

DATA PROTECTION IN A NATIONAL REGISTER-BASED STATISTICAL SYSTEM

L. THYGESEN

Danmarks Statistik

SUMMARY

The main source for social and demographic statistics in Denmark is data from administrative registers. A coherent statistical system has been developed on this basis since 1970 and was used, for example, to carry out the 1981 census of population and housing. For this system to function, the statistical service must have access to identified personal data from registers and a common, unambiguous personal code number must be used.

To establish and maintain such a system, great importance must be attached to a data policy which provides the public with a guarantee that personal data will not be misused. In this paper a number of problems relating to data policy are discussed in the light of Danish experience.

0. INTRODUCTION

It has been recognized for very many years that information on individuals must be used when compiling social statistics (e.g. results of population censuses). This has been necessary to enable the basic data to be checked and any errors found in them rectified. From an early stage, statisticians were aware that this required protection of an individual's identity.

The development of data processing methods over the past few decades has widened the scope for using personal data for statistical purposes and, in so doing, made it possible to analyse social phenomena which previously could not be explored.

At the same time, however, there is in most countries increasing concern about the uses to which computerized personal data may be put. This is mainly attributable to the fact that public authorities have created large personal registers in order to be able to administer legal provision of ever increasing complexity. The existence of registers

has made the individual citizen afraid that he might be kept under surveillance and has led to demands for restrictions on the use of the registers.

Although statistics consist only of statements about groups of people and not about individuals, the use of personal data in statistics has not gone unnoticed in the public debate. On the contrary, statistics have been one of the favourite targets of criticism, perhaps because it has not always been understood that they can serve a useful purpose.

In Denmark the use of administrative registers for statistical purposes is probably more widespread than in any other country; nowadays virtually all the personal statistics compiled form a coherent system based on the administrative sources. The whole system is founded on the coordination of data using common personal identifiers.

A statistical system like that in Denmark focuses attention on problems of data policy. It is vital that the statistical services have access to the administrative registers and this access can be ensured only if the public remains confident that personal data can be used for statistical purposes only.

Starting with the Danish experience, this paper discusses the problems associated with data protection in a statistical system which is based mainly on administrative registers. By way of introduction, Section 2 briefly examines the principles underlying the Danish statistical system. This is followed by a discussion of the aims of data policy and the means which can be used in a number of practical problems related to dissemination of statistics, data collection and data processing.

1. THE COMPILATION OF STATISTICS IN DENMARK

The compilation of statistics in Denmark is more centralized than in many other countries. The central statistical office (Danmarks Statistik), which is responsible for most social statistics of a general nature, is an independent public body which was set up under a special law.

Since the beginning of the 1970s, one of the aims of the strategic planning undertaken by Danmarks Statistik has been to set up a coherent system of personal statistics based on information from administrative registers kept by various authorities. This stems from that fact that the actions of

public administrative authorities in Denmark are based to a large extent on information in registers relating to the "objects" of administrative action, e.g. citizens, businesses and buildings.

1.1. The administrative registers

In 1924 a statutory duty was laid on every municipality in Denmark to set up a local population register, i.e. a file containing information about all persons living in the municipality. These registers were to contain identifying information such as occupation, name, birth date and place of birth. Apart from these, the most important items of information were the address, family circumstances, nationality.

The municipalities had to continuously update the files using information on births, marriages, deaths, etc. obtained from various public authorities while the individual citizen was obliged to report any changes of address directly to the registration office.

A major reform of the population register system took place in 1968. The municipal registers continued in being but in addition a central population register, or CPR, was created. This is a computerized register covering the whole of the Danish population. The central register and the local registers are updated as part of one coordinated administrative process.

An important part of the reform was the introduction of a permanent and unique identification number for every citizen: The Person Number. This number was regarded as a practical necessity for the operation of the central population register. In addition, it was to be introduced in every area of public administration, thus replacing the large number of numerical systems which had hitherto been used by the various administrative departments.

