# The importance of the archive statistical idea for the development of social statistics and population and housing censuses in Denmark

## *(Betydningen af den arkivstatistiske idé for udvikling af social statistik og folketællinger)*

Lars Thygesen, Statistics Denmark



This paper describes how Danish statistics production was reformed during the period c. 1966-1981, especially as regards 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 new and unforeseen statistical needs. Both of these trends were in line with the ideas advanced by Svein Nordbotten from the start of the 1960s. While this was not entirely accidental, there has also been a certain amount of parallel thinking.

In order to describe the development and Nordbotten's influence on it, I have drawn on all accessible, published sources describing the history, not least Poul Jensen's books on the history of Danish statistics (Jensen 2000). However, the sources do not in all cases reveal the discussions that took place and the deliberations behind decisions. I have therefore permitted myself, having been involved in part of the development, to rely on my own deficient memory and evaluations of these aspects. Unavoidably, this means that the exposition is seen from a participant's perspective, and the paper does not pretend to be an objective historical study.

## *1. The impact of Nordbotten on statistical thinking in Denmark*

For more than a hundred years the Nordic statistical cooperation has been an important source of inspiration for development of new ideas and methods in Danish statistics. The cooperation was organised around periodical Nordic Statisticians' Meetings and Nordic Chief Statisticians' Meetings. In this framework, Svein Nordbotten made a couple of very visionary presentations about how the new electronic processing technology could and should influence the way we think about statistics.

The first one was his famous speech in 1960 Nordic Statisticians' Meeting in Helsinki (Nordbotten 1961), later published in part in Nordbotten (1966). In 1966, Nordbotten made his points even clearer in the Nordic Chief Statisticians' Meeting in Copenhagen (Nordbotten 1967). This is the clearest expression of his ideas I have been able to find.

Nordbotten's analysis was based on the new opportunities opened by the emerging electronic data processing (EDP) techniques. In 1960 these techniques were still rather immature, but Nordbotten anticipated their potential and realised that they would change the cost structure and thus make feasible to reuse information in innovative ways. It would become economically advantageous to store micro data in archives, e.g. from administrative registers, with a view to linking and recombining them as needs arose. Until then, the typical way of using administrative data was to collect them in aggregated or partly aggregated form and then progressively summarising them to create macro data designed in advance for just one or a limited set of purposes. Similarly, surveys would be carried out in such a way that accommodated a certain set of tables and uses. These techniques were of course basically sound, as it is always extremely wise to have the intentional use in mind when designing data collection. The novelty was that reuse for other purposes could be thought into the process, which would greatly enhance the value of the statistics. This would lead to a vision of the statistics as a whole system consisting of "archives" with links between them.

In the Nordic Statisticians' Meeting in Copenhagen 1964, Ingvar Ohlsson, Director General of Statistics Sweden, presented a vision in for the future development of statistics, to a large extent based on Nordbotten's ideas (Ohlsson 1964). He explained how some of the ideas could be carried out in practical work. For instance, regarding the information on attained level of education, one would start from census data, keep it in a register with personal identification and update it with information on new graduates from educational institutions. The intensive use of data linkage was foreseen, and response burden could be reduced. He also referred to the complaints from respondents that "we have already given this info, why do you ask again".

Ohlsson's presentation greatly enhanced the interest in the ideas, and the meeting was attended by 85 statisticians from Denmark, among them Poul Jensen, later to become a key figure in the development of the new statistics. In the audience were also Kjeld Bjerke and Aage La Cour, Division Heads at the Danish Department for Statistics who in 1965-66 became members of the commission that prepared the Act on Statistics Denmark. The meeting is characterised in Jensen (2000) as a break-through of the new ideas. Nevertheless, strong opposition against the new ideas persisted for several years to come in the other Nordic countries, and in some quarters at Statistics Denmark as well.

During the same years a new Act on Statistics Denmark was being prepared, to be enacted in 1966. It gave a new and more independent status to the organisation, which had until then formally been a government department. But it also aimed to modernise the statistical methods and gave Statistics

Denmark an ideal platform for engaging in the register-based statistics. In the remarks to the Law, it was foreseen that in future, a full population census could be taken without asking any questions to the public, only recombining already existing data from administrative sources. This was an extraordinary degree of foresight from the side of the fathers (and mothers) of the Law, and the provisions of the Act have greatly helped in making the dreams coming true, some 15 years later. When Statistics Denmark started its life in 1967, the new head, N. V. Skak-Nielsen proclaimed this policy as a major goal in his program statement[1].

