined in the way we want (de jure instead of de facto place of residence in the population register), and the time references are not always as precise as we want them to be (the time references of the register may not coincide with the "census day" of the statistics). However as these data are almost free for use for the NSIs, we can use our resources to supplement the information not covered by the administrative data, and make corrections just as we have to do when we are using traditional methods.

10. Summing up the result may be that for a given amount of money the quality requirements are better fulfilled by using data from administrative registers than by traditional data collection by means of questionnaires. Most NSIs seem to have experienced budget cuts, without receiving any priorities or signals on how to reduce the statistical programme or maintain the quality of the statistics. Instead we have observed a growing interest in statistics on new subjects and on new combinations of variables. In many countries there has been a growing demand for statistics on small population groups more frequent than the ten-year-based population censuses.

11. Administrative agencies are professional in formulating transparent rules and procedures and in handling applications, distributing money, collecting taxes or producing other administrative services. The agencies are also under economic pressure. In this situation, the administrative agencies and the NSIs should cooperate. The NSI may offer experiences and capacity in collecting, editing and storing data. Together they might develop a system that is satisfactory for both administrative and statistical purposes.

Example: Norwegian Education Statistics

12. Nearly all education in Norway is regulated by one or more "school acts". One common paragraph in these acts gives the right to the Ministry of Education (or the institution the ministry gives that authority to) to collect data from the educational institutions about pupils, teachers, buildings, costs, income and so on. The Ministry needs this information to govern the school system, and in some cases to control that the money from the government to the school is used as intended.

13. The Ministry, however, is not professional in collecting and editing data and presenting statistics and is therefore cooperating with Statistics Norway. The NSI is, on behalf of the Ministry, allowed to collect data from the schools, both for administrative and statistical purposes. The Ministry may have a copy of the "raw data" because the Ministry is the institution that by the law has the authority to control the individual school. In most cases, however, the Ministry wants the NSI to produce the statistics needed for their governmental purposes. The Ministry can subsequently concentrate its efforts on analysing the statistics and forming policies.

14. Even if the data are collected on the basis of the school act, Statistic Norway has, through the cooperation with the Ministry, the possibility to influence the content and the collection method of data from the schools. This means that the NSI is able to influence the factors that are affecting the quality of the statistics. The difference between the administrative and the statistical data is reduced to almost nothing.

15. These data may be called "administrative data" because they are collected from administrative records in the schools for administrative purposes (governing and control). The NSI produces detailed statistics for the Ministry, as well as official statistics for the whole of society. The official statistics are produced independent of what picture the Ministry may want to give of the school situation, even if the statistics are based on their data. Therefore the statistics are produced in accordance with the UN's Fundamental Principles of Official Statistics.

16. In simple terms, a conclusion may be expressed as follows: Using traditional data collection methods, the NSI is asking: "We asked our question, but do we get the right answer?" Using register-based (administrative) data the NSI is asking: "Surely we got the correct answer, but how does it correspond to our question?" The NSI is working towards the same goal but from a different starting point.

Chapter 2. HISTORICAL DEVELOPMENT OF REGISTER-BASED STATISTICS

17. All the Nordic countries have long-standing experiences 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. It is important to note that administrative data records can never fully replace direct data collection by sample surveys, but that these two methods complement each other.

18. The history of the use of administrative register in statistics in the Nordic countries is briefly illustrated in table below. It is limited to variables used in population and housing censuses.

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

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

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

19. In the period 1964 to 1969, Central Population Registers were established in all Nordic countries, introducing unique personal identification numbers. In the years that followed, several other administrative registers were established. Administrative data as a source for the production of statistics was introduced in the early 1970s. Registers were first used in several subject matter statistics, beginning with population statistics and income statistics. Subsequently, new register-based statistics were developed in all countries.

20. The time elapsing from administrative registers being established to the point when the data are satisfactory for census use may vary from one subject matter to another, and also between countries.

One example is the development of employment statistics. In Denmark, Finland and Sweden, the process lasted for a few years, but in Norway the situation was different.

Example: Development of register-based employment statistics in Norway

21. A Norwegian register of employees was established in 1978. From 1983, small area statistics on employees were published as a supplement to the Labour Force Survey. The statistics did not however cover all relevant units (employed persons) and only comprised a few variables. During the late 1980s and the 1990s, new relevant registers were established, and the scope of the register-based labour market statistics was gradually broadened. In 2001, the data coverage and quality were sufficient for publishing register-based census variables on employment.

22. In 1981, Denmark was the first country in the world to conduct a totally register-based census and Finland followed in 1990. From 1980, the censuses in Norway and Sweden have been partly register-based and these countries are planning for their first register-based censuses in 2011. The main difference between the countries is that Denmark and Finland had dwelling registers in 1980, while the dwelling register in Norway was established in 2001. In Sweden, the plan is to establish a dwelling register in 2008.

23. The step-by-step development is, however, the same in all countries: First, subject matter statistics were tested and published in different areas. Register-based variables were introduced in the census as soon as the quality was considered sufficient. When statistics had been developed for all areas relevant for censuses, a totally register-based census could be conducted.

24. Recent developments in information technology enable the time span between establishing administrative registers and their use in official statistics to be reduced. However, even for registers established today, the process will take some time. Reporting procedures have to be implemented and do not normally work perfectly from the first day, especially when reporting from the public is involved.

Chapter 3. GENERAL PRECONDITIONS

25. In the light of the Nordic experience, there are certain preconditions that facilitate the extensive use of administrative sources in statistics production.

3.1. Legal base

26. Legislation provides a key foundation for the use of administrative data sources for statistical purposes. National legislation must reflect the broadly held view that it makes good sense to take advantage of existing administrative data sources rather than re-collect data for statistical purposes. All Nordic countries have a *national statistics act* that gives the NSI the right to access administrative data on unit level with identification data and to link them with other administrative registers for statistical purposes. Furthermore, the statistics act provides a detailed definition of data protection.

27. In some of the Nordic countries there is an obligation for producers of statistics primarily to use data collected in other contexts. The NSI is obliged first to examine whether the data exists in administrative registers before starting a direct data collection process. All the data must be collected in a manner that is economically efficient and causes minimum inconvenience and incurs least costs to the respondents.

28. In addition, EU regulations are in agreement with statistical acts. *Council Regulation on Community Statistics, article 16* says that in order to reduce the burden on respondents, the national authorities and the Community authority shall have access to administrative data sources, each in the fields of their own public administrations, to the extent that these data are necessary for the production of community statistics[2].

29. All the Nordic countries have an *act on processing of personal data*, which contains provisions on the processing of personal data. The act aims to ensure that the establishment and the use of computerised registers containing personal information are undertaken in such a way that the legal rights of the individual citizen with regard to the protection and integrity of his/her data are not violated.

30. According to this act, processing data for statistical purposes is allowed even if it was not the main aim of the data collection. Once data have been processed in a NSI, they must not be used for purposes other than statistics and research (the principle of "one-way traffic"). A statistical authority may grant access to confidential data for scientific research or statistical surveys. Data collected for statistical purposes are confidential irrespective of the source. The data collected from administrative sources are confidential in the possession of statistical authorities even if these data are public in the possession of administrative authorities. When handling personal data or business data, both direct and indirect identification shall normally be excluded.

3.2. Public approval

31. The existence of more and more administrative registers in society may of course trigger discussions on privacy issues. If the public attitude should become negative, politicians may become reluctant to establish new registers or upgrade existing ones. Statistical use of administrative data normally involves linking data from a number of different registers, which may give the impression that the NSI knows "everything" about every single citizen. ("Big Brother Syndrome")

[2] Council Regulation (EC) No 322/97 of 17 February 1997 on Community statistics.

32. On the other hand, people know very well that administrative authorities are collecting the same data that the NSI uses for statistical purposes. For example, the tax authorities hold information on everybody's employers and earned income, work pension institutes on all the working periods of employees, labour market authorities on all unemployed persons, and pension institutes on old age and other pensioners. In this situation it is very difficult to motivate people to report the same data for statistical purposes.

33. It is 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 administrative data for purposes of statistics production. It is vital that the national register legislation is up-to-date and that the work of the register authorities is open and transparent. Indicators of the situation in a country may be the response rate for official statistics, the politicians' attitudes to proposals on establishing and using registers, and the experiences from the public debate.

34. Over time, the situation has varied between the countries. In Finland and Denmark people seem to believe that the statistical use of register data is rational and they also have a strong faith in the NSI. In the 1970s, there was a big debate in Norway about administrative registers and statistics, but since then the situation seems to have been the same as in Finland and Denmark. In Sweden, the discussion has been more or less ongoing since 1970. However, a common observation in all the countries is that an open discussion and debate, explaining the rationale and benefits of register use, has always been considered a key principle.

35. From the public point of view, using administrative data for statistical purposes has the following advantages:

- Cost efficiency.
- Persons, households and firms can live in peace, they are not harassed with unnecessary inquires.
- Data security is better because of the reduction in the number of persons handling the data, both internally (in the NSI) and externally.
- Only the computer "sees" all the data, because it is all in an electronic format.

36. These advantages seem to be accepted by most citizens as good arguments for statistical use of existing administrative data. However it is important that the statistical agency always remains on its guard in this respect. It is very easy to loose the confidence of the general public, but a major effort to regain it.

3.3. Unified identification systems

37. One major factor that facilitates the statistical use of administrative data records is the use of unified identification systems across different sources. In the absence of such unified systems it is much more difficult to link different registers, which is absolutely central to register-based statistics production. A minimum requirement is to have a unified identification system for base registers.

38. In the Nordic countries, unified personal identity codes (personal identification numbers) are currently present in nearly all registers used in the production of statistics. Similarly, almost equally unified identification systems are in use for other base registers like the business register and the registers on addresses, buildings and dwellings. It may be possible to link different registers even without unified identification codes, but this is certainly more laborious and time consuming.

3.4. Comprehensive and reliable register systems developed for administrative needs

39. The compilation of administrative data registers has been initiated from the needs of the functioning of society and development of administration. It has been closely tied with the development of social security, taxation systems and other administrative needs. These are mostly systems ruled by the state, and therefore it has been necessary to establish registers on a state level. Very often the purposes of the

administrative systems are connected and therefore register information is exchanged between the institutions. Such processes give useful corrections to administrative registers, and hence improve the quality of register-based statistics.

40. Both the citizens and the authorities in the Nordic countries have always shown a high level of confidence in the accuracy of register sources, which after all are the basis for many individual rights and duties. For instance, the official domicile of every individual resident in the country is determined on the basis of register information. Likewise, an extract from the population register serves as a basic document that is needed when applying for a passport, getting married or divorced, or when a funeral is held or an estate is distributed. It is in the interest of the individual to make sure that all the data within administrative register systems are indeed accurate. Good data quality for administrative uses is of course very important also for statistical use, even if "administrative quality" does not automatically imply "statistical quality".

41. One reason for the successful development of administrative registers in the Nordic countries may be that all countries are rather small and homogeneous. In larger and more heterogeneous countries it may be more problematic to establish administrative registers at state level due to technical, organizational and other constraints. But even if a countrywide register system is clearly preferable, local registers may also be useful in the production of statistics, provided that they contain the variables needed, that the definitions of the variables make them comparable for statistics and that the NSI has a legal base to get access to these registers.

