

The use of administrative sources for censuses: Merits and challenges

Lars Thygesen

Thygesen Statistics Consulting, Holmevej 77, 2860 Søborg, Denmark

E-mail: lthygesen@gmail.com

Abstract. The paper discusses the prospects for increased use of administrative sources as the basis for population and housing censuses. The background for this discussion is the case story of how Danish statistics production was reformed during the period c. 1966–1981, especially with regard to social statistics and population and housing censuses. The new paradigm combined increased reliance on administrative sources with a philosophy of a statistical system that was more flexible towards meeting new and unforeseen statistical needs.

Keywords: Administrative sources, registers, linkage, census, population, housing, enterprises, statistical system, identifiers, confidentiality

1. Introduction

This paper describes how Danish statistics production was reformed during the period c. 1966–1981, especially with regard to social statistics and population and housing censuses. The new paradigm combined increased reliance on administrative sources with a philosophy of a statistical system that is more flexible towards meeting new and unforeseen statistical needs. Both of these trends were in line with ideas advanced by Svein Nordbotten from the start of the 1960s [8]. Nordbotten was probably the first to foresee the great potential of using administrative sources for statistics and even foresee a scenario where censuses could be based on these sources.

2. The development of the Danish register-based statistical system

Since the origins of official statistics, which is often defined to be around the mid-18th century, administrative data have been heavily used as a source of statistics. In fact, many of the first recognised statistics were based on church registries.

However, a new way of using these administrative sources in Denmark developed during the period

1966–1981 and thoroughly reshaped statistics. This development is summarised in the milestones shown in the Table 1.

The Act on Statistics Denmark was enacted in 1966 [4]. It gave a new and more independent status to the organisation, which until then formally had been a government department. But it also aimed to modernise the statistical methods and to give Statistics Denmark an ideal platform for engaging in the admin-based statistics. It provided a strong legal basis for the development of a register based, or an *archive statistical system*.

First, the Act has a provision granting Statistics Denmark full access to all registered information held by public authorities; it has been established in a court of justice that this access comprises identification numbers such as the Person Number. Second, the Act gives the institution formal authority to influence the registers run by public authorities so that they can be useful as statistical sources: “Statistics Denmark . . . will supervise or assist in the establishment and utilisation of central public registers which serve to perform administrative duties for the public sector, business and industry, and which can be used for statistical purposes” [4], Section 1.3.

In the remarks to the Law, it was foreseen that in the future, a full population census could be taken without

Table 1
Milestones in the development of Register Statistics since 1966

Statistics Denmark's program and work plan are published annually. Some of the main points of development, especially with regard to the use of register methods, are presented below.