It is not certain that the "father" of the 1966 Law, Viggo Kampmann[2], knew about Nordbotten's writings, but he certainly thought along similar lines[3]. However, the members of the committee preparing the Law were well acquainted with the ideas, as 5 out of 7 members had attended Ohlsson's and Nordbotten's lectures at the Nordic Statistical Meeting in Copenhagen in 1964: Kjeld Bjerke, Aage la Cour, Henning Friis, Erling Jørgensen, and Axel Holm, (Jensen (2000) p. 72f).

Kampmann was determined that the coming census in 1970 had to be the last conventional one to be taken in Denmark and he clearly foresaw a system much like what was later to come. Kampmann was a visionary politician with excellent knowledge of statistics as well as administration and registers.

## *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 developed during the period 1966-1981 and thoroughly reshaped statistics. This development is summarised in the milestones shown in the Annex. The chief architect of the development of register-based statistics in Denmark was Poul Jensen, who worked in Statistics Denmark 1955-67 and 1969-95, the last 20 years as Director.

The Act on Statistics Denmark[4] was enacted in 1966. As mentioned above, it laid down a solid foundation for the development to take place in the following decades. It provided a strong legal basis for the development of a register based, or an archive statistical system. Firstly, the Act has a provision granting Statistics Denmark full access to all registered information held by public authorities; it has been established in court of justice that this access comprises identification numbers such as the Person Number[5]. Second, the Act gives the institution the 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"[6].

As regards 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

[1] Jensen (2000), vol. 1, p. 89
[2] Chairman of the Administration Council, former Prime Minister and Minister of Finance
[3] Skak-Nielsen (2010)
[4] Danmarks Statistik (2006), Section 6
[5] Jensen (2000), vol. 2, p. 99-104
[6] Danmarks Statistik (2006), Section 1, 3

1970, based on the 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 in 1975. 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-alike" statistical exercise would be carried out as per 1976, combining the available administrative sources to produce as much as possible of what users wanted a census to provide. The decision of not having 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 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 around 1980. It should be noted that this was a huge task involving close and often difficult cooperation with many administrative authorities, and some of the necessary improvements could only be partly implemented by 1981, others went on well into the 1980s.

During the 1970s several social statistics domains were reengineered using administrative sources: Education, social assistance, crime. This development continued during the 1980s, when sick 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 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 so-called workplace 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. Firstly, the registration of existing workplaces 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 workplace, and this piece of information was to be included exclusively for statistical purposes and must not be used for administrative purposes[7]. This was because the tax authorities did not recognise any need for this information, a decision they have later regretted.

[7] Jensen & Thygesen (1985)

In 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, as Denmark was under obligation to provide a full set of Census tables to the EU. The census consisted in a process of linking existing statistical registers and producing the desired outputs, and this was completed on time and with a quality that has not been challenged. Thus Denmark became the first country to conduct a full register based census, fulfilling international standards.

Since this first register based 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 could be produced at any time. This is seen as one of the big advantages of the method, as there is nothing to indicate that census data are only needed every 5 or 10 years – but users have until now just been forced to accept to use outdated statistics as the basis for their planning purposes because it was not feasible to have better data (Jensen 1977). Statistics Denmark still produces an extraordinary census package for Eurostat on 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; 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 at 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 paid 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 a higher degree of detail, for which the initial processing has not been designed. But the introduction of paid 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 data themselves; 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.