3.5. Cooperation among administrative authorities

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

43. The governments in the Nordic countries have given political support to the NSIs' efforts in developing a register-based statistical system. The statistics acts underline such a policy. In practice, this means strong support to the NSIs in their negotiations with administrative authorities on access to administrative data. The collaboration between NSIs and the register authorities often take part in the frame of a national forum of register keepers, but also bilaterally.

Chapter 4. USE OF ADMINISTRATIVE DATA IN PRODUCTION OF STATISTICS

44. In this chapter we compare the two main types of data sources used in the Nordic countries: *Sample surveys* and *register-based data*. Furthermore, we discuss different ways of using administrative data in statistical production.

4.1. Comparing sample surveys and registers

Advantages and disadvantages of the sample survey method

45. The main advantage of survey data is the direct control that can be exercised over the data content, since it is possible to ask questions on precisely those subjects on which information is sought. There is also the fact that data from sample surveys, when modern techniques in the form of Computer Assisted Personal Interviewing (CAPI), Computer Assisted Telephone Interviewing (CATI) etc. are used, in general can be processed and published much more rapidly than register data.

46. On the other hand, survey-based statistics represent a relatively costly method of collecting data. Another problem is data quality. Non-response may cause biases in the statistics, and interviewed households, persons etc. are not always able or willing to give correct answers.

Advantages and disadvantages of the register method

47. The particular advantage of register-based statistics is that they in principle provide total coverage even if collection and processing costs are relatively low. By using registers we can produce more detailed statistics than by using sample surveys, for instance statistics for small areas and for very detailed classifications.

48. Register data offer a large potential because different registers can be linked together on the basis of clearly defined keys (person numbers, business numbers, address codes etc.). Furthermore, data from the individual administrative registers are usually consistent and of high quality, at least for administrative purposes. When register data are used for statistical purposes, and especially when data from several registers are combined, problems concerning consistency and data quality may be revealed. Use of multiple sources is further described in 7.2.

49. The register method may have a parallel to non-response in sample surveys: the under-coverage. An example from Norway: The register of the population’s highest level of education has a growing under-coverage because immigrants are not obliged to give details of their education when entering the country. Every ten years, Statistics Norway therefore has to conduct a special survey on education for immigrants from the last ten years.

50. Another disadvantage of register-based statistics is that statisticians are to a large extent bound by the definitions and administrative practice of the authorities responsible for the registers. Data from administrative sources may be regarded as giving the "authorities’ view of the world". Furthermore, the statistics producer is not very close to the actual data collection, as a result of which he may not know enough about the precise data content, data processing, data quality etc. The best way to handle this problem is by having close cooperation with the register keepers.

4.2. Totally register-based statistics

51. Some statistics are produced by using register data only. Examples include population statistics, various business statistics, income and wealth statistics, education statistics, crime statistics and dwelling

statistics. In some cases, data from other sources, for instance sample surveys, may be used for quality assessments, but not to modify the register-data used in the statistics.

4.3. Combined use of sample surveys and register data

Register data supplemented by data from sample surveys

52. Often the statistical variables required cannot be obtained from one administrative source, but it is necessary to combine data from several registers (use of multiple sources is described in 7.2). One example is the register-based employment statistics where information on jobs is collected from a number of administrative registers. Not all data are comprehensive, and inconsistencies between data sources may occur. Methods are developed to construct a model or an optimal group of decision rules to derive a parameter value that corresponds to the statistical concept in question. For instance, data from the Labour Force Survey may be used to calibrate the model used in the register-based labour market statistics. For the variable *current activity status,* quite a number of data sources are used (for instance in Finland more than 20). Information on the same person is thus available from different administrative sources. In models for selecting the most reliable data source, data from sample surveys may be entered.

53. Imputation for item non-response in register data is most often treated by using supplementary data from other registers. In some cases however, survey data are used in imputation. One example is working hours for the self-employed in the Norwegian regional employment statistics where data from the Labour Force Survey at a macro level is used.

Register data used in sample surveys

Sampling frames

54. Registers are used as sampling frames for surveys. The registers most frequently used are the population register and the business register. For instance the population register contains important demographic data such as age, sex, marital status and place of residence, which are useful in defining a population and drawing a sample. By using data from other registers in addition, samples may be drawn for desired sub-groups, for instance employed persons or students. From business registers, information on for instance industry and location may be used to draw samples for enterprises or establishments.

Improving information

55. Information from registers is used to improve the quality of data from sample surveys. Errors occurring in the data collection can be searched by comparing register data and survey data at micro level (*editing)*. Register data can be used in *coding.* For instance, if the name and address of a person's workplace (establishment) is collected by interviewing, the industry code can be found in the business register. Unit non-response or missing information can be handled by using register data (*imputation*).

56. In *estimation*, register data can be used in calculating weights, for instance by using post stratification. Population distributions derived from the registers, such as age distributions, are thus estimated correctly. Furthermore, the error of the estimates is lowered if the sample is biased with respect to the population, and the reliability of the estimates can be improved considerably if there is a strong correlation between the register data and the sample survey variables.

Additional information

57. Register data is used to supplement the information contents of surveys. For instance, demographic data and data on education in surveys are often collected from registers. One reason for this is of course that it saves time in the interview situation, and time in this context is a direct cost factor. In addition, this method

provides, in a number of cases, more precise information, for instance it is normally easier to obtain reliable information on wages and income from registers than by asking persons themselves.

Non-response studies

58. Plenty of information is available from registers on persons or other survey units refusing to be interviewed or not being reached for various reasons. The reliability of sample surveys can be assessed by means of these kinds of non-response studies.

Statistics combining sample surveys and register data

59. In some cases it is necessary to combine data from administrative registers and sample surveys to produce statistics. One example is income statistics for households. In countries that don't have a household register, survey data on household composition are combined with register data on income. For instance, in the Danish Household Budget Survey and in the Danish part of EU-SILC (statistics on income and living conditions), register data are used to a large extent as a supplement to survey data obtained by interviews carried out in households. The register information used consists mainly of income data together with some information on occupational activity, education and housing.

60. In combining data from registers and sample surveys, *weighting* is one main type of method to obtain consistency at macro level. One example is the Dutch system for population and housing censuses, where such weighting methods have been extensively used.

61. Administrative registers do not always contain all variables relevant for statistical production. In such cases additional information may be collected from sample surveys. *Imputation* at a micro level is then an alternative to combining statistics at macro level. The method consists of importing estimates based on information in relevant sample surveys. Different kinds of models may be used, most of them involving some kind of stratification and random distribution of the variable within each stratum. The outcome of such imputation procedures is mainly valid at macro level.

Chapter 5. REGISTERS - AN OVERVIEW

62. In this chapter, the register concept is discussed and different kinds of registers are described. Furthermore, a description of the administrative data sources used for statistical production in the Nordic countries is given. Of course the registers available are not exactly the same in all countries, but the types of registers are very similar. Some differences between countries are indicated where relevant.

5.1 What is a register?

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

64. As a rule, a register will contain information on a complete group of units, a *target population* (e.g. persons, buildings, firms). These units are defined by a precise set of rules (for instance *resident population in a country*), and the attributes are updated in line with changes undergone by the units. Updating is further discussed in 5.7.

65. The definition of registers given here refers exclusively to the content of the data collections we are dealing with. Information in registers can be stored in a variety of ways. It must be pointed out, however, that the administrative registers used for statistical purposes in the Nordic countries are all computer-based and data are available in a machine-readable form.

66. A key requirement is that each unit in the register can always be uniquely identified. This is best achieved by using a system of *identification codes* (keys), but identification is also possible without such a code if sufficient information on the units is available (for persons: name, address, date of birth etc).

67. Ideally, for statistical use, registers should be as *comprehensive* and *accurate* as possible, but even incomplete registers can be used for statistical purposes.

68. Basically a register may be a list of all units in the population and nothing more. In practice most registers also comprise some additional attributes for each unit (data). The term *data register* is used to separate registers with data from the mere lists of units. In this report the term *register* is used in both senses.

69. *Administrative registers* are registers primarily used in administrative information systems. That means that the registers are used in the production of goods and services in public or private institutions or companies, or that the information is a result of such production.

70. In the Nordic countries, most administrative registers used for statistical purposes are countrywide registers operated by the state or jointly by local authorities. However, private registers are also used, for instance registers operated by insurance companies and employer organizations.

71. *Statistical registers* are created by processing data from administrative registers. Statistical registers could be based on a single administrative register, but they are more frequently based on combined data from several administrative sources.

72. Data files produced by traditional censuses and other total enumerations are not registers. Such *statistical data files* are not subjects of an updating process; i.e. the intention is not to update the population of units or the information about the units when changes occur. At every reference point of time or reference period (census day or census year), all information has to be collected. But when cross-sectional data files are established, whether based on register data or traditional census data, the way of producing statistics is quite similar.

73. As part of a register-based statistical system, administrative registers are sometimes called *primary registers*. However, some primary registers are in fact statistical registers kept by the NSIs; the main example being registers on education. In this case, no central register is kept for administrative purposes. Statistical registers are also referred to as *secondary* registers.

74. The set-up of register-based statistical systems is described in more detail in chapter 7.

5.2. Base registers and specialized registers

75. Administrative registers are subdivided, according to their purpose, into *base registers* and *specialized registers*. The term "base registers" is also used to describe statistical registers.

Base registers

76. *Administrative base registers* are kept as a basic resource for public administration. The function of these registers is typically to keep stock of the population, i.e. to record the birth and death of units and to keep track of what units are present at any given time in the population. In addition, base registers have to maintain identification information to be used by other administrative registers.

77. *Statistical base registers* are based on the corresponding administrative registers. These are registers of great importance for the whole register system. Their principle tasks are to define important populations and contain links to other base registers.

78. In practice, an administrative base register will contain more information than only identification of the units, typically data of common interest to a number of administrations. A Central Population Register (CPR), for instance, will have information for each person on sex, day of birth, marital status, children, etc.

79. Base registers must be able to provide data for different populations. For instance, the main population of the CPR is persons resident in a country. The CPR should, however, also give information on people working in the country but living elsewhere. Further, in the Nordic countries we have experienced that even units that are no longer active should be kept in the base registers for historical reasons. This means that the CPR should also comprise of dead persons and emigrants.

80. In all Nordic countries, three administrative base registers are defined:
- Register on persons (population register)
- Register on properties (real estate, buildings and dwellings)
- Business register

81. Correspondingly, there are three statistical base registers in all countries. In Sweden, a fourth statistical base register is defined: The activity register. Registers on activities (jobs, other labour market-related activities and educational activities) do exist in the other countries as well, but they are not defined as base registers, (See 5.6).

82. Administrative base registers are organized in different ways among the countries. For instance, in Finland, data on dwellings and buildings are a part of the population information system, but in Norway and Denmark these data are included in the same information system as real estate and addresses.

83. Although important in the daily work, these differences do not really affect the main set-up of the register-based statistical systems. The reason for this is that all base registers in the Nordic countries have linking keys to related registers and also to the other base registers. The linking keys are:

- Personal identification numbers
- Numerical addresses
- Business identification numbers (for establishments and enterprises)

Specialized registers

84. Unlike administrative base registers, specialized administrative registers serve one specific purpose or a clearly defined group of purposes. The authority maintaining the registers is normally the main user of the information and the data contents are naturally influenced by this fact. Specialized registers often receive information on the population and some basic data from a base register, but supply other data themselves.