The main reason for setting up the CPR was the wish to avoid duplicate registers and the use of extra resources which that involved. Another reason was the prospect of a tax reform involving the introduction of a PAYE system, which would be difficult to operate without a very reliable system for identifying persons living in Denmark.

Information from the CPR is used by the public administrative bodies in almost all areas relating to the individual citizen. This means that there are many opportunities to identify and correct or remedy errors and defects in the information contained in the register.

In the years following the creation of the CPR, the use of computers by the Danish authorities greatly increased and large personal registers were created to administer the collection of taxes, the payment of pensions, etc. All these registers, which are valuable sources for statistics, use the Person Number as identifier.

In 1977 a law was introduced which set up a nationwide buildings and dwellings register to be used by the municipal authorities. It was thought that the information in the register on the size and layout of dwellings could be used as a basis for future population and housing censuses. Special care was taken to ensure that the details of addresses given in the CPR and in the building and housing register corresponded exactly to one another so that information on individuals and information on housing could be linked automatically. Experience has shown that this system offers a high degree of reliability.

Finally, mention should be made of the central business register, which contains basic data on both enterprises (legal units) and establishments (local units). The register, set up under a statutory provision in 1975, is kept by Danmarks Statistik. The law in question presupposes the existence of a unique numbering system for establishments which can be used by public administrative authorities and other bodies.

1.2. The principles underlying the system of personal statistics

The Danish system of personal statistics has developed progressively since 1970 in parallel with the creation of the administrative registers on which it is based. The first step was to reorganize the annual vital population and its movements, with information broken down by sex, age, place of residence (municipality), etc. Statistics on income, employment, etc. then followed. The basic development work was not completed until 1981.

The system's effectiveness was demonstrated by the 1981 population and housing census which was carried out without sending questionnaires to the public but solely by collating information already available in the system. This kind of census could, in principle, be carried out every year.

The statistical system is concerned mainly with persons. It also contains information on the dwellings where these persons live and on the places of employment where they

work.
by mee
number
for tl

Each
emplo
const
situa

The
of t
chan
enab
will
trac
man

The
(pe
rel
ind
is
edu
th
Tw
st
su

Mc
ar
co
m
c
p
o
c
c
i

2. DATA POLICY OF THE CENTRAL STATISTICAL OFFICE

The type of statistical system established in Denmark requires that the statistical processing of the registers takes place centrally.

It is often suggested that the administrative authorities which have control over information should themselves carry out statistical surveys, thus avoiding the need to pass on personal data. One point in favour of this approach is that those keeping the register are normally themselves among the most important users of these statistics and therefore know what is required and, at the same time, are very well acquainted with the characteristics of the register and are therefore in the best position to update it. But, as was indicated in the previous section, if administrative information is to be made generally suitable for statistical purposes, it is essential that the information from several sources be combined at an individual level. The data policy of the Central Office must therefore guarantee access to personal data.

Since public opinion and policy-making bodies are very much on their guard against the unauthorized use of personal data, the recognition that data protection is vital for the central statistical office must be the starting point when devising a strategy for data policy. It is essential that the general public and politicians have no misgivings whatsoever about handing over personal data to statisticians.

A leak of personal data would naturally have disastrous consequences for the work of the statistical service. The suspicion that a leak might occur must be clearly dispelled. Indeed, there must be no possibility of one occurring. Experience has shown that even totally unjustified suspicions which are publicized in the news media can be very harmful and difficult to allay.

It is not sufficient for management to recognize the importance of data protection and lay down rules regarding which personal data may be used and how they should be processed. Every member of staff at the central office must be made aware of the importance of data security. It is here that data security must be guaranteed.

In day-to-day work involving the handling of data and contact with customers, decisions must be taken on many questions which may seem trivial or innocuous but which, when combined, constitute the data protection of the

statistical office. It is a difficult task for a management to ensure that every one understands and accepts the reasons for the data protection rules and that they take them seriously.