## *3. The basic ideas of Nordbotten – how they were implemented*

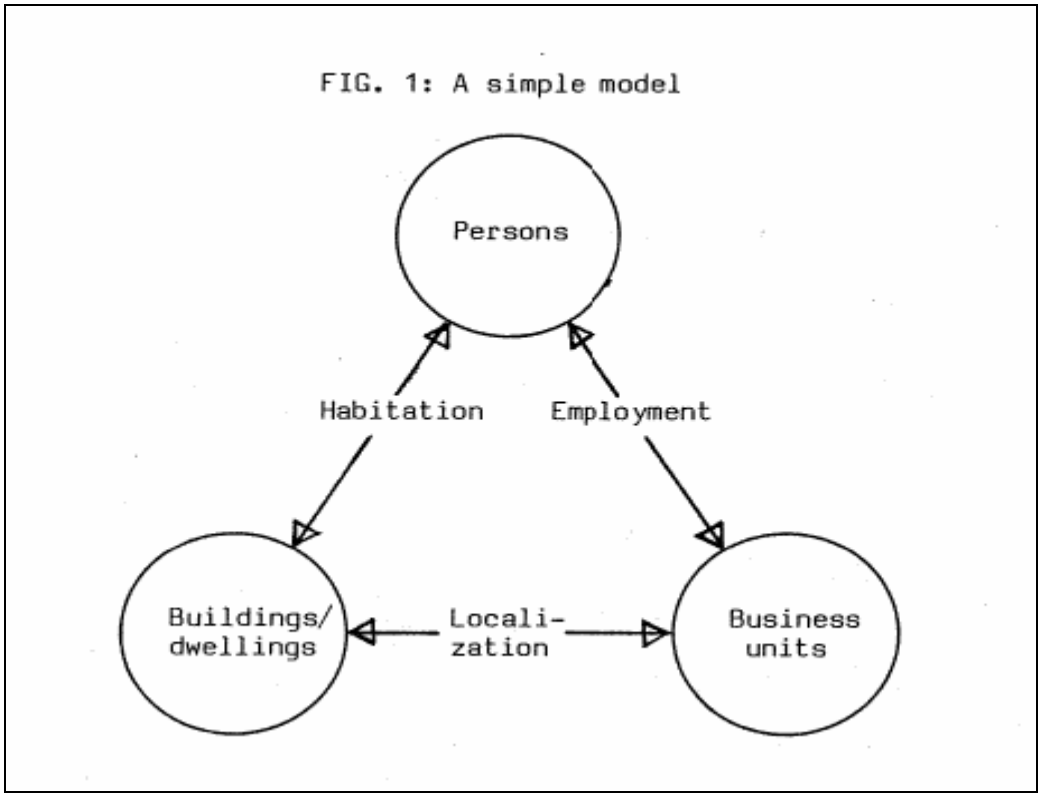
Nordbotten helped make Danish statisticians aware that, as a result of the electronic data processing revolution (or evolution), the balance was changing dramatically between the value of information and the cost of making it available. The cost is reduced, and the value increases manyfold because of the possibility of integration and data linkage from different sources. Data and statistics already collected (the capital) enter positively into the value of the present statistics provision: the more data and statistics you already posses, the more value may be created by a new piece of information, because it can be integrated. All of this may come true, if data are properly organised so they can be used in future, and make use of data from the past. This has several implications. This chapter will try to illuminate how the ideas were implemented in the early years of restructuring, mainly 1970-1981.

## 3.1 The notion of a system

Traditionally, statisticians tend to see their statistics on a specific subject as islands in a stream. Each statistics 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.

Nordbotten thinks of the different branches of official statistics as one coherent system: This is prerequisite to using data from different domains for unforeseen purposes. This idea progressively invaded the Danish statistical production during the period 1970-81. In the early 1970s, the linking across domains was mostly addressed within the statistics on persons, but gradually the focus shifted to establishing links between persons and families on the 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 (1980). Thygesen (1983) 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"; a simple model for the system was offered:




Each corner of the triangle in Fig. 1 above 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 object, most important of which is the Person Number.

A special kind of registers was called Systematized Data Modules (Thygesen 1983, p. 235). These are not aimed at one branch of statistics, but they contain data for general use in several fields. 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 is at the heart of Nordbotten's early writing, and this was also recognised as a key objective of the Danish development. Even though the statistical registers were aimed at each their branch of statistics, they could produce many other outputs and they were combined across domains when a need was recognised.

One of Nordbotten's most important observations was that until now, statisticians had had to concentrate completely on fulfilling needs that were already well established, but this was not satisfactory seen from society's point of view. As costs of storing and processing data were reduced, new opportunities arose and should be taken: "... it is essential that information requests anticipated but not received must be taken into account when collecting and processing. A significant portion of the processing may result in statistics which never are used. In a statistical file system the variable costs per unit, statistic or table will be more dominant because the data already are filed and it will therefore be more efficient to satisfy special needs when they occur. The files will probably encourage to make experiments as for example in the field of sequence-analytical estimation in order to reduce processing costs[8]."