85. *In* the Nordic counties, almost all specialized registers use the same identification key as the corresponding base register. In this way, linking is normally quite straightforward. However, in a few cases it is not obvious what base register should be regarded as the corresponding one. The register of vehicles is one example. Vehicles could actually have been defined as a basic unit, but the register is not considered important enough to be classified as a base register. So vehicles owned by private persons are linked to the population register and company-owned vehicles are linked to the business register. The register of vehicles is then regarded as a register on persons or a business register depending on the context.

86. As for base registers, specialized statistical registers are established based on one or several administrative registers. The registers listed in 5.3-5.6 are administrative registers, but corresponding statistical registers do exist.

5.3. Registers on persons

Base register

87. All the Nordic countries have a Central Population Register (CPR) with a unique identifier (personal identification number). The register contains some basic information on all persons (sex, age, martial status, nationality etc.). Further, references to parents and spouses are included, making it possible to establish the *family* unit. Another important variable is the dwelling address. In countries with a dwelling register[3], the *household* (dwelling-household) unit is established by linking persons and dwellings.

Other registers on persons

88. There are numerous administrative registers used for statistical purposes in all Nordic countries, all using the personal identification number from the CPR. Registers are to be found in almost all areas and *include* the following:

- Taxation registers
- Income and wealth
- Wages and salaries
- Labour market registers
- Jobs

[3] In 2006, all countries except Sweden, where a dwelling register is being planned.

- Unemployment, labour market training
- Register on educational attainment
- Student registers
- Social security registers
- Health registers
- Criminal registers
- Registers on vehicles

89. *These* are the most important registers on persons, but the list is by no means exhaustive.

90. *Labour* market registers and student registers may alternatively be classified as *activity registers,* (see 5.6).

5.4. Property registration

91. *Administrative* registers on properties are organized somewhat differently among the Nordic countries, but there are still many similarities. Denmark and Finland have dwelling registers that have been in operation for a number of years and accordingly used for statistical purposes. In Norway, a total dwelling register was established in 2001 and register-based dwelling statistics are being developed. In Sweden, a dwelling register is currently (2006) being planned.

Registers on real estate

92. These registers cover real properties and parcels of land and include details on rights and mortgages attached to the property. In addition, they contain information on the location of the properties as well as the identification of the owners.

Registers on buildings and dwellings

93. Registers on buildings and dwellings may be regarded as extensions of the property registers. The *information* covers the age, size, installations and use of buildings and dwellings. The registers are updated with reports on alterations, new buildings and demolitions.

Address registers

94. These registers contain geographical characteristics attached to all addresses, for instance what administrative units (municipality etc.), as well as other geographical regions, the address belongs to, and if the address is located in a densely or sparsely populated area. Furthermore, information on the geographical coordinates of the address is included.

Registers are interlinked

95. Registers on real estate, buildings and dwellings and addresses are *interlinked*. In this way it is *possible* to obtain data on the exact location, and thereby geographical details, of buildings and dwellings.

96. Numerical address is the linking key between the *population register* and the register of buildings and dwellings as well as between the *business register* and the registers of buildings. In this way the exact *location* of place of residence and place of work is obtained.

5.5. Business registers

Base register

97. All Nordic countries have developed central business registers covering units in all kinds of activities, *including* central and local government units. The main types of units are enterprises and establishments, and the register holds a unique identification number for these units. These identification numbers are used in most administrative business registers.

98. The central business registers contain basic information of the units: type of ownership, identification of owner(s) where applicable, industry, location and size.

99. There have been substantial improvements in the system for business registration in recent years. We will use the business registers in Norway as an example.

Example: Business registration in Norway

100. Until the mid 1990s, there were several business registers in operation. These registers were not coordinated with regard to units, identification numbers and populations, and were updated separately. In 1995, The Central Coordinating Register of Legal Units was established. This register collects all basic information in one data system so that all units are assigned an identity number in the same series, regardless of their objectives. This is similar to the personal identification numbers assigned by the Central Population Register. The units required by law to register with The Central Coordinating Register of Legal Entities are those registered in the Register of Employers, the Value Added Tax (VAT) Register, the Register of Companies, the County Governors' Register of Foundations and Statistics Norway's Register of Establishments and Enterprises.

101. The use of unified organization numbers reduce the obligation to report data as well as registration work since the individual unit will have to report each event to one register only. It is subsequently the register receiving the information that is responsible for transmitting the report to The Central Coordinating Register of Legal Entities for updating.

Other business registers

102. Several administrative business registers are available in all Nordic countries, the most important being registers maintained by the tax authorities (for instance business taxation registers, registers on value added tax and employers' contributions and employers' annual control data). Another major data source is trade registers. There are also a number of registers for sub-populations of the business register, for instance school registers.

103. The existence of a unique identification number facilitates the linking of business data from different registers. For instance, turnover may be collected from the VAT register and the number of employees from registers of employees and wage registers.

5.6. Activity registers

104. The main types of activity registers are *job registers*. There is no single administrative register covering all kinds of jobs, so statistical job registers are dependent on different administrative sources in different countries. The main types are taxation registers, different registers on employees and registers on employment pension systems. Additional information for special groups of employees may be collected from

other administrative registers, for instance persons employed by central or local government, teachers and conscripts.

105. The unit of job has two identification keys, the business number (for the establishment or enterprise) and the personal number (for the employed persons). Characteristics of the employed person (sex, age, place of living, education etc.) and the establishment (industry, location etc.) are in this way added to the job register.

106. Employment statistics are usually based on the unit of person. The main job from the statistical activity register (for a point of time or period) is then chosen to characterise the employed person.

107. Job registers are also used in business statistics. Establishments and enterprises are characterized by the number of employees and other details of the staff (for instance sex, age and educational distributions).

108. *Registers on unemployment spells and labour market training* and *student registers* are other types of activity registers. For students, the registers contain the personal identification number as well as the business number of the school and in addition information on the studies (for instance type of education and information on part-time/full-time studies).

109. Statistical activity registers are considered as *base registers* in Sweden only, but are of great importance in all Nordic countries.

Example: Four statistical base registers in Sweden

110. Statistics Sweden differentiates between the three administrative base registers and the four statistical base registers. The population of job activities and other related activities is regarded as an important target population that should be analysed by, for instance, demographic methods. This, and the fact that quite a lot of administrative data is related to activities, is the reason why the Activity Register is considered as a statistical base register in Sweden.

5.7. Updating of registers

111. The basic idea of registers is that they are 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. new-borns and immigrants in the CPR) while those exiting the target population are "removed" from the register (or more precisely classified as "not active"). The data describing the units may also change and therefore have to be updated as well. For persons, changes may be recorded in such details as marital status, number of children, educational level, occupation, etc. Likewise, the floor area of a building may change; the owner of a motor vehicle may change, etc.

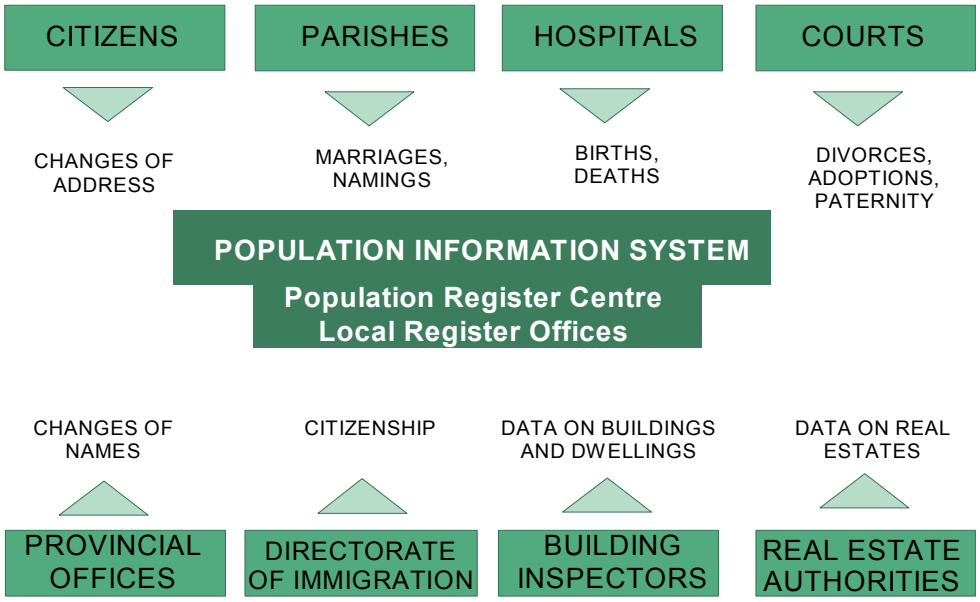
112. Updating can take place with a greater or lesser time lag with respect to the event that gave rise to the change. In some cases, reporting may not coincide directly with the event, so ascertaining the correct data at certain time intervals must suffice. Data that are not capable of systematic updating are not regarded as true register data.

113. A key principle of keeping registers is that all relevant data are collected when they arise and on this one occasion only. This is effective both from an economical point of view and to prevent redundancy in the register system. Indeed the greatest advantage of a good register system is that the need for processing is restricted to those units and those attributes that actually have *changed*. Units and attributes that remain unchanged are not processed at all. The more stable the data contained in a register, the less work involved in register maintenance.

Example: Procedure for updating the Finnish population information system

114. The Population information system 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.

115. As can be seen from the figure, the responsibility for updating the population information system rests mainly with different authorities. All that remains for individual citizens is to notify changes of address, which must be done within seven days of moving through one of the country's 37 local register offices. 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.



Chapter 6. REQUIREMENTS TO DATA FROM ADMINISTRATIVE SOURCES

116. The general preconditions for using administrative data for statistical purposes are discussed in chapter 3, and the use of administrative data and the combined use of registers and different surveys are described in chapter 4. This chapter describes in more detail some demands that must be met if official statistics are to be successfully based on administrative registers.

Contents

117. First and foremost, administrative registers must contain data covering the most important subject areas in a statistical system for elucidating patterns and trends in society. An important precondition for statistics based on registers giving comprehensive coverage, is that the data contained in administrative registers should be extensive and should cover many variables relating to the relevant units. Gaps either necessitate the supplementary collection of information using traditional methods or limit the content of the statistics.

Units and identifiers

118. Three central units are essential to the structuring of the statistics: persons, enterprises/establishments and dwellings. This is reflected by the set-up of the corresponding base registers discussed in chapter 5. There are of course many other units for which it is important to have records. Some of these units are *events*, which relate to persons, for example demographic events (births, deaths, marriages etc.) .

119. For statistical use, it is important that the units are well defined. This is not a problem for "natural" units such as persons or motor vehicles. Also, the concept of a home or dwelling is fairly easy to deal with, although there are some borderline cases in which the existence or specification of a home can only be determined on the basis of detailed rules. Some units manifest themselves as combinations of natural units, e.g. families and households.