- The Act on Statistics Denmark enacted in 1966, enhanced the institution's independence and strengthened access to administrative records of public authorities.
 - Starting in 1968, a statistic was established covering turnover and other factors in all economic activities, based on information from the Value Added Tax (VAT) Administration combined with data from Statistics Denmark's Central Business Register.
 - In 1968, the Central Population Register (CPR) was founded as a base register for the entire public administration. It was managed by the Ministry of the Interior. It maintains basic information such as name, address, marital status and family relations and a unique identifier, the Person Number (PN).
 - In 1971 a statistic was introduced covering employment in business enterprises and, starting in 1974, including total wages paid, based on employers' returns to the Supplementary Pension (ATP) System and tax authorities, respectively.
 - Once these statistics had been introduced, the general business enterprise censuses, which had been undertaken for 1925, 1935, 1948 and 1958, were discontinued.
 - On May 1, 1970, total figures were released or published for the first time, based on the CPR, showing the population in the individual municipalities broken down according to sex, age and marital status. In the period 1970–73, other population and vital statistics were reorganised on the basis of the CPR, which dispensed with manual reporting from the local authorities.
 - Starting in 1970, income statistics were switched to a register base in conjunction with the introduction of taxation at the source. Starting in 1976, the statistics were extensively reorganised, amongst other things by linking the income data with data from an employment classification system which was also used for other purposes.
 - In 1971, Statistics Denmark established a Sample Archive Register for use in the development of statistical methods based on registers. It held a longitudinal sample of the population, covering diverse subjects from many sources.
 - Following a decision made in 1971, in 1973, Statistics Denmark took over the production of statistics on pupils and students from the Ministry of Education. The individual data reported by the educational establishments were recorded in a Register of Education and Training Statistics, so that the regular returns of the establishments only needed to cover changes with respect to the previous year.
 - In 1974, the Minister of Economic Affairs decided that a traditional questionnaire-based population census would not be carried out for the census date 1975/6. Instead, a register-based population census (or "census-like statistics") was undertaken in 1976 on the basis of CPR data combined with a range of other register data relating to various aspects of economic activity.
 - Starting in 1977, a statistic relating to the Social Assistance Act was established based on person-oriented information from the Joint Local Authority Finance System, supplemented with data from questionnaires returned by non-affiliated local authorities and combined with family data from the population statistics. This statistic gave rise to a very heated public debate on the question whether it was proper to use person numbers for the production of statistics and led to a court case between Statistics Denmark and five local authorities who refused to provide data. In 1981, Statistics Denmark won its case at the Supreme Court.
 - On April 1, 1977 the first figures for buildings and dwellings were produced on the basis of the newly established Central Register of Buildings and Dwellings (BBR). On January 1, 1980 the first register-based housing census was conducted, combining data on homes with data on the occupants.
 - In 1979, it was decided to establish a Register of Statistics on Workplaces based on existing register data and limited collection of supplementary data with the assistance of the tax authorities and employers. This finally put an end to total data collections from the population in the form of traditional population censuses.
 - With effect from 1979, the statistic on criminal convictions, which had previously been a summary compilation based on questionnaire information, was reorganised on the basis of annual reports from the Central Criminal Register of the Danish Police Force.
 - In 1980, the Law Model was established in cooperation with the Ministry of Economic Affairs. The model, accessible for authorised staff in government departments, contained micro data on a sample of the population including a broad range of data for calculating the effects of contemplated legislation.
 - On January 1, 1981 the first Population and Housing Census was conducted entirely based on register information.
 - Over the period 1982–84, a number of social statistics registers were set up based on extracts from the joint local authority benefit payment systems throughout the country. This removed a considerable burden of manual work while at the same time dramatically increasing the scope for statistical uses.
 - In 1988, Statistics Denmark opened a service allowing researchers to make their analysis based on micro data; the service was only on site, but starting in 2001 researchers could have remote access.
 - In 1992, the Register of Motor Vehicle Statistics was set up based on extracts from the Central Motor Vehicle Register of the Danish Police Force. The register forms the basis for figures on the ownership and use of motor vehicles.
 - In 1992, a new health statistic was introduced based on extracts from the National Register of Patients, combined with a range of background data from existing statistical registers relating to, among others, the population, social patterns, income, employment, education and housing.
 - In 1994, an extended statistic was introduced on persons covered by employment training schemes and retirement schemes, based on person-oriented data reported by the administrative authorities.
 - In 1994, after several years of extensive developmental work, an integrated, modernised business register system was established with a number of content-related and technical improvements.
 - In 1994–1995, an extensive reform of wage statistics was carried out which facilitated the production of comprehensive and consistent person-oriented figures for all wage-earners in both the private and the public sector.
-

asking any questions to the public, only recombining already existing data from administrative sources. This was an extraordinary degree of foresight shown by the fathers (and mothers) of the Law, and the provisions of the Act have greatly helped in making the visions come true, some 15 years later. The Government was determined that the census in 1970 had to be the last conventional census in Denmark.

With regard to the development of social statistics and population and housing censuses, the first important step was the establishment of the statistics on population size and its distribution from 1970 based on the Central Population Register, CPR; this was followed in 1973–74 by the vital statistics based on extracts of events in the CPR combined with information from medical records on births and deaths.

Following the Population and Housing Census 1970, which was the last traditional census with questionnaires to all households to be carried out in Denmark, the usual periodicity of such censuses would have the next one occurring in 1975. Following intensive hearings among principal census users, the Minister of Economic Affairs, who had the authority to decide on censuses according to the Act, decided on the advice of the Board of Statistics Denmark that no census would be taken in 1975. Instead, a “census-like” statistical exercise would be carried out in 1976, combining the available administrative sources to produce as much as possible of what users wanted a census to provide. The decision to not have the census in 1975 was based on the precondition that a full census from registers would be feasible by 1980/81.