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. One of the reasons for this approach was that processing power was limited 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 in a way a similar machine 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 in activity to present times.

[8] Nordbotten (1967)

**Fig. 2: The Law Model**

*"...and when our expert has coded the proposed law changes into the terminal, we will in a few seconds witness the computer calculating the effects for a typical family - - based on a simulated model population"*[9]


## 3.3 The potential of new use of administrative data

Both Nordbotten and Ohlsson stressed that reuse of administrative sources would replace direct data collection. This is exactly what increasingly happened in Denmark from 1970 onwards, culminating with the full scale register based housing and population census in 1981. 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 doesn't 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.

[9] Drawing by Claus Ib Olsen from Nordic Statistical Secretariat (1990)

## 3.4 The filing or archive system

Three quotations from Nordbotten (1967) outline the idea of the file system:

- "By a statistical file system is meant a system in which the data collection is done continuously and independently of the traditional statistical fields and observation times, and in which the processing of filed data is carried out when needs occur."
- "The data may be divided into three groups: The *active files,* the *historic files* and the *statistical registers.* In the active files the most recent observations are stored while the historic files are used for all historic data. The latter files will grow and the problem of optimum amputation will therefore have to be considered. The statistical registers will be used as a basis for collection and will contain information such as the identification numbers, names and addresses of the units."
- "In order to obtain an up-to-date statistical file, collection, editing, and filing should be more continuous instead of the present periodical pattern."

The Danish system is in many ways an attempt of realising these ideas. As illustrated in Fig. 3, the active files are currently created by processing all the different sources. Collection may take place spread out 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 dream of seeing it all as one huge data warehouse. Nordbotten stated that the total database should be seen as a *data box* with 3 dimensions: Statistical unit, variable and time. During the last 10 years, Statistics Denmark has made a real movement in that direction.

Identifiers are kept in all these registers, 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 arising, improved methodology, budgetary problems making it impossible to continue certain statistics, etc. No matter what is the cause, these changes give rise to problems in some of the uses for which the system is aimed. More about this in paragraph 4.4 below.