120. It is more problematic when other types of units are considered, and here business units are a good example. A firm or a workplace is an abstract concept. Its existence can only be established and its nature defined in accordance with very precise rules. For instance, from administrative data it may be difficult to determine whether a new enterprise has been established or if an existing one has undergone some major changes.

121. Registers should, ideally, contain information on all units in the country of the types specified, i.e. they must be comprehensive. Incomplete coverage will often be systematic, for example, if certain municipalities are not covered, the statistics produced on such a basis will immediately be affected by biases that it may be difficult to compensate for.

122. *Identifiers* play a very considerable role both in the maintenance of administrative registers and in their statistical use, particularly in linking information from various sources. Identifying codes should ideally not be changed for the period a unit exists. This is best achieved by using codes containing no information. When a personal identification number contains the date of birth and sex, there may be changes during the lifetime. The most frequent reason is that the date of birth is being corrected in the CPR, mainly for emigrants. Problems may also arise when a person arriving in the country is initially given a temporary identification number and subsequently a permanent personal identification number. In such cases, the NSI has to establish links between identification codes that refer to the same person. One way of solving this problem is to replace the administrative identification number by a statistical identifier in statistical registers.

Time references

123. The time dimension plays a very important role in statistics, revealing patterns and trends in society, and in all areas it is necessary to be able to make comparisons over time. It is therefore vital for statistical usability that reliable time references are contained in registers.

124. The most important is the *dates of changes or events*. Among the main events of interest are the "birth" and "death" of units, but it is also important to date other changes affecting units. What we are concerned with here is the real point in time at which an event took place bringing about a change in a data item, for instance the date of a removal or the date of a change of industry for a business enterprise.

125. In practice it is often difficult to pinpoint the real date of an event. The industry within which an enterprise operates is a variable that may change gradually, as the activities of the enterprise shift their emphasis, and it is perhaps not possible to determine exactly when a new activity becomes dominant. However, even for variables that in principle can be observed at any given time, and in which a change can therefore be dated with reasonable accuracy, problems arise in practice in the recording process. This may be due to the fact that the date of an event is not necessarily important to the administration using the data: all it needs is to know the most up-to-date value at any given time. In other cases problems may be due to the fact that it is not possible or practicable to record events in real time. It may only be possible to ascertain data changes from responses to enquiries, for example once a year in conjunction with a tax return.

126. In addition to dates of events there is a need for *registration dates*, i.e. an indication of when the data value in question was entered in the register. The ideal situation therefore is that any item of information in the administrative register should be accompanied by two dates. Registers in reality often deviate significantly from this ideal.

127. If it is not possible to record dates with accuracy, approximations of course remain an option and are much better than nothing at all.

Data on events

128. In some cases there is a requirement that it should be possible to view the data changes recorded in the register as a model of events, which are described in the statistics. This applies for example to population statistics, in which it should be possible to derive reports of events in the population (removals, marriages etc.) from data changes in the Central Population Register (CPR). For the purposes of migration statistics for instance, it must be possible to distinguish real events (removals) in the register from corrections and data changes of a more technical nature.

129. Actual corrections of wrongly reported data however pose a problem for statistics, even when it is possible to distinguish them from real events. Corrections have to be regarded in many cases as modifications of events reported earlier: an address correction may be a correction to a report on a removal. A correction may also cancel an event reported earlier, which occurs in the CPR system in particular when a person, having notified the population register that he is to emigrate, changes his mind and remains in the country.

Stability

130. An important characteristic of statistics is to describe a process over time, i.e. to show how a particular magnitude develops from month to month and from year to year. It is of great importance, therefore, that concepts in the administrative registers remain constant over the longer term. Otherwise major problems can arise in securing comparable figures from one period to the next. In some cases, but far from all, it is possible to adjust for changes with greater or lesser precision. Paradoxically, a problem may arise even if the registers become more reliable, for even then we get data discontinuity. Major problems may be

posed for statistics when legislative or regulatory changes result in alterations to the data content of administrative registers. On the one hand, it may be difficult or impossible to assess the long-term trend in a particular magnitude if different definitions are used in the base material. On the other hand, problems may arise in deciding what changes in data values are to be viewed as reflecting actual events and what changes merely represent new concepts or definitions.

131. The statistical consequences depend on what type of statistics is involved. If it is in fact *statistics for monitoring of legislation*, the function of which is to show how the administration of a law affects ordinary people, the statistics of course merely have to go along with and adopt the concepts of the new legislation. *General statistics*, on the other hand, seek to elucidate certain concepts that are not defined in legislation, e.g. unemployment, a concept whose definition is to be found in an international convention. If the statistics relate to the payment of social benefits, and the rules for these are changed, it may be difficult or impossible to compensate for the change in the statistics. An attempt must at least be made to estimate the significance of the changes, so that time series can to some extent be chained together.

132. There are also instances of administrative registers changing without this being due to any changes in the basic rules. Typically these are changes introduced for technical reasons or to achieve rationalization gains. A problem may arise, for example, if it is decided that a particular item of data is no longer necessary for administrative purposes.

133. Clearly the changes described are serious for the statistics if they are undertaken without sufficient consideration for its requirements. It is therefore important that statisticians are involved in the preparation of the changes so that at least their consequences can be assessed before they are introduced.

Quality

134. The quality requirements imposed by statistical use coincide to some extent with the requirements that must also be met in serving the primary purpose of the registers; the information must be reliable and recorded with sufficient precision. In addition, data must be relevant to the statistics. Quality aspects are discussed in chapter 9.

Cooperation with register-keepers

135. In producing statistics based on administrative data it is not possible to exercise the same control over the content of basic data as in the production of questionnaire-based statistics. We cannot be sure that the registers cover the units of relevance with the same degree of precision or that data are defined in accordance with the needs of users.

136. It is of course desirable for statisticians to exert a certain influence on data content, but it must not be forgotten that registers are kept for quite specific administrative purposes and that the task of the register keepers is to serve those purposes in the best way possible. They cannot therefore pay too much attention to demands that "only" serve statistical purposes. For this reason, statisticians should not expect to make very extensive demands for additional data, different definitions and the like, and to have those demands met. Register keepers must pay strict attention to efficiency in their own operations and use of resources. Statisticians must therefore be modest and only put forward major requests when an adjustment to a register can be expected to yield very substantial benefits to society. Very few instances of extra data being collected by way of registers exclusively for the purpose of statistics are known in the Nordic countries.

137. A national statistics act is a useful instrument for good cooperation with register owners, (see 3.1).

Chapter 7. FROM ADMINISTRATIVE DATA TO STATISTICS - DATA PROCESSING

138. Before administrative data can be used for statistics they must undergo statistical data processing, which will render them suitable for the purpose. In this process, data are organized into *statistical registers*, which are data compilations with identification keys in which the population and the data content of each register are correlated so that they meet the needs of a particular statistical area. Data in a statistical register may be drawn from one or more administrative registers.

7.1. Editing

139. One of the most important and resource-consuming processes in the creation of registers is the verification of the basic data entered and the correction of errors, also called editing. Normally data are first controlled by the authorities in charge of the administrative register and later by the statistical office.

140. The purpose of the editing done by the register keepers is to improve data quality for administrative use. This means that the focus is on the variables that are most important for the work of the administration. Variables that are included in the register mainly for statistical purposes may not be controlled as carefully. It is vital that statisticians have good knowledge of these editing procedures and their limitations.

141. In the editing done by the NSI, the focus is on the statistical use of data. Normally each register is first controlled by using internal logical checks to remove register errors. No external data are used in this process. Coverage problems and other register errors may also be revealed by combining data from several registers (more about use of multiple sources in 7.2).

142. The statistical register should ideally be “clean” and consistent, so that there are no conflicts between individual data. Improbable information must be investigated. In this process it is not unusual to discover system errors in the administrative registers or in the extracts the data supplier makes before the statistical office receives them.

143. Errors may consist of a failure in the information to meet the *formal requirements* for the information type in question. In such cases, where errors often arise in groups, it must be decided whether to change the value in question so that the information coincides with the permitted intervals or whether to exclude the information from the statistical basis, since no meaningful correction can be undertaken.

144. Isolated errors can in certain cases be detected and corrected by direct reference to the source registers. Registers cover the total population, so the possibilities for contacting individual units to collect information are very limited. This is indeed the case when the units are individual persons. When we are dealing with business registers it is possible to collect additional information by contacting individual enterprises, but this is a very resource-consuming process. The main method for control and corrections is therefore computerised, logical editing.

145. Time references must be processed in order to provide the best possible statistical model of the unit population and its data at the desired points in time or in the desired periods. Administrative registers do not normally conform to such a model, since the main purpose normally is to provide the most updated data on each area.

146. It is crucial that any systematic types of error occurring in the administrative registers from which data are extracted for statistical purposes should be corrected. This can only be achieved by close contact with the authorities responsible for the register. There have been many instances in which error types

discovered by statisticians were subsequently corrected in administrative registers, to the benefit of both the register and the statistics.

7.2. Use of multiple sources

147. When data are collected for statistical purposes using traditional collection methods, the questions and guidelines for answering them are formulated in such a way that information is obtained that accords with the definitions used in the statistical area concerned.

148. Using administrative registers as a basis for statistical data does not provide the same scope for control. It is possible to search the information available and select those items that best suit the intended purpose. It will often be the case, however, that even the most suitable information does not cover the immediate need to a sufficient degree. This may be due to divergent definitions, limited coverage, errors or inadequate updating. In that situation, one solution is to draw on information from several administrative sources. The creation of statistical registers therefore often involves the linkage of basic data at unit level. On the basis of the various data, an attempt must be made to find the best estimators for the statistical variable.

149. Where the need for a specific item of information for one or more statistical functions cannot be covered immediately by direct reference to the information available, the problem may be solved by constructing *derived variables*. Such a variable is obtained by chaining together different items of information that provide the information required, if not with absolute precision at least with good approximation.

150. In setting up a derived variable, the starting point is to establish a definition of the concept that it is desirable to use in the statistics. Administrative information that can contribute to meeting the definitional requirements is then sought. A distinction can be made between three types of variable formation: *creation of new variables*, *prioritisation of information for a given variable* and *adjustment or correction of information*.

Creation of new variables

151. The creation of a new variable may be the solution if the necessary information is available to a sufficient degree but is not contained in a single item of administrative data. Several items of information, collected and processed in accordance with administrative rules, can each contribute to providing an overall description of single individuals in relation to the desired statistical concept.

152. When a new variable is formed, it is essential that the necessary components are present for all units to a high degree of quality in terms of the intended purpose.

Examples of derived variables in Danish statistics

153. *Example 1:* Information on business ownership can only be found for persons who are included in the Central Business Register. These are employers and persons subject to VAT. A small group of business owners are not contained in this population, for example persons operating within the field of passenger transport and providers of certain health services. In this case, information gathered pursuant to the obligation to submit accounts is used, and greater significance is attributed to profits or losses recorded by the business.

154. *Example 2*: The separation of wage and salary income from other types of income taxed in the same way as wages, is somewhat influenced by the fact that the distinction is of no great relevance to the assessment work of the tax authorities. It may be concluded however that, despite the deficiencies indicated, the result concurs well with the definition laid down.