The most severe lacunas in the database existing around 1975 were: a) the lack of housing information, b) the lack of information on the workplace addresses of the persons, which would, *inter alia*, allow statistics on commuting needs, very much used for transportation planning, and c) the lack of specific kind-of-activity information for employed persons, as this information was only linked to the legal business units. Other deficiencies were incomplete information on citizens’ education and occupation. It was decided that these lacunas must be filled before the next Census term in 1980. It should be noted that this was a huge task involving close and often difficult negotiations with many administrative authorities; new legislation on a dwellings register would be required and some of the necessary improvements could only be partly implemented by 1981. Others would come well into the 1980s.

During the 1970s, several social statistics domains were reengineered using administrative sources: Edu-

cation, social assistance, and crime. This development continued during the 1980s, when health insurance, hospital treatments and a number of other social benefits were added. All of these statistics could subsequently be seen as part of a coherent system, allowing ad hoc as well as regular combination of data from the different domains. In addition, also the – rather few – traditional surveys which were still taken at regular intervals, notably the Labour Force Survey, were also seen as parts of the system and data could be linked for the surveyed persons or households.

In 1978, the Central Register on Buildings and Dwellings was established. It was run by the municipalities under the coordination of the Ministry of Housing. Statistics Denmark had been intensively involved in the preparation and the specification of the register during the preceding years in order to ensure that it would meet the statistical needs, especially for the Census. One of the economic arguments that allowed the Act on the Central Register on Buildings and Dwellings to be passed by Parliament was the fact that very costly censuses could be avoided in the future.

Another key activity aimed at filling the data needs regarding the link between persons and their physical workplaces. This was organised in the “work place project” from 1978 onwards. The basic source was the tax registers holding information on employments, including the link between employer and employee. This had to be enhanced with registrations of the workplace in cases where an employer had more than one such place. First, the registration of existing work places had to be rendered complete and included in the Business Register. Secondly, the tax authorities had to include an extra piece of information in their register on employments, namely the work place, and this piece of information was to be included exclusively for statistical purposes and must not be used for administrative purposes [6]. The tax authorities did not recognise any need for this information, a decision they have later regretted.

By 1981, the house had been built. All the basic elements were in place. The final test was to be the full scale Population and Housing Census that users expected to be published that year. A privileged user was the European Union (EU), as Denmark was under obligation to provide a full set of census tables to the EU. The census consisted of a process of linking existing statistical registers and producing the desired outputs. This was completed on time and with a quality that has not been challenged. Thus, Denmark became the first country to conduct a full registerbased census, fulfilling international standards.

Since this first registerbased census, we have not talked very much about population censuses in Denmark. There has been a continuous census, in the sense that any census outputs can be produced at any time. This is seen as one of the big advantages of the method. There is nothing to indicate that census data are only needed every 5 or 10 years – but users have until now been forced to accept to use outdated statistics as the basis for their planning purposes because it had not been feasible to have more up-to-date data.

Statistics Denmark still produces an extraordinary census package for Eurostat every census term, but no special publishing is made in Denmark. Instead, most of the information of a regular census is made available with annual periodicity through the dissemination database *StatBank Denmark*. This information is fairly broad and detailed, e.g. geography down to municipality level, population, household and family characteristics regarding housing, commuting, occupation, incomes, education, etc. More specialised outputs, such as small area statistics tailored to individual needs, are produced for a fee recovering all additional costs necessary to produce the outputs from the register system. It is necessary to maintain this distinction between free and at cost statistics, since the archive statistical method makes it possible to compile infinite amounts of statistics and needs are virtually endless. In many cases, it is necessary to strengthen the quality assurance of the data when users request new combinations of data or data with a higher degree of detail, for which the initial processing has not been designed. But the introduction of charging a fee for the statistics on a larger scale – stipulated in the Act on Statistics Denmark – ran counter to the culture of official statistics, especially in the early days, thus creating internal tensions, and many external users found it difficult to accept.