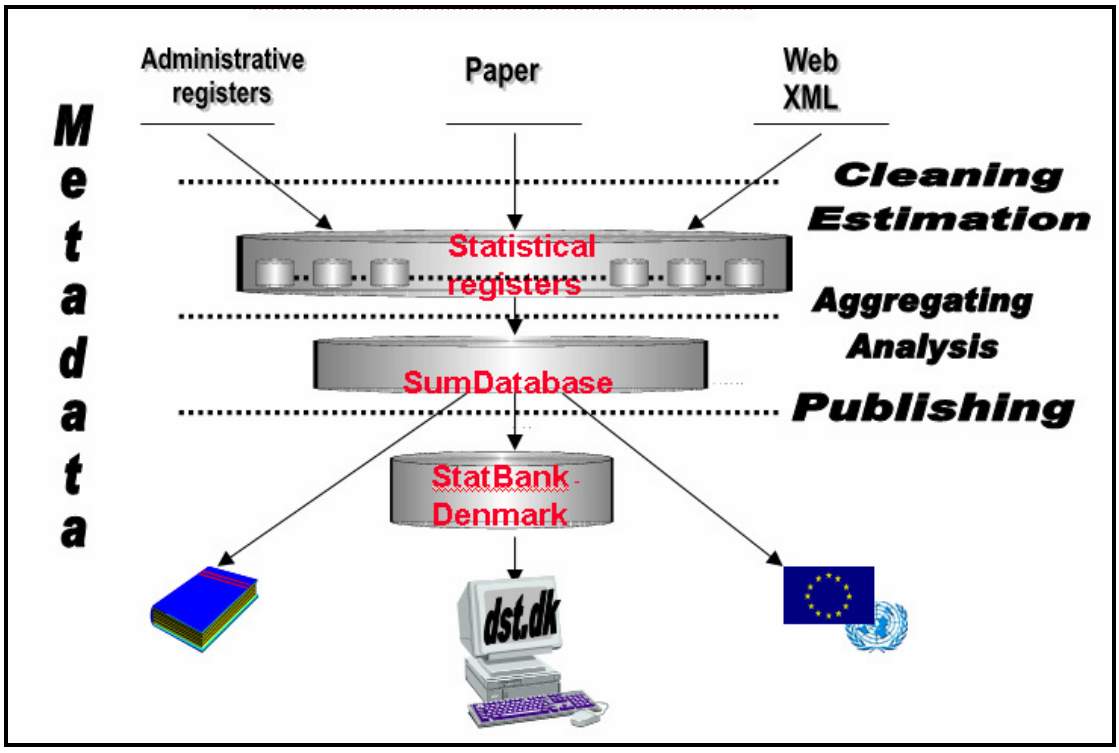



Fig. 3: A simple model of the archive system (from a PowerPoint presentation, SEUGI Paris 2002 by Lars Thygesen)

The figure also shows how aggregated data are stored in a macro statistics databank, StatBank, which was also foreseen by Nordbotten.

## 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[10]. 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 must 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 solution was chosen that the base registers are not run 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 Ministry of Taxation (businesses)[11]. The reason why this organisation was chosen is that data from the statistics bureau cannot be used for purposes other than statistics and analysis. Therefore data may not be passed on from statistics to administrative uses, where individual data collected by Statistics Denmark would be used in the administration vis-à-vis citizens or businesses.

[10] Nordbotten (1967), p. 39-42 discusses these base registers and also includes an "Employers register", which is included in the Business register in the Danish model.
[11] The organisational setup has changed since the 1980s but the basic idea remains unchanged.

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 warranted by 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 such a 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 above, the Danish Statistical system was built around three *statistical* base registers – for persons, dwellings and businesses – each reflecting the events (deaths, births, etc.) in the "real" administrative base register behind it but being enriched with basic statistical knowledge where possible. The statistical base registers were a kind of "shadows" of the administrative registers, linked to subject-specific statistical registers.

## 3.6 Micro data and longitudinal analysis

Nordbotten foresaw innovative analysis that would become possible by linkage of data across domains and perhaps especially over time. This would allow 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. Elsebeth Lynge was responsible for the first mortality study revealing considerable differences in mortality over a period of time (1970-75), depending on previous occupation (Danmarks Statistik 1979). Another path-breaking study combined Statistics Denmark's register data with data from the Cancer Registry to illuminate the occupational differentials in cancer incidence and prospects (Lynge & Thygesen 1990). Since then a large and ever growing number of studies have been carried 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 environments can conduct research directly on anonymised micro data sets created specially for their research purpose; 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 2 M€ in 2009, which is really proof of Nordbotten's idea.

## 3.7 Confidentiality and data protection

Nordbotten's early articles (1961, 1967) did not emphasise very much the importance of addressing the risk of misuse of data – they focused more on possibilities and opportunities, and the value of new information as compared to costs. However, it is – and was – obvious that when the statistical system increasingly was built on huge collections or archives of identified data on persons, enterprises, etc., that could be combined as needs arose, the potential risks would multiply. The public's perception of the risks could be expected to become a threat. Therefore, the importance of keeping the confidence of the public would have to be seriously addressed, and much effort would have 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 comprised 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 realising Nordbotten's vision has been making the organisation itself and all its members 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. This is a very common management problem when introducing deep change – but that didn't make it any easier.

The statistical textbooks that were taught at universities didn't say anything about registers. On the contrary, they told about distinct surveys, and how statisticians had to make absolutely clear in advance exactly which tables had to come out as the result. So this was the conventional wisdom; how could the new paradigm be acceptable, according to which the statistics should be designed so as to be able to accommodate also unforeseen, future needs?

The internal discussions regarding the paradigm shift was hot 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 basis 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. Discussions with the EU statistical office Eurostat were difficult for many years. Statistics Denmark of course wished to fulfil the European obligations by using the results from the register based system. In principle this was OK, 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 in the European statistical system. Many colleagues expressed 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?[12] 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, touched upon in the following paragraph. Several countries reported on bad experience 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 are now endeavouring to create similar systems. It was a great step towards acceptance, when Eurostat supported the translation of, and published the book *Statistics on persons in Denmark – A register-based statistical system*[13].

## 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 had right of access to register data kept by other public authorities. However, it was not necessarily so that this was 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 problems of public opinion 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 use of registers, as one would have expected; the reason why statistics were aimed at was probably that many people do not understand the usefulness of statistics, while public administration is seen as necessary. 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 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.