Prioritization of information from multiple sources

155. There are situations in which the desired information is to be found, but in different administrative registers with variable coverage and/or of uneven quality. Variable coverage is due to the fact that certain registers are set up to perform particular administrative functions relating to a less well defined section of the population. The quality of the information depends on its relevance to the administrative purpose it is to serve.

156. An example is the variable *current activity status* (employed, unemployed, not in the labour force) that could not be collected from a single register. For persons having more than one activity at the same time, employment shall, according to the statistical definition, have priority. However, if information on unemployment is regarded to be more reliable, at least where time references are concerned, then persons registered with both employment and unemployment at the same time should be classified as unemployed.

Correction of information

157. Correction of register information is undertaken if possible where errors or deficiencies arise. Such corrections can to a certain extent be carried out on the basis of expected links between items of information from the same source, but it may be necessary to use information from other sources.

158. Correction on the basis of supplementary information from other sources may provide usable information in areas where the available material is prone to error or is deficient in other respects. It may however be dangerous to undertake systematic corrections if limitations of the method applied are not considered carefully.

Use of combined data over time

159. The use of combined data may often require a very complex set of rules for the processing of the information. It may be difficult to maintain a complete view of what is really happening and this may result in insufficient attention being paid to the significance of individual elements. A change in the basic information which, seen in isolation, is relatively insignificant can have very unfortunate consequences in the processing of the combined information.

160. It is important that the set of rules for processing is described in detail with a precise indication of what information is to be used, how it is to be used and what conditions apply to its use. Furthermore, it is important to allow for possible changes in the content of the information so that the rules can continue to be effective if such changes occur.

161. It may be difficult to obtain prior knowledge of all relevant changes, so it is essential to test the processing as it proceeds, for example by checking that the various conditions affect the final result to the same extent as previously. Deviations may not in themselves indicate that something has changed, but they do indicate that there may be a need for closer investigation.

162. The use of derived variables formed on the basis of information from different sources will be particularly vulnerable to administrative changes. The use of multiple sources will increase the risk that changes of relevance to the statistical purpose may arise. The changes may be known, but it is not always possible to allow for them, for example if an item of information has been omitted. In this case it may be necessary to adjust the conceptual definition in use in order to render it compatible with the prevailing circumstances. That means departing from the definition previously used and the result will be data discontinuity.

163. Administrative changes may also be positive in the sense that new information arises or existing information is improved in the way it meets statistical requirements. Changes for the better will of course

also give rise to data discontinuity, but the gain from the improved basis may more than compensate for the inconvenience of the loss of continuity.

Presentation of statistics based on multiple sources

164. The actual definitions associated with the derived concepts may often be somewhat complex since they reflect the set of rules applied. This may create unnecessary confusion in the minds of the users of the statistics, which is why they appear in a more simplified form in their presentation, as is also the case with traditional data collection. When information from multiple sources is used, a more differentiated picture can be obtained. For example, a tax register may provide information on wages earned by a particular individual and the Business Register may show that the same person has been active as a business owner. The set of rules must determine how the information should be weighted, whether such a person should be classified in terms of employment status as an employee or as a self-employed person. For presentation purposes, the register-based employment status of a person is defined as that person's main activity. Further details can of course be made available if special investigations call for an analysis of the demarcations between different groups.

7.3. Developing a register-based statistical system

165. A system of register-based information has been developed in all Nordic countries. This means that all statistical registers, in principle, are regarded as being a part of the same system and not as single registers. Such a view has implications for all stages of statistical production: data collection, data processing, quality control and dissemination. The cornerstones are the existence of base registers keeping track of the target populations and the fact that all major registers can be interlinked. In this way, different subject matter statistics may be partly based on the same data sources. For instance, demographic variables are produced in the statistical population register and are used in all kinds of social statistics.

166. In practice, this does not mean that all variables are linked with all other variables available in the statistical system. When specialized registers are used for statistical purposes, they are always linked to the corresponding base register(s). Linking or not between specialized registers, is a matter of appropriateness. For instance, if variables will never be published or otherwise used in combination, there is no need for linking. In general, however, multiple use of register data will improve the quality of all data involved.

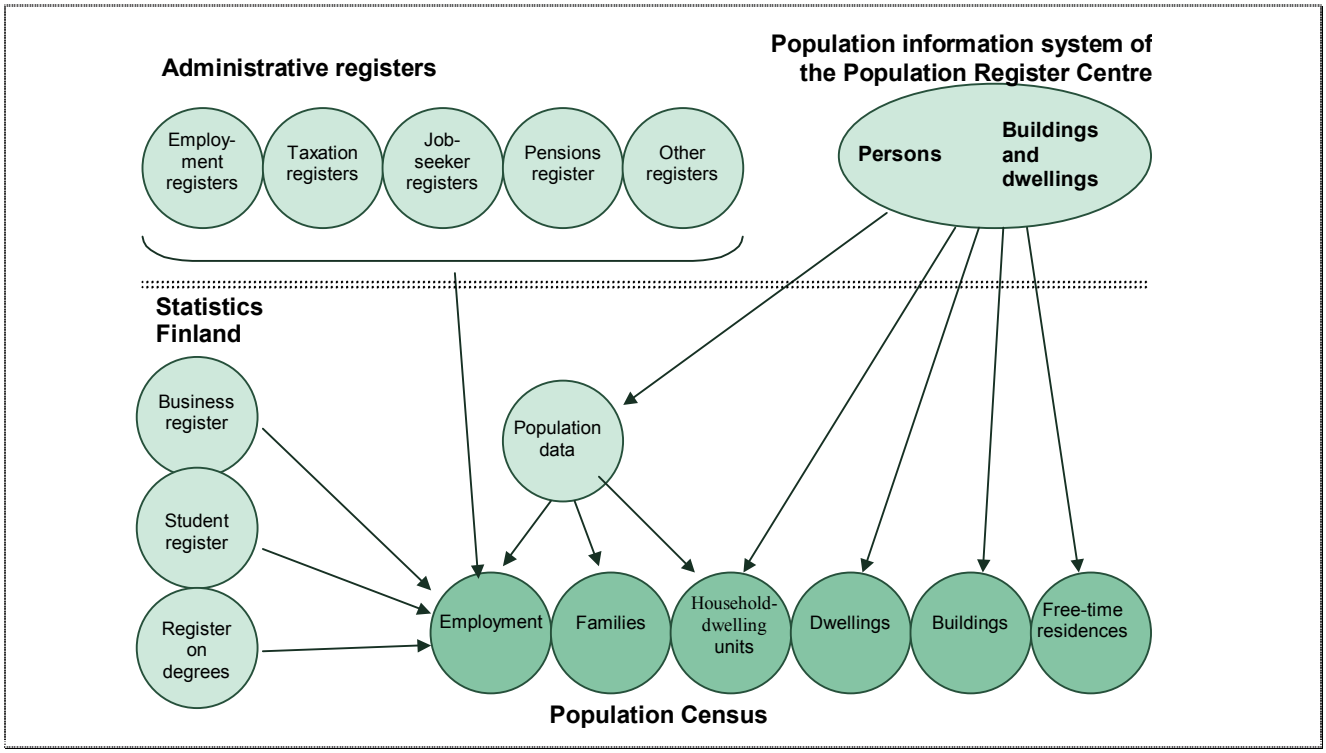
167. As pointed out in chapter 2, developing a register-based data system has in all Nordic countries been a step-by-step process over a rather long period. Statistical registers have been established in several areas and by 2011 all Nordic countries, at least according to national plans, will have a totally register-based population and housing census system. And even if we call it a census system, the same data sources are used in the corresponding subject matter statistics.

168. A register-based statistical system should never be regarded as completed once and for all. As new user needs arise and new administrative registers are established, new information should be integrated in the system.

Register-based population and housing censuses system

169. The population and housing census provides the best example of the use of administrative records in statistical production. Here we will use the Finnish system as a main example.

Example: Register-based population and housing census system in Finland



170. The base registers cover the statistical units relevant to a census: *persons* resident in the country, the *buildings and dwellings* in the country as well as all *enterprises and their establishments* (business register). All statistical units can be linked to one another by means of the identification systems: *persons* can be linked to *families* and *households*, 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 maps using geographical coordinates, because all buildings have been provided by coordinates.

Census statistics and subject matter statistics from the same data source

171. The register-based population census system is comprised of several statistical sub-systems, including the following:

- Population statistics
- Population structure
- Population changes
- Family and household statistics
- Regional employment statistics
- Income statistics
- Education statistics
- Completed education and degrees
- Students
- Buildings and dwelling statistics
- Housing conditions statistics

172. This means that the users get annual (and in some areas even monthly or quarterly) updates of most variables included in the census statistics.

173. In this section we will describe in some more detail two of the sub-systems: regional employment statistics and building and dwelling statistics. Both examples refer to Finland.

174. Register-based employment statistics are among the most important statistical systems in all Nordic countries. 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, status in employment, industry, location of workplace, and occupation.

175. The first set of employment data based on this system was produced in 1980 in Denmark and during the 1980s in Finland, Norway and Sweden. A number of different registers are used in compiling the annual employment statistics. Most of them are needed for defining main type of activity at reference day (for instance end of the year) or during a reference week, which is indeed the primary task of employment statistics.

Example: Register based employment statistics in Finland

176. The system of Finland is an example, because systems are slightly different in all Nordic countries.

177. Data on jobs 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).

178. In connection with data collection from various administrative sources, a separate register is compiled on all jobs in the current year. In addition, the Business Register and other registers on public sector establishments are consulted to obtain information on data on industry, ownership type etc. For people employed in enterprises with several establishments, the data on place of work are only available for the last week of the year on the basis of a separate inquiry conducted by Statistics Finland.

179. Data for activities other than ordinary employment is collected from several administrative registers: on unemployment and persons in labour market measures, conscripts and pensions. Statistics Finland also compiles a student register from a variety of sources.

Example: Statistics on housing and households in Finland

180. The building and dwelling data are used to compile statistics on *building production,* the *building and dwelling stock* and *housing conditions*. In addition, register data on buildings and dwellings are used as a sampling frame for various surveys concerning dwellings or households, such as the rental inquiry and the household budget survey.

181. The data files on the building and dwelling stock and housing conditions are produced once a year for all municipalities and sub-areas within municipalities.

182. 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-based dwelling data through the domicile code. Taxation data for the household-dwelling unit are linked with the household-dwelling data set.

183. *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.

184. 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 residence, which includes separate codes for village and town district. These people are removed from the household-dwelling files.

Chapter 8. DISSEMINATION AND USERS' NEEDS

185. One advantage of register-based data is that they offer information of the entire target population on a regular basis (annually or in some cases even monthly or quarterly). This means that register-based data is very suitable for producing statistics on small groups, both *small social groups* and *small geographical areas*. In this chapter, we will draw the examples from small area statistics.

186. In *flow statistics*, data on the same individual is chained together for consecutive time periods (longitudinal data). This makes it possible to follow statistical units over time. One example is statistics describing the transition from education to working life. Registers are especially suitable as sources for flow statistics, since they comprise the whole population at any point of time, for instance every year.

187. Longitudinal data is also very useful for *research purposes*. Several registers, for example the population register, comprise dates of changes and events. This way it is possible to estimate the precise date transitions or the duration of statuses and activities. In all Nordic countries, micro data based on registers is available for researchers.