If this aim is to be achieved, the central office must pursue an active data policy. It is dangerous if data protection provisions are regarded as a necessary evil which is forced on the statistical office from outside, e.g. by a data inspection board. The central office must therefore take the initiative itself and lay down data security criteria. This must be done in the knowledge that the precautions may involve extra expenditure and inconvenience, which have to be weighed against the benefits obtained.

It is important that data security should not be regarded as something static, i.e. as a set of rules laid down once and for all. In the first place, observance of the rules must be continuously monitored or the rules will gradually lose their importance and come to be regarded as ridiculous. Responsibility for this monitoring work should be assigned according to the division of responsibility in the various departments of the organization. In addition, the effectiveness and suitability of the rules must be kept under constant review so that any necessary amendments can be made.

While data protection within the organization must be maintained at a high level, it is management's task to inform the public and those who make decisions about what is being done. Many people have wildly exaggerated ideas about what data Danmarks Statistik handles and what these data can or must be used for.

The first step is therefore to ensure that the public really has no reason to feel anxious because the statisticians deal in a careful and considerate way with confidential data ("internal" marketing of data security). Next, data security must be publicized "externally" i.e. steps must be taken to ensure that the public is not afraid of the statistical organization. Efforts must be made to create an understanding that statistical registers are not harmful if effective data protection is provided. The organization cannot be "sold" externally unless it has its own house in order.

Data protection is at one with the main aim of the central office's work, namely, to give users of statistics the best possible service. There will often be an apparent conflict between the two aims. But each must be given due weight if the central office is to continue to have access to the

administrative registers. If the statistics produced by the central office cannot be regarded as effective or useful, it is no use claiming that the data protection provisions work well.

The following subsection contains a discussion of the balance to be achieved between these two aims in the case of the dissemination of statistics, data collection and data processing. First, a short account will be given of certain aspects of Danish data protection legislation.

2.1 Legislation on registers

At the end of the 1970s, as in many other countries, Denmark introduced legislation on registers which lays down general rules for the establishment and management by public authorities of computerized registers containing personal data. The law specifies that a Data Surveillance Authority ("Registertilsyn") should be set up with the task of supervising registers and the special data protection rules laid down for each individual register.

Under the law, every instance in which several registers are linked must be notified to the supervisory board, which can lay down conditions governing the way in which this record linkage is made.

In some respects the law lays down special rules for registers used solely for the compilation of statistics. Thus, registers can be linked without notifying the supervisory board if it is done for purely statistical purposes. In addition, the general right of citizens to know what information on them is registered is waived in the case of statistical registers. The reason for this is that statistics cannot intrude on an individual's privacy since information on individual citizens cannot be identified in the tables and because personal information in statistical registers may not be used for administrative purposes. In this connection, it should be pointed out that the law forbids the passing on of personal information from statistical registers. This is a statutory formulation of a principle which has been observed by Danmarks Statistik for many years.

By and large the law has made it easier for Danmarks Statistik to carry out its work. Clear rules have been laid down for a number of questions, and it is an advantage for Danmarks Statistik that it does not provide the only guarantee that information will not be misused.

any statistical work. The possibility of providing very detailed information which cannot be published increases the risk of such disclosures taking place unless measures are taken to prevent it.

The following measures might be taken:

- limiting the number of recipients; data should be supplied only for a specific purpose and be regarded as confidential by users; any analytical results intended for publication must be submitted to the statistical office in advance.
- limiting data contents; the degree of detail must be defined on the basis of a concrete assessment of the aims specified and of the possibility that users might recognize the "subjects"; thus, for example, sample data might be supplied instead of full information;
- the suppression of certain data or the introduction of "interference" in the data without fundamentally altering the results which can be derived from them (see for example "Report on Statistical Disclosure and Disclosure-Avoidance Techniques, US Department of Commerce 1978): this procedure, however, has not been adopted by Danmarks Statistik.

Moreover, the