[12] Nordic Statistical Secretariat (1983), the actual wording given here sharpened according to the author's memory
[13] Statistics Denmark and Eurostat (1995)

## 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. On the one hand, 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. On the other hand, problems may arise for statistics on the number of events, in deciding which changes in data values are to be viewed as reflecting actual events and which changes merely represent 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 of course 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 chained together.

## 4.5 Functional organisation

In the first decades, the functional organisation was not adopted in Statistics Denmark, as outlined by Nordbotten: "Today data for the survey of manufacturing and the wage statistics are collected separately. It will probably in many cases be more efficient to integrate the collection, *i.e.* collect as much as possible from a respondent when we have made a contact instead of repetitive visits or requests".

This way of working would imply the separation of the data collection as a function to be taken care of by someone else than the compilation of statistics, and this ran counter to a widespread belief about how to organise statistics. However, during the last five years, aspects of the functional organisation are gradually gaining ground. Especially in business statistics data collection is becoming a speciality. This goes hand in hand with the increasing use of electronic reporting using XML, encouraged by the government.

## 4.6 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 about uses of the data that we never thought of. 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 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.

The metadata must at the very least contain descriptions of methodology, definitions of concepts used, definitions of variables and classifications. The metadata must cover all archive data and be maintained 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 on the strategy of Statistics Denmark.

## 5. Concluding remarks

In his first presentation of his ideas, Nordbotten stated: "In 1970, there will probably not be any need to collect a census in the traditional way"[14]. Today we realise that he was far too optimistic, although of course the statement can be interpreted in different ways. It is true that, e.g., the 1970 census in Denmark was not collected in the traditional way as it made use of data from the CPR, but every citizen still had to fill in questionnaires. The full transition to a register based census was not completed until 1981, see Annex. In the other Nordic countries the transition took even much longer, and outside the Nordic region there still seems to be a long way to go.

The reasons why the optimism of 1960 could not be realised are many, some of them explained in the preceding paragraphs. Many factors, human as well as technical and economical, contributed to the longer development time. One crucial factor has been confidentiality and the need to make sure the public accepts the development in all stages. We must also admit that even today, we have not yet realised the full potential: We still have a tendency to think "in boxes", considering only part of the available data at a time and not seeing how we can create new important knowledge by combinations of data. There is still scope for more work on realising Nordbotten's vision.

## References


Danmarks Statistik (1970-1980): *Arbejdsplaner (Work Plans)*, 1970-1980. Statistics Denmark, Copenhagen.

Danmarks Statistik (1979): *Dødelighed og erhverv 1970-75 (Occupational mortality 1970-75).* Statistical Inquiries no 37, Statistics Denmark, Copenhagen 1979.

Danmarks Statistik (1982): *Personstatistik på registergrundlag (Register Based Statistics on Persons).* Copenhagen 1982

Danmarks Statistik (2006): *Act on Statistics Denmark*. In *Statistical Code* (Lovsamling om statistik), p. 127-130. http://www.dst.dk/pubpdf/10035/natlegi. Copenhagen 2006

Jensen, P. (2000): *Dansk Statistik (Danish Statistics).* Vol. 1-2. Statistics Denmark, Copenhagen 2000


[14] Nordbotten (1960) p. 43

Jensen, P. & Thygesen, L. (1985): *Linkage of Records on Objects of Different Kinds. Methodological Problems and Practical Experience*. Bulletin of the International Statistical Institute, Vol. LI, Book 2. Amsterdam.

Lynge, E., & Thygesen, L. (1990): *Occupational Cancer in Denmark*. Scandinavian Journal of Work, Environment and Health. Vol. 16, suppl. 2. Helsinki 1990.

Nordbotten, S. (1961): *Elektronmaskinene og Statistikkens Utforming i Årene Framover* (*Computers and the Future Form of of Statistics*). The Statistical Conferences of the Nordic Countries in Helsinki 1960. Statistical Reports of the Nordic Countries, Vol. 7, p. 135-141. Helsingfors 1981

Nordbotten, S. (1966): *A Statistical File System*. Statistisk Tidskrift 1966:2. Stockholm

Nordbotten, S. (1967): *Om arkivstatistiske systemer*. Annex 4, in: Statistical Reports of the Nordic Countries, Vol. 14. Copenhagen 1967.