It should be noted that there is no price charged for the data; customers only pay the full cost of any additional effort that would not have been necessary to produce the “regular” outputs stated in the statistical program. This can include more thorough editing, necessary to be able to give good estimates at a finer level of breakdown, e.g., small area statistics.

3. The basic ideas of the new statistics paradigm

A fundamental principle of the described development has been to ensure that data are properly organised so they can be used in the future, and to make use of data from the past. This section will try to illuminate how this was implemented in the early years of restructuring, mainly 1970–1981.

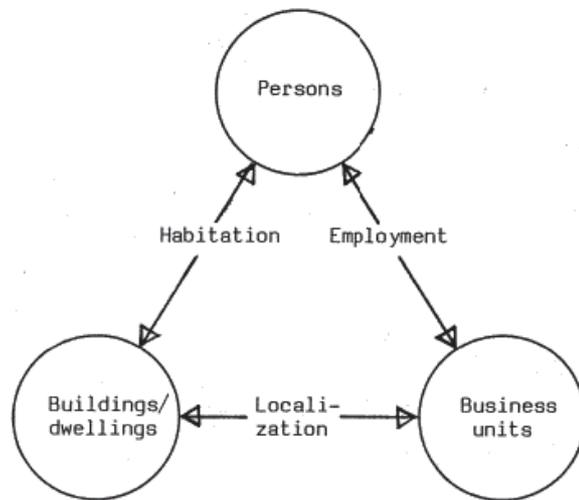


Fig. 1. A simple model.

3.1. The notion of a system

Traditionally, statisticians tend to see their statistics on a specific subject as islands in a stream. Each statistic is produced on its own, with little or no connection to other statistics. They are collected, edited, analysed and disseminated as a workflow that does not need to be linked to other systems.

In the “new” paradigm, we think of the different branches of official statistics as one coherent system: This is prerequisite to using data from different domains for unforeseen purposes. The idea progressively invaded the Danish statistical production during the period 1970–1981. In the early 1970s, linking across domains was mostly addressed within the statistics on persons, but gradually the focus shifted to establishing links between persons and families on one hand, and business units or dwelling and real estate units on the other. This was the biggest challenge in building the system.

The notion of a statistical system was explicitly discussed in Danmarks Statistik [1]. Thygesen [13] talks about a “Socio-demographic statistical system” and states: “Since the early seventies a full register-statistical system has been a goal in the strategic planning of Danmarks Statistik”. Figure 1 shows a simple model for the system. Each corner of the triangle in Fig. 1 consists of a number of more or less independent statistical registers, each containing the data necessary for one field of statistics, e.g. personal income statistics. The registers can be linked together by means of the unique and common identifiers of each type of ob-

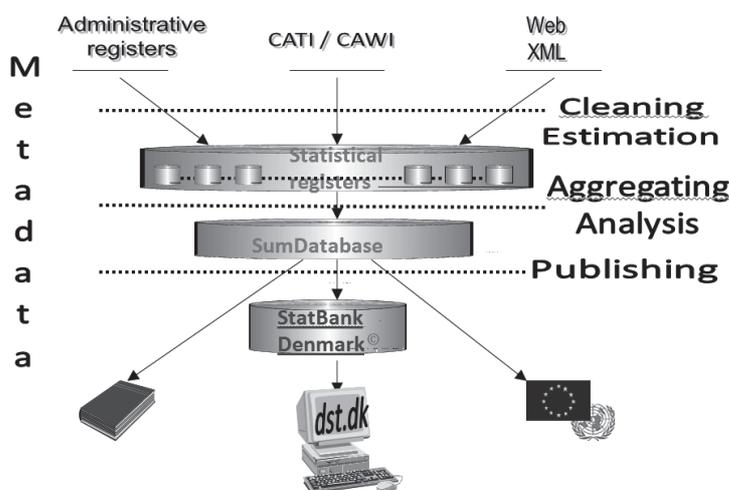


Fig. 2. A simple model of the archive system.

ject, the most important of which is the Person Number.