8.1. GIS and register-based statistics

188. Traditionally, the most important regional units in statistics have been administrative areas. 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. Further, many users need information on smaller areas or regions crossing municipality boundaries.

189. All Nordic countries have developed what could be described as a *point-based* statistical system. This means that the locations of buildings are specified using map coordinates, giving the exact location of each statistical unit using developed linking system. This building-based code system with its co-ordinates has provided a solid foundation for reliable and flexible statistical areas.

190. The adoption of map coordinates for buildings has also paved the way for a more flexible determination of statistical areas. With most generic Geographical Information System (GIS) packages, this point-in-polygon analysis can be performed with data located to points (buildings) and polygons (statistical areas).

191. The most commonly used non-administrative areas in statistics production are as follows:
- A statistical classification between urban and rural areas[4]
- Localities (urban settlements)
- Municipal sub-areas
- Post code areas
- Grid squares

192. For instance, all variables in population and housing censuses could be specified using map coordinates and could therefore be made available to users on non-administrative areas. However, since small area data may be very detailed, issues of confidentiality and data protection must be considered carefully.

[4] The actual classification may differ somewhat between countries.

8.2. Micro data to researchers

193. Increased use of micro data requires improved possibilities for providing better data to meet the needs of users. It is vital for NSIs to assure that the wealth of stored micro data can be fully utilized by researchers and other authorised users. By and large, the access to micro data means that investments made in official statistics give a higher return.

194. Technological advances in hardware, software, data documentation and the Internet have largely increased the possibilities for improving the access to micro data. However, as the issue of confidentiality and protection of individual data goes hand in hand with use of micro data, a balance is needed between use of data and confidentiality. The main principle is that researchers shall not have access to more data than is needed for a specific project, and that the access is restricted to this specific project and not given for general purposes.

195. Micro data suited for researchers must be standardized and of high quality. The Nordic countries have compiled a number of integrated, statistical registers suitable for analyses and research purposes. The longitudinal integrated register *Louise* from Sweden, containing anonymized micro data on persons and families regarding their education, income and employment might serve as an example. This register includes annual data on all adults in Sweden from 1990 and is updated each year. In Finland, a Longitudinal Census Data File has been established, including all censuses from 1970 (every 5 years), and an annual Longitudinal Employment File has been created from 1987. Samples for researchers may be drawn from such integrated data files and they offer rich possibilities to carry out different analyses.

196. Over a number of years, the NSIs in Norway, Sweden and Finland have distributed anonymous micro data to a large number of research institutions and authorities using magnetic tapes, CD ROMs, DVDs or other formats. The volume has increased simultaneously to the increase in the number of releases/assignments. Denmark has in the past only allowed access to micro data on-site at Statistics Denmark.

Example: Research databases in Denmark

197. Researchers are experiencing an increasing need to analyse multidimensional problems in the course of which data from a number of different statistical registers and/or from many periodic versions of the same register are linked. This results in high start-up costs for a research project.

198. It has however been found that different projects sometimes call for the same sets of background data, which have to be generated again every time. This applies, for example, to labour market data, data on health and data on education. Taking advantage of this situation in order to reduce the costs of individual research projects, Statistics Denmark has set up some databases especially for research purposes. In these cases a large amount of the work in establishing the basic data is done once and for all, so that researchers can get started on their projects without having to undertake all that preparatory work. The databases are continuously updated.

199. To strengthen register-based research Statistics Denmark gives access to tailor-made sets of data compiled from various registers including the research databases mentioned above. Authorised research environments and analysis institutes can receive online access to non-identifiable micro data; that is data where identification numbers for persons and businesses have been removed.

Example: Micro data access in Sweden

200. In 2005, Statistics Sweden developed a new system for remote access to micro data (MONA). With this system users are given secure access to databases at Statistics Sweden from practically any place that can provide Internet access. Data are processed and analysed through a rich set of applications and resulting data sets are automatically sent to the users. The main goals for the MONA system are:

- To increase accessibility to micro data for external users at the same time as security and secrecy is reinforced.
- To keep all types of micro data for research on site at Statistics Sweden enforcing control of where, when, who and how data are used.
- To have instantly upgraded data when needed without any requirements to produce new sets of discs or tapes for redistribution.
- To present an easy-to-use front end for the end users based on well-known standard techniques and components such as server-based computing.
- To present a complete system with powerful servers and a rich set of applications with no requirements on expensive equipment and software costs for end users.

201. It is clear that the Nordic countries are committed to improving access to high-quality micro data. One important issue in future development is to compile several new thematic registers tailored to better meet the needs of the research community. To accomplish this, considerable work is needed, engaging both methodologists and subject matter experts. Another future trend is to develop techniques that allow linkage of data from different sources, both within and outside the Nordic NSIs. In addition, some of the Nordic NSIs are designing Statistical Data Warehouses, which will enable them to build integrated registers and cubes in ways that allow continuous updates of data. It goes without saying that all these development trends are subject to the fulfilment of legitimate disclosure control.

Chapter 9. QUALITY AND STANDARDIZATION

9.1. Quality

202. How do register-based statistics comply with the demands for quality listed in chapter 1? Lowering *costs* and keeping the *response burden* as low as possible are among the main reasons for introducing register-based statistics.

203. *Comparability and coherence* are primarily taken care of at a micro level in the register-based statistical system. Further, the Nordic NSIs have made efforts to harmonize register-based statistics with statistics based on other sources.

204. *Accessibility and clarity* will only to a limited extent depend on the sources used. However, since register-based data systems are quite complex and are used in a number of statistics, metadata have to be comprehensive and well organized. In general, the quality requirements imposed by statistical use of register data coincide to some extent with the requirements that must also be met in serving the primary purpose of the registers. Sufficient metadata from the register keepers is necessary for the judgement of the quality (see chapter 6).

Relevance

205. A very important quality requirement to be met by basic data is that they must be *relevant* to the statistics. This means that the data must relate to the concepts that it is desirable to elucidate in the statistics. It is a lot easier to check for relevance in statistics based on data specially collected for the purpose. When the statistics are to be based on data collected for a different (administrative) purpose, whatever is necessary or desirable for that purpose usually has to suffice.

206. In favourable cases however, the statistician may be fortunate enough to exert an influence on the content of registers.

207. It is also possible to let different data sources supplement one another through linkage, so that the desired information is provided by comparing data from different registers that document the subject more or less perfectly.

208. A statistical system that is exclusively or to a very large extent based on the registers of public authorities may end up giving a picture of the world as seen through the eyes of those authorities. Concepts or sub-populations that do not exist in the frame of reference of the tax authority or social affairs administration also may be missing in the statistics. To exclude this risk it is important that there should be certain base registers, such as the Central Population Register (CPR), whose task is to record all units without reference to any specific administrative purpose. We also know for sure that many records from different authorities can be linked so as to provide statistics that are not tied to any single administrative viewpoint. At the same time it is important to bear in mind that the registers cannot give us a complete picture of society and that it is therefore necessary to collect data covering aspects that are not covered by the registers. These additional data can be collected through interview or questionnaire surveys, and should ideally include identifiers that make it possible to compare them with the register data.

Accuracy

209. Another important requirement to be met by basic data for statistics concerns accuracy. This applies irrespective of whether the data comes from questionnaires or registers. There must be a high degree of certainty that the recorded data faithfully reflects the circumstances investigated.

210. It can of course be argued that, as long as the errors are not systematic, they will not result in biased distributions since errors in opposite directions have a tendency to cancel one another out. When we look at statistical *links* between several variables, however, such errors become a problem since they can distort the assessment of those links.

211. If the errors tend systematically in a particular direction, the problems are immediately magnified, but their systematic nature may make it easier to correct them.

212. Finally it is important that data be recorded with a *degree of precision* suited to the needs of the statistic, i.e. the scale used should be sufficiently detailed. For example, the degree of specification in occupational information must be fairly detailed if the information is to be used for epidemiological studies where the risk of disease is to be viewed in terms of job-related effects.

Timeliness and punctuality

213. Delays in the updating of administrative registers in many cases necessitate the publication of several editions of the figures: *provisional figures*, which have to be issued because of the need for current figures but which lack precision due to inadequate updating, and *final figures* which are more complete but only appear some time later. Sometimes there may be a number of interim versions. The existence of several versions of the figures may however cause great confusion and inconvenience to users and should for that reason be avoided if possible. This is of course a well-known problem also for statistics produced from other sources, but may be specifically relevant for statistics based on registers where updating is substantially delayed.

Quality studies

214. Combining data from surveys and registers offers good opportunities for quality studies. Comparisons are made at both macro and micro level. For instance, in population and housing censuses, relevant sample survey data (for instance from the Labour force survey and surveys on living conditions and housing) is used to assess the quality of the variables in question.

215. At the time when register-based statistics was something new, the assumption was that data from sample surveys should be used to measure the quality of register data. This is still the case, but today register data is also considered useful in assessing the quality of sample survey data.

9.2. Metadata and documentation

216. "Statistical metadata is descriptive information or documentation about statistical data, i.e. micro data, macro data, or other metadata. Statistical metadata facilitates the sharing, querying, and understanding of statistical data over the lifetime of the data."[5]

217. Over the last few decades the need for metadata in the statistical production process has been increasingly evident. Most statistical offices are striving to introduce metadata systems, or improve existing ones.

[5] UN/ECE Workshop on Statistical Metadata [METIS, 1999], Working Paper no. 2

218. The nature of the metadata differs significantly between register-based data and surveys with their own data collection. It is also necessary to distinguish between the documentation of *registers* and the documentation of *register-based statistics.* When using existing registers to create new registers, *register* documentation is crucial. This type of documentation is characterized by:

- The volume of the metadata which can be very high
- Every administrative source must be documented
- Changes in the administrative system must also be documented
- The variables can be complicated so documentation must be precise
- A large amount of data processing is done to create units and variables, and this processing should also be documented

219. This means that the metadata system must suit the requirements of the register system and register-based statistics.

Documentation of administrative sources

220. Suppliers of the data (register owners) submit *record descriptions* etc., which indicate the structure and content of the data being delivered. Furthermore, the NSI should have electronic access to the *questionnaires,* including *instructions,* to be stored in the metadata system.

221. All changes must be carefully noted and stored to make it easy to judge the comparability over time. A metadata system should therefore contain a *calendar,* i.e. a formalized metadata system in which it is possible to search for information related to *time*, *register* and *variable*.

Documentation of sources within the register system

222. In constructing a new statistical register we often use variables from other statistical registers already available in the register system. In this case it is essential to have easy access to the existing metadata to be able to search and select suitable variables. It is important to have the possibility to transfer the metadata from the existing sources to the new register's documentation. To avoid duplicate work and instead conveniently use metadata again, documentations must be strictly formalized according to common rules and stored in a universally accessible way.

Documentation of changes over time

223. Four types of events can affect register statistics, and in order to avoid incorrect interpretations of time series from register-based surveys we need to know the following:

- Have changes taken place in the administrative system that makes up the sources, whereby administrative concepts have been given new definitions?
- Have changes taken place in the way the register has been formed, e.g., new sources or new estimation methods?
- Have there been changes to the classifications that are used in the register?
- Have any external changes taken place that could have affected the statistics indirectly?