Nordic Statistical Secretariat (1977): *Censusproblematikken. Indledning fra Danmark (The Census challenge. Introduction from Denmark)*. The Meeting of the Chief Statisticians of the Nordic Countries in Reykjavik 1976. Statistical Reports of the Nordic Countries, Vol. 32, p. 11-20. Copenhagen 1981

Nordic Statistical Secretariat (1981): *Anvendelsen og afhængigheden af administrative registre i statistikproduktionen (The Use of and Dependency upon Administrative Registers in the Production of Statistics)*. The Meeting of the Chief Statisticians of the Nordic Countries in Copenhagen 1979. Statistical Reports of the Nordic Countries, Vol. 39, p. 28-36. Copenhagen 1981

Nordic Statistical Secretariat (1983): The Joint Meeting of the Nordic Statistical Societies at Elsinore 1979. Statistical Reports of the Nordic Countries, Vol. 40. Copenhagen 1983

Nordic Statistical Secretariat (1990): Det 18:e nordiska statistikermötet i Esbo 1989. Hundraårsjubileum. Nordisk Statistisk Skriftserie Nr. 53. Helsingfors 1990.

Ohlsson, I. (1964): *EDB-systemets betydelse för statistikproduktionen och statistisk analys.* The joint meeting of the Nordic Statistical Societies in Copenhagen 1964. Statistical reports of the Nordic Countries Vol. 12, p. 21-35. Copenhagen

Skak-Nielsen, N. V. (2010): Personal communication

Statistics Denmark and Eurostat (1995): Statistics on persons in Denmark – A register-based statistical system. Eurostat, Luxembourg 1995.

Statistics Denmark and Eurostat (1995): Statistics on persons in Denmark – A register-based statistical system Luxembourg 1995

Thygesen, L. (1978): *Personal Identification Numbers and Population Statistics in Denmark*. Statistics Denmark, Copenhagen 1978.

Thygesen, L. (1983): *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.

## *Annex. Milestones in the development of Register Statistics since 1966*[15]

Statistics Denmark's activity and the main changes and plans in it are detailed in the institution's plan of work, which is published each year. 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, enhancing the institution's independence and strengthening access to administrative records of public authorities.
- With effect from 1968 a statistic was established covering turnover and other factors in all economic activities, based on information from the 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 public administration. It was managed by the Ministry of the Interior. It maintains basic information such as name, address, civil status and family relations and a unique identifier, the Person Number (PN).
- In 1971 a corresponding statistic was introduced covering employment in business enterprises and, from 1974 on, 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.
- At 1 May 1970 total figures were produced for the first time, based on the CPR, of the population in the individual municipalities broken down according to sex, age and marital status. In the period 1970-73 the other population and vital statistics were reorganised on the basis of the CPR, so that manual reporting from the local authorities could be dispensed with.
- From 1970 on, income statistics were switched to a register basis in conjunction with the introduction of taxation at source. From 1976 the statistics were extensively reorganised, amongst other things by linking the income data with data from a 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 taken in 1971, Statistics Denmark in 1973 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 (see Act on Statistics Denmark). Instead, a register-based population census was undertaken for 1976 on the basis of CPR data combined with a range of other register data relating to various aspects of economic activity.
- From 1977 on, a statistic relating to the Social Assistance Act was established based on person-oriented information from the Joint Local Authority Finance System, supplemented

[15] Adapted from Statistics Denmark and Eurostat (1995)

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 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, which refused to provide data. In 1981 Statistics Denmark won its case at the Supreme Court.

- At 1 April 1977 the first figures for buildings and dwellings were produced on the basis of the newly established Central Register of Buildings and Dwellings (BBR). At 1 January 1980 the first register-based housing census was conducted, combining data on homes with data on the occupiers.
- 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 contained micro data on a sample of the population including a broad range of data for calculating the effects of contemplated legislation.
- At 1 January 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 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, incomes, 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 development work, an integrated, modernised business register system was established with a number of content-related and technical improvements.
- In 1994-95 an extensive reform of wage statistics was carried out which will facilitate the production of comprehensive and consistent person-oriented figures for all wage-earners in both the private and the public sector.