A special kind of registers was called Systematized Data Modules [13, p. 235]. These are not aimed at one branch of statistics, but contain data for general use in several domains. One example is the Employment Classification module. These modules underpin the coherence of the system.

3.2. Reuse of data for multiple purposes

The idea of reuse of data for multiple (unforeseen) purposes was recognised as a key objective of the Danish development. Even though the statistical registers were aimed at each branch of statistics, they could produce many other outputs and they were combined across domains when a need was recognised.

Until now, statisticians had to concentrate completely on fulfilling needs that were already well established and described, but this was not satisfactory from society's point of view. As costs of storing and processing data were reduced, new opportunities arose.

In Denmark, two initiatives in particular underlined the wish to reuse and combine data across domains for multiple purposes: The Sample Archive from 1971 and the Law Model from 1980 onwards.

The Sample Archive (originally called the Mini Population Register) held data on a representative sample of the population from a great variety of sources and all kinds of social and population statistics. It was aimed at trying out different combinations of those data, especially for analysis – a test bed for new combinations. The processing power was quite limited and expensive

in those days, so using full population registers could be almost prohibitive. As this situation changed over the years, the importance of the Sample Archive diminished and eventually vanished.

The Law Model is similar but with a much greater practical importance. It was established in 1980 in cooperation with the Ministry of Economic Affairs. The model contained micro data on a sample of the population including a broad range of data for calculating the effects of contemplated legislation.

The system was managed in collaboration and was hosted at Statistics Denmark, and its importance was underlined by the fact that government departments, especially the Ministry of Finance and the Ministry of Taxation, demanded and got guarantees that it would be up and running 24 hours a day, 7 days a week. The model remains active.

3.3. The potential of new use of administrative data

From 1970 onwards, reuse of administrative sources replaced direct data collection. An often quoted, unofficial Statistics Denmark estimate from around 1980 holds that 95% of the organisation's data are based on administrative sources rather than on surveys. Although such an estimate does not really make a lot of sense (what is the unit of measurement, is it bits or information value?), it is still an expression of the importance attributed to administrative registers by Statistics Denmark.

3.4. *The filing or archive system*

Figure 2 illustrates the flow in the Danish system. The active files are currently created by processing all the different sources. Collection may take place over time but the creation of “versions” of the statistical registers normally takes place at the same interval as the frequency for the statistics in question, e.g., annually for crime statistics. Until recent years, the creation of the registers was carried out in something like a “stove-pipe” fashion, each register being created on its own; however, there were streams of data between the pipes, and as the figure indicates, there was a vision of seeing it all as one huge data warehouse.

Identifiers are kept in each register, so the registers can interchange information when needs arise. Typically, they are kept as annual versions. Differences may exist between one year and the next, both in the attributes actually kept on each individual (person, company, dwelling, etc.) and in the definition of each attribute. Differences may be due to changes in the administrative registers feeding the statistics, or they may be caused by new needs, improved methodology, budgetary problems that make it impossible to continue certain statistics, etc. No matter what the cause, these changes give rise to problems in some of the uses for which the system is designed. More about this in Section 4.4.

The figure also shows how detailed, but aggregated data are stored in a macro statistics databank, StatBank Denmark.

3.5. *The base registers*

In a register based statistical system it is of key importance to be able to keep track of the most important “objects” that are described by the system: Persons, dwellings and business units. It is necessary that these objects are well defined and used in the same way in all data and statistics. It is also important for statistics that all the authorities providing input data for the statistics agree on and use the same basic registrations. It is necessary to define and keep track of birth, migration and death of units. This is done in so-called base registers, which also administer the common identifiers to be used throughout all systems.

It has been discussed in many countries whether such base registers should be controlled by the statistics authority or by some administrative authorities, and in the latter case, if this ought to be centralised into one “register administration”. In Denmark the so-

lution was that the base registers are run not by the statistics authority but by three different administrative authorities: The Ministry of the Interior (persons), the Ministry of Housing (buildings and dwellings) and the Business Authority (businesses).¹ The reason why this organisation was chosen is to ensure that data from the Statistics Bureau not be used for purposes other than statistics and analysis. Therefore, data may not be passed from statistical to administrative uses, i.e., where individual data collected by Statistics Denmark would be used in the administration vis-à-vis citizens or businesses.