224. An events calendar should therefore exist where events that could affect different registers should be documented and stored in one place. This will make things easier both for those working to create the register and for the users of the completed register. The risk for misinterpretations of the data will thereby be reduced.

Classification and definitions database

225. Industrial classification, product category, education, occupation and regional codes are examples of important statistical *standards* and *classifications*. The administrative sources contain data on these hierarchical classifications and this information is used to create variables within the register system. These classifications are changed at regular intervals. As value sets (sets of all codes/categories) are also large, a classification database is needed to manage all the codes and keys between the different versions. This classification database is an important resource when the variables in a register are documented.

Comprehensive metadata system

226. There should be a system that integrates all formalized metadata, including a calendar, and a classification and definitions database. In addition to this, a system is needed to manage documents with other types of metadata.

9.3. Coordination

227. An important part of register-statistical methodology work consists of structuring and improving the whole system. Coordination aims essentially to minimize duplicate work, simplify the combined use of data and increase consistency between different registers. We can define four types of coordination:

- Organizational
- Technical
- Content related
- Methodological

Organizational coordination

228. Register-based studies need an organization that supports extensive cooperation between those responsible for the different parts of the register system. The creation of an appropriate formal organization is important, but not sufficient. The organizational culture also needs to be one in which cooperation over organizational boundaries is seen as inherent.

229. The register-based data systems are organized somewhat differently among the Nordic countries. The Swedish model is one example.

Example: Organizational coordination in Sweden

230. Statistics Sweden has a central organizational unit that carries the overall responsibility for the register system. This unit works with structuring and coordinating the registers into one functioning system.

231. The work with individual statistical registers within the existing system and with statistics based on statistical registers is carried out by local organizational units. The responsibilities must be clearly regulated between the central and the local units in the NSI.

232. The data received from administrative sources are compared with data from surveys relating to the same units. This type of combined editing is an important task, which contributes to the quality of the statistics. To avoid duplicate work it should be carried out in close cooperation with all those working with the involved statistics. Combined editing can either be organized in common projects or by the creation of special organizational units.

233. The development of new integrated (statistical) registers means that certain organizational issues come into focus. To avoid conflict, there should be regular cooperation between the units involved.

Two main factors in having a cultural change from "kingdom ownership" to subject data and statistical products to "common ownership", are the signals from the directors of the NSI, and the demand of integrated statistical data from researchers and other important users. However, the NSI also has to recognize the need for subject specialists to process the administrative data for statistical use, and for the communication with the register keepers.

Technical coordination

234. The most important individual factor for technical coordination is a well-functioning, uniform and standardized *metadata system*, (see 9.2).

235. Standardized *naming conventions* and uniform *data formats* for key variables, such as linkage variables, contribute to the coordination. There should be an unambiguous link between a variable name and a *variable definition* within the register system. A variable name that occurs in several registers must always use the same definition. The opposite, variables with the same definition but different names in different registers, should never occur.

236. Coherence demands coordinated *physical access* to data in the cooperating registers. A minimum requirement is that the registers that form the register system are compatible with the commonly used processing and analysis tools to make sure that the tools can be used to jointly process the registers. It is an advantage if at least a part of the combined processing can be carried out internally in the databases involved.

Content related coordination

237. The coordination of content is based on the following:
- Standardized populations
- Standardized variables
- Standardized groupings of important classifications

238. All populations should be created using the statistical base registers. One important use of the base registers is the creation of *standardized populations*, which are then used to create register populations in the other statistical registers of the system.

239. These standardized populations should be of high quality with regard to coverage and the important tabulating variables. They are therefore created some time after the period or point in time that they refer to. As far as possible, all administrative information should be received and processed before the population is defined. The standardized populations should then not be changed, and other statistical registers should use these standardized register populations.

240. The register system should be based on two basic types of standardized register populations:
- The turn of the year population – the units existing at the turn of the year
- The calendar year population – the units that existed at some time during the year

241. *Coordination of variables* is another vital issue. An important coordination tool is a central classification database containing statistical classifications that are used within the statistical system. It should include classifications and keys between different versions of a classification that has been established as the standard nationally or internationally.

242. The variables that are relevant for standardization are those used recurrently within several registers. What organizational unit inside the NSI is responsible for what variables must be clarified. This will normally be the unit that is receiving the variable. This responsibility includes the processing, naming and

documentation of these variables. Correspondingly, the units that create important derived variables are responsible for the naming and documentation of these.

243. Many variables must be grouped or divided into class intervals. The groups and class intervals must be made according to common rules, or the tables from different sources will not be comparable. Age and income categories, for example, should be formed in the same way in different registers and sample surveys. The larger classifications, such as industry, education and occupation must be grouped before statistical tables can be formed. Grouping these variables requires good knowledge of the variables. Those responsible for these variables within the organization should therefore create a number of standardized groupings that can be adapted for different needs. Both aggregated divisions and more detailed divisions are needed for different purposes. The groups created should also have standardized names.

Coordination of statistical methodology

244. A common approach and common terminology are important for the development of register-based statistics.

245. "*Base registers should be used when defining populations*" and "*everyone should support the base registers*" are two fundamental principles that contribute to the consistency of the system's populations. Anyone who is linking and matching several registers must trust the fact that every register has been processed, edited and documented to make processing work easier and to assure quality.

246. Calculations may need coordinating so that the estimates from different registers are consistent. If, for example, a variable that exists in several registers has partial non-response, the non-response should be dealt with in the same way in the registers concerned.

Chapter 10. REGISTER-BASED POPULATION AND HOUSING CENSUSES

247. A main observation in all Nordic countries is that the development of register-based censuses is a step-by-step process. First subject matter statistics were tested and published in different areas. Register-based variables were introduced in the census as soon as the quality was considered sufficient. When statistics had been developed for all areas relevant for censuses, a totally register-based census could be conducted.

248. In other words, the census is actually an integrated part of a system of register-based information. The same statistical registers are used as sources for censuses and for subject matter statistics. This means that almost all points made in this report on register-based statistics in general are also valid for register-based censuses. In this chapter we will discuss in some more detail topics that are especially relevant for censuses.

10.1. Developing a register-based census

What is a register-based census?

249. At one end of the line we have traditional censuses collecting data by use of enumerators and questionnaires, using no register information at all. At the other end we have the totally register-based census. Some countries use mixed mode data collection with a combination of data from registers and questionnaires (as a total count or a sample survey). However, even countries conducting mainly traditional censuses may use register information to some extent, for instance an address list for persons or households for mailing out census forms.

250. For some countries, such as in the Nordic region, the goal has been to develop fully register-based censuses. For other countries, the goal may be to use registers to a certain extent, even if it is not possible or desirable to conduct a fully register-based census. And for some countries, a register-based census may not be an option at all.

251. What is required in order to call a census "register-based"? The main requirement must be the existence of a population register and a dwelling register. After all, resident persons and housing units are the basic units in a population and housing census. The link between persons and their dwellings is equally important, giving the household unit. These requirements constitute the cornerstones for developing a register-based census.

Quality assessments and decision processes

252. When is a statistical register or a system of registers good enough to be used as a main source in a census, or more generally for statistical purposes? We will use the Norwegian experiences as an example.

Example: Developing register-based censuses in Norway

253. The Central Population Register (CPR) was established in 1964, based on the 1960 Census. The census in 1970 was used to check and update the CPR, and also to establish a register of education. In the 1980 Census, administrative registers were used to rationalize data collection. Data capture by postal mail only was introduced. The questionnaires were substantially shortened and only sent to people aged 16 years or older. Information on demography, education and income was register-based. Registers were also used during the coding and editing of the data. The census was further used to update the register of education.

254. Several studies investigated the feasibility of a 1990 Census based on registers only. As an experiment, existing registers were used to produce census statistics for 1980, and the results were compared with the results from the ordinary census of 1980. The conclusion was that even if labour market data were available in existing registers, they were not of sufficient quality for census purposes. Little or no information on households and housing was available from registers so the decision was to collect labour market data and data on households and housing by questionnaires from a sample of the population. The information from the sample survey was combined with information from various administrative and statistical registers, resulting in a high-quality statistical register.

255. In planning the 2001 Census it was clear that all data on persons could be based on registers. However, there existed still no dwelling register, so data on households and housing had to be collected by a full coverage survey with a questionnaire for every occupied dwelling. Based on a decision in the Parliament, a project was started to extend the Register of Ground Properties, Addresses and Buildings to include dwellings. Furthermore, the official addresses in the CPR were extended to include precise dwelling numbers for persons living in multi-dwelling buildings. Information on dwellings and the link between resident persons and dwellings were captured by the 2001 Census and later entered in the registers. This means all registers necessary for conducting fully register-based censuses in the future are established.

256. To measure the quality of register data and make decisions on when quality is sufficient, several methods may be used. One method specifically related to censuses is to produce alternative register-based "census" data and compare it with the latest ordinary census. This way it is possible to assess what parts of the register system are of a sufficient quality and what parts have to be improved in order to conduct register-based censuses.

257. Another observation is the mutual relationship between registers and censuses in the Nordic countries. Censuses have been used in establishing registers, and register data is subsequently used as a source for censuses. In order to transfer data from censuses to administrative registers it has been necessary to use the legal bases for the administrative registers involved, in addition to the Statistics Act.

10.2. Comparability with international recommendations

Essential features of population and housing censuses

258. Register-based population census data largely correspond to the recommendations for population and housing censuses issued by the United Nations Economic Commission for Europe and the Statistical Office of the European Communities. According to these recommendations there are five essential features.

259. *Individual enumeration*: A register is defined as a systematic collection of *unit-level data* organized in such a way that updating is possible. The registers used in censuses comprise all persons and dwellings, so the condition of individual enumeration is met.

260. *Universality within a defined territory*: The register-based population census includes all persons registered as residents in a country on the census day. The housing census includes all conventional dwellings and institutions, but non-conventional dwellings are not included.

261. *Simultaneity*: Since all registers are updated regularly, all units can be enumerated at the census day.

262. *Small area data*: Register-based data are well suited for producing small area statistics.

263. *Defined periodicity*: Most of the census variables are published annually. Complete censuses can be conducted whenever necessary, for instance every ten years.

Units and variables

264. According to the recommendations, data are to be collected on the following statistical units: *persons*, *households*, *family nuclei*, *living quarters*, and *buildings*. Register-based data can be produced on all these units. The *dwelling household* concept is used, that is all persons living in the same dwelling belong to the same household.

265. Concerning *census* variables, the situation may differ somewhat between the countries. In general, however, with a fully developed register-based census system, all *core topics* can be produced.

266. However, some *non-core topic*s cannot be produced from register data. For instance, *ethnicity* is not included. This is a variable that has necessarily a subjective dimension and must be based on individual self-declaration. This kind of information is normally not included in administrative registers[6]. Other examples are *computer literacy, mode of transportation to work* and *durable consumer goods possessed by the household.* Some variables are available in some countries, but not in others, examples being *religion, language* and *time usually worked.*

10.3. Advantages and disadvantages of register-based census production

267. The advantages and disadvantages of producing register-based censuses are of course the same as for register-based *statistics* in general. Some issues are, however, especially relevant for census statistics.