However, Statistics Denmark has been, and still is, deeply involved in the construction and development of the three base registers. Our involvement in this is mentioned in the Act on Statistics Denmark, Section 1.3. The advice of the organisation has carried much weight, as the purpose of producing statistics is very closely related to the role of the base register: What is essential to statistics is also beneficial to making the base register useful to a very broad range of (administrative) uses.

As has been shown, the Danish Statistical system was built around three *statistical* base registers – persons, dwellings and businesses – each reflecting the events (deaths, births, etc.) in the “real” administrative base register behind it, but enriched with basic statistical knowledge where possible. The statistical base registers are a kind of “shadows” of the administrative registers, linked to subject-specific statistical registers.

3.6. *Micro data and longitudinal analysis*

Innovative analysis has become possible by linkage of data across domains and perhaps especially over time. This allows analysis to be based upon time profiles for individual objects, i.e., longitudinal analysis.

In Denmark, work on this started soon after the first registers with Person Number identifiers were developed. Pioneering was a mortality study revealing considerable differences in mortality over a period of time (1970–1975), depending on previous occupation [2]. Another path-breaking study combined Statistics Denmark’s register data with data from the Cancer Registry to illuminate the occupational differentials in cancer incidences and prospects [7]. Since then a large and ever growing number of studies have been carried

¹The organisational setup has changed since the 1980s but the basic idea remains unchanged.

out, mostly by external researchers under the umbrella of Statistics Denmark's research arrangements, since 2000 the Remote Data Access for Researchers. Under this arrangement, authorised research can be conducted with anonymised micro data sets created especially for the purpose of each research project; data may also include survey data specially collected for the research project. Access takes place through a secure network, protected by considerable organisational and technical safeguards. It is a paid service, as it goes far beyond the statistical program financed by the Government. A large number of researchers and institutions conduct a wide range of research projects, amounting to a revenue of 3 M€ in 2014.

3.7. Confidentiality and data protection

It is obvious that the potential risks multiply when the statistical system increasingly is built on huge collections or archives of identified data on persons, enterprises, etc., that can be combined as needs arise. The public's perception of the risks could be expected to become a threat. Therefore, the importance of keeping of the trust of the public has to be seriously addressed, and much effort has to be dedicated to work on safeguards to protect confidentiality.

During the 1970s, Statistics Denmark therefore intensified the measures for data protection. Internal data security regulations imposed strict organisational and technical measures. Rules were enforced limiting the internal as well as the external access to data, following a need-to-know principle. In 1978, a new Register Law laid down general rules and introduced a Data Inspection Board to supervise protection of personal data in all public authorities. This had a beneficial effect on the authorities' awareness of confidentiality and data protection, and also on public trust.

The guiding principle of Statistics Denmark's data protection policy has been balancing the need to create and give access to knowledge against the need to keep personal and business information well guarded and protected.

4. Difficulties in the practical implementation

4.1. Convincing ourselves

One of the challenges of the new paradigm was to get the organisation and all its members to agree that this is the right way to go. After all, there were many

champions of the "old" system, where data collection and processing was made for one or a limited set of purposes, using conventional survey techniques. This is a very common management problem when introducing deep change – but that did not make it any easier.

The statistical textbooks that were used by the universities did not say anything about registers. On the contrary, they talked about distinct surveys, and how statisticians have to make absolutely clear in advance exactly which tables will be produced as a result. This was the conventional wisdom; how could a new paradigm be acceptable, where the statistics should be designed to be able to accommodate current and also unforeseen, future needs?

The internal discussions regarding the paradigm shift were heated for many years and reduced the pace at which such a big reform could take place. The lawmakers behind the Act on Statistics Denmark had laid down a good foundation and the management of Statistics Denmark was firm in the belief that this was the way to go, which was absolutely essential to success. In recent years the register-based strategy has been accepted almost unanimously inside the organisation and among Danish users of official statistics.

4.2. Convincing the rest of the world

When the archive statistical system was first introduced during the 1970s, it was met with widespread international disapproval and suspicion. For many years, discussions with the EU statistical office, Eurostat, were difficult. Statistics Denmark of course wished to fulfil the European obligations by using the results from the register-based system. In principle, this was acceptable, as the treaty allowed freedom of method of compiling the statistics as long as results had sufficient quality, including relevance. However, there was little understanding in practice. On several occasions, register-based statistics were labelled as inferior to "real statistics". At the time, there were only few other countries – notably the Nordic countries – who used, or were able to use similar techniques. Among international colleagues in other national statistical offices – with the notable exception of the prominent chief statistician of France, Edmond Malinvaud – the method was generally regarded with suspicion.

Statistics Denmark invested in trying to change this attitude towards register-based statistics. This was done in meetings and conferences, especially within the European statistical system. Many colleagues ex-

pressed the belief that the truth is a census form. It was argued in several international meetings that people trust in statistical offices and are willing to give them correct information, whereas they will never give correct information to administrative authorities.

One of the pioneers of register-based statistics in Denmark, Jørgen Wedebye, gave the following statement in the Nordic Statisticians' Meeting in 1979: "On what grounds can anyone claim that there has ever been one single piece of correct information recorded on a census form?" There is in fact little evidence or reason to believe that citizens really trust public statisticians not to pass on information to administrative authorities.

Another obstacle to international acceptance of the archive statistical method was confidentiality. Several countries pointed out bad experiences going back in history, especially in relation to the persecution of Jews during WW2.

Gradually, the international dismissal of registers and archives has diminished, as many countries have created or are now endeavouring to create similar systems. It was a great step towards acceptance, when Eurostat supported the translation and publication of the book *Statistics on Persons in Denmark – A Register-Based Statistical System* [11].

4.3. Confidentiality and access to data

One of the most serious challenges to setting up an archive statistical system has been the need to establish effective and credible rules and procedures to protect confidentiality.

According to the Law, Statistics Denmark has the right of access to register data kept by other public authorities. However, this was not necessarily acceptable to all citizens or to the public authorities who were the guardians of the data.

A cornerstone in the Danish register-based statistical strategy was giving up the collection of data for the population and housing censuses after 1970. It is well known that traditional censuses created big public opinion problems in many other European countries at the time, and the fact that this could be avoided in Denmark helped the strategy. On the other hand, some citizens and opinion leaders were strongly against the increasing use of large computerised registers. The threat of a "Big Brother" society was launched in a fierce public debate in the late 1970s, especially focusing on statistics rather than the administrative uses of registers, as one would have expected. Statistics were a target probably because many people do not at all under-

stand the usefulness of statistics, while public administration is seen as a necessary evil. Eventually a lawsuit was filed against Statistics Denmark in 1978–81 by five municipalities. They claimed that Statistics Denmark was not entitled to collect data with Person Number from registers kept by the social service authorities regarding social assistance. The case was finally won by Statistics Denmark in the Supreme Court.

Since this case, there has been little public criticism of the use of registers and personal identifiers for statistics. The public's general trust in Statistics Denmark is high, also by international standards.

4.4. Data breaks

The reliance on administrative sources has given rise to problems of continuity of time series. Major problems are posed for a statistic when legislative or regulatory changes result in alterations to the data content of administrative registers. First, it may be difficult or impossible to assess the long-term trend in a particular indicator if different definitions are used in the base material. Second, problems may arise for statistics on the number of events, in deciding to which extent changes in data values are to be viewed as reflecting actual events or are a consequence of new concepts or definitions.

Altered data in administrative registers may be due to changes in the legislation applicable to the field, and the statistical consequences depend on what type of statistics is involved. If it is in fact a statistic for the monitoring of legislation, the function of which is to show how the administration of a law affects ordinary people, for example, the statistics merely have to go along with and adopt the concepts of the new legislation. Statistics on cash benefits must thus, at any given time, reflect the rules applicable to the field and it may then be difficult to assess aspects such as behavioural changes.

If we are concerned with a more general statistics, on the other hand, it is not acceptable that it should be impossible to compare the statistical concepts before and after the legislative change. General statistics 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 (e.g. daily cash 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 linked together. In some cases, it may help estimating the concept combining several administrative sources relating to the subject.