Advantages

268. *Reducing costs* are particularly relevant in a census context. When conducting a register-based census there is no longer any need to design and test census forms, 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 NSIs no longer need to recruit large numbers of people to sift through the returned forms. In connection with the 1980 census in Finland, for instance, some 2500 employees were hired and trained.

269. The *reduced response burden* on the population is another significant advantage indeed. For example in Finland in the 1980 Census people 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.

270. All the census statistics are available *annually.* This means that users will not have to wait 10 years for an update of *census* statistics. To put it another way: census statistics and annual subject matter statistics could be compared directly.

Disadvantages

271. One 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, meaning that *some topics have to be dropped* from the register-based population census system, (see 10.2). Sample surveys can be used to collect data that are not available through registers. Imputation methods may be used to integrate these data into the census file.

---

[6] *Country of birth* and *country of birth of parents* will normally be available from administrative sources. These variables may, in combination with for instance *citizenship* and *religion*, give some information related to ethnic group, but they do not fully cover the contents of the variable *ethnicity*.

272. Additionally, some restrictions are imposed on the definitions of units and variables. A private household is defined as all persons living in the same dwelling (*dwelling household*). *Housekeeping units* (persons living in the same dwelling with joint board) cannot be created using registers. The population is *normally* counted according to *legal* place of residence (according to the population register) and not *de facto* place of residence.

273. Using administrative registers as a source for censuses may have an effect on the timeliness. The time needed to produce the census variables (including the time used for updating the administrative registers) differs from one subject to another. For instance, in Norway demographic variables could be published three months after census day, but for income variables the production time is almost 18 months. These differences are due to different routines for updating the corresponding administrative registers. So for register-based censuses the statistics will be timelier for some variables and less timely for other variables compared to traditional censuses.

274. When census data are collected from registers, the census is *no longer a collection tool for meeting emerging information needs in the society*. In countries conducting traditional censuses, new variables may be added to the questionnaire to collect new data items that are considered to be very important. This flexibility is lost when data are no longer collected by means of questionnaires. However, also in a traditional census this is a very limited scope, since extending a census form is a very expensive operation. So irrespective of census mode, the alternative for collecting additional data will for most countries be by using sample surveys.

Chapter 11. CONCLUDING REMARKS

The same statistics based on another type of data

275. To make statistics from administrative sources means that data collection, editing and other kinds of data processing are done by methods other than the traditional ones. The statistical product, however, is still the same. Instead of asking persons, households or enterprises for information, we ask the owners of the administrative registers. Instead of making quality controls of data received from the individuals, we have to modify administrative data based on our knowledge of differences in definitions and coverage between the administrative sources and the statistical needs. In both cases human resources are required, but the type of qualifications needed may differ. Once the registers are established and the information has passed a quality limit, the register-based solution generally gives lower costs than the traditional one.

Some base registers are necessary

276. Three base registers seem to be necessary for a statistical system based mainly on administrative sources: a population register, a business register (enterprises and establishments) and a register of addresses, buildings and dwellings. A unique identification for each unit in the base registers, and links between them are necessary. The unique identification should also be used in other administrative registers for these units (educational registers, taxation registers and so on).

277. Register-based statistical systems may be established without these three base registers, but in our experience the lack of such registers will give problems with the quality of the statistics produced.

Statisticians should give active support to base registers

278. Administrative registers are normally established and managed by the relevant authorities, but often the NSI provides active support in initiating and organizing basic registers. In some cases it has even been necessary for the NSI to take responsibility for establishing and managing base registers as well as other administrative registers. In such cases, as soon as the register is established for administrative use, and a reasonable quality and well functioning routines for managing the registers are established, the responsibility for the register should be taken over by another official and qualified authority. However, even then the NSI should be active in the technical support (reporting quality problems) and in the political support (call for resources needed) of the register.

Many users of a register give better statistics

279. Establishing and managing administrative registers is very expensive. If many institutions (official as well as private) can argue that their goals can be fulfilled more reasonably by using an administrative register, the government may be willing to put money in it.

280. Even the administrative users can contribute to the quality of the register. If for instance one user reports that a unit that should be in the register is not present, the inclusion of that new unit will improve the quality for all other users, including the NSI. The NSI may inform the register keepers of quality problems, but in general not report on errors on individual units in the registers. In the active support of the register, the NSIs must assure themselves that they are not breaking the national statistics act or the UN Fundamental principles on official statistics.

Requirements for register data from the statistician

281. Both units and variables need to be well defined, and the definition should be easy to understand for the person giving the information to the register (the reference person or the civil servant asking the questions and writing down the answers). For each subject it is preferable to have one single register covering all actual units with all information needed. However, in practice we have to accept that information on one subject has to be collected from more than one source. To integrate data from two or more sources is one of the challenges for the statisticians. If it is difficult to identify the statistical unit from the administrative sources we must create some estimation procedures. Otherwise the data may be useless for statistical purposes. The time reference (point of time or period) must be clearly defined, and an event should ideally be marked with both the date it took place and the date the information was entered into the register.

282. Administrative registers shall serve administrative purposes, and the rules and procedures that are valid for the register, may change over time. The effect of changes on the quality of the statistics in general, and for the time series in particular, should be taken into consideration when the register is adjusted according to new administrative rules. The best way to safeguard the statistical needs is to have good relations between the persons responsible for the registers and persons responsible for the statistics. These contacts and experiences have to be documented and given to new persons in the register authority and in NSI when a person's position is changed.

Register data is a "common good" inside the NSI

283. Data from one administrative source may be direct input to more than one official statistics. The source may also be used in the quality control of other statistics, and the more administrative data are used the better the quality will be for the statistics.

284. All statistical data, register-based or not, are part of the statistical system, and all data may be used in subject matter statistics or in larger statistical data systems like population and housing censuses. Therefore, data are "common goods". The relations between the NSI and the administrative register may be organized through one unit in the NSI or (in practice) one person, but the responsible unit or person is acting on behalf of the whole NSI, and does indeed have to be conscious of this both in internal and external communication.

Concepts in register-based statistics

Administrative data source: Comprise in principle all kind of sources used for administrative purposes. In this report all administrative data sources mentioned are registers.

Administrative register: Register primarily used in an administrative information system. This means that the registers are used in the production of goods and services in public or private institutions or companies, or that the information is a result of such production. Administrative registers used for statistical purposes are normally operated by the state or jointly by local authorities, but registers operated by private organizations are also used.

Base register: *Administrative* base registers are kept as a basic resource for public or private administration. The function is to keep stock of the population and to maintain identification information. *Statistical* base registers are based on the corresponding administrative registers. Their principle tasks are to define important populations and contain links to other base registers.

Derived variable: New variable formed by using existing variables.

Link: One or several connecting variables that identify individual units. Links (or *keys*) are used when several registers are matched.

Primary register: Most often equivalent to administrative registers, but also used for statistical registers in areas where no central administrative register exists.

Register: Systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, updating, correcting or extending the register.

Register-based census: When all data is collected from statistical registers, we call it a (totally) register-based population and housing census. A census based on combined data from registers and questionnaire is called a partially register-based census.

Register-based statistical system: Statistical registers are included in a common and coordinated system.

Register-based statistics: Statistics produced by using register data only. Data from other sources (sample surveys) may be used indirectly, for instance for imputation, calibration of models or quality assessments.

Register keeper: See *register owner*.

Register owner: Authority responsible for an administrative register. Also called *register keeper.*

Secondary register: See Statistical register

Specialized register: Register, which unlike base registers, serves one specific purpose or a clearly defined group of purposes. Specialized registers often receive information on the population and some basic data from a base register, but supply other data themselves.

Statistical register: Register processed for statistical purposes. A statistical register could be based on one or several administrative registers. Statistical registers are also referred to as *secondary registers*.

Survey data: Data collected directly for statistical purposes. Surveys may cover the total population, but are more frequently based on a sample.

Unit: A population consists of a number of units (or objects).

References

Houbiers, M., 2004. Towards a Social Statistical Database and unified estimates at Statistics Netherlands, in: Journal of Official Statistics, Volume 20, No. 1, 2004, 55-75.

Longva S., Thomsen I. and Severeide P.I. (1998): Reducing Costs of Censuses in Norway through Use of Administrative Registers. Reprint from International Statistical Review, Vol. 66, 223-234

Schulte Nordholt, E., M. Hartgers and R. Gircour (eds.), 2004. The Dutch Virtual Census of 2001, Analysis and methodology. (Voorburg/Heerlen, 2004).
(http://www.cbs.nl/en-GB/menu/themas/dossiers/volkstellingen/publicaties/2001-b57-pub.htm)

Schulte Nordholt, E., 2005. The Dutch virtual Census 2001: A new approach by combining different sources, in: Statistical Journal of the United Nations Economic Commission for Europe, Volume 22, Number 1, 2005, 25-37.

Statistics Denmark (1995): Statistics on Persons in Denmark – a register-based statistical system

Statistics Denmark (2005): New developments in the Danish system for access to micro data. Working paper for the joint UNECE/Eurostat Work Session on statistical data confidentiality in Geneva 9.11 November 2005.

Statistics Finland (2004): Use of Registers and Administrative Data Sources for Statistical Purposes – Best practices in Statistics Finland.

Statistics Norway (2005): Editing and imputation for the creation of a linked micro file from base registers and other administrative data. Working paper for the UNECE Work session on statistical data editing in Ottawa 16-18 May 2005. (http://www.unece.org/stats/documents/2005/05/sde/wp.8.e.pdf)

Statistics Norway (2006): The role of censuses in a country with a register-based statistical system: Norwegian experiences and plans. Paper for the CES Plenary Session in Paris 13-15 June 2006. (http://www.unece.org/stats/documents/ece/ces/2006/32.e.pdf)

Wallgren A. and Wallgren B. (2006): Register-Based Statistics - Administrative Data for Statistical Purposes. John Wiley & Sons, Ltd

*****

Register-based statistics in the Nordic countries

Review of best practices with focus on population and social statistics

The Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) have a long tradition in using administrative registers in the production of official statistics. They also have a long tradition in cooperating in the field of statistics. The National Statistical Institutes of the Nordic countries decided to share their experience and knowledge with the international statistical community, by producing comprehensive documentation of their best practices in this field.

In recent years, an increasing number of countries in the UNECE region are considering the possibility of producing statistics based on administrative registers. The present volume offers strategic and planning officers in the National Statistical Institutes an understanding of what register-based statistics are, covering also the necessary technical and administrative capacity, and the possible applications of these methods to produce official statistics. The emphasis of the publication is on the use of administrative registers to produce demographic and social statistics.

In publishing the present volume, the United Nations Economic Commission for Europe (UNECE) would like to support the Nordic countries in sharing their experience in this field with the international statistical community at large. The volume represents a valuable tool for all National Statistical Institutes that are planning to produce official statistics based on administrative registers. It also supports the implementation of the 2010 round of population and housing censuses since an increasing number of countries are planning to use administrative sources to supplement or replace the traditional data collection in the field. Guidelines on how to use registers for population and housing censuses are also included in the new set of "Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing", prepared by UNECE in cooperation with the Statistical Office of the European Communities (EUROSTAT).

Printed at United Nations, Geneva
GE.00-00000-Month 2007-9,999
Sales No. 0.00.00.0.0

ISBN 00-0-000000-0
ISSN 0000-0000

ECE/STAT/--/--

Bar code aera