4.5. Metadata

Statistical metadata, i.e., descriptions of what the data really are and mean, are important in any statistical system. However, when we look at an archive-statistical system they become even more crucial. In such a system, it is necessary at any time to accommodate uses of the data that we never imagined. It is not enough to present the indicators decided in advance together with definitions and quality information regarding those. The metadata for the archive data must be sufficiently good and rich enough to allow a user to determine, with a certain margin of error, whether the data in the archive could be recombined to create new knowledge, e.g., to confirm or reject certain hypotheses.

At a minimum, the metadata must contain descriptions of methodology, definitions of concepts used, definitions of variables and classifications. Breaks in data time series and other quality problems must be documented. The metadata must cover all archive data and be updated when changes occur.

In all statistical offices it has proven very difficult to direct enough attention to the metadata, as the main task of statisticians is to “get the statistics out” on time. Statistics Denmark is no exception. Internal as well as outside users claim this is the most important hindrance to making full use of the potential of the system. Therefore, it remains an area of high priority in the strategy of Statistics Denmark.

5. Concluding remarks

The development of a Population and Housing Census and a full statistical system based mainly on administrative sources has taken some 15 years to accomplish, but it has been a process where many useful products have been released along the way. The vision

of the system has been relatively clear from the very beginning, and the strategy to achieve it has been persistent. The conditions for realising the strategy have been favourable, as the administrative systems as well as the legislation give an excellent foundation. Nevertheless, it has been necessary along the way to influence legislation and adjust contents of administrative registers, and to constantly fight to obtain and keep public trust.

References

- [1] Danmarks Statistik (1970–1980): *Arbejdsplaner (Work Plans)*, 1970–1980. Statistics Denmark, Copenhagen.
- [2] Danmarks Statistik (1979): *Dødelighed og erhverv 1970–75 (Occupational mortality 1970–75)*. Statistical Studies No. 37, Statistics Denmark, Copenhagen 1979.
- [3] Danmarks Statistik (1982): *Personstatistik på registergrundlag (Register Based Statistics on Persons)*. Copenhagen 1982.
- [4] Danmarks Statistik (2006): *Act on Statistics Denmark*. In *Statistical Code (Lovsamling om statistik)*, p. 127–130. <http://www.dst.dk/pukora/epub/upload/10035/lovsam2006.pdf>. Copenhagen 2006
- [5] P. Jensen, *Dansk Statistik (Danish Statistics)*. Vol. 1–2. Statistics Denmark, Copenhagen 2000.
- [6] P. Jensen and L. Thygesen, *Linkage of Records on Objects of Different Kinds. Methodological Problems and Practical Experience*. Bulletin of the International Statistical Institute, Vol. LI, Book 2. Amsterdam, 1985.
- [7] E. Lyng and L. Thygesen, *Occupational Cancer in Denmark*. Scandinavian Journal of Work, Environment and Health. Vol. 16, suppl. 2. Helsinki 1990.
- [8] S. Nordbotten, *Elektronmaskinene og Statistikkens Utforming i Årene Framover (Computers and the Future Form of Statistics)*, 1961. The Statistical Conferences of the Nordic Countries in Helsinki 1960. Statistical Reports of the Nordic Countries, Vol. 7, Helsingfors 1981, pp. 135–141.
- [9] S. Nordbotten, *A Statistical File System*. Statistisk Tidskrift 1966: 2. Stockholm, 1966.
- [10] S. Nordbotten, Om arkivstatistiske systemer. Annex 4, in: *Statistical Reports of the Nordic Countries*, Vol. 14. Copenhagen 1967.
- [11] Statistics Denmark and Eurostat: *Statistics on persons in Denmark – A register-based statistical system*. Eurostat, Luxembourg 1995.
- [12] L. Thygesen, *Personal Identification Numbers and Population Statistics in Denmark*. Statistics Denmark, Copenhagen 1978.
- [13] L. Thygesen, *Methodological Problems Connected with a Socio-Demographic Statistical System Based on Administrative Records*. Bulletin of the International Statistical Institute, Volume L Book 1, Madrid 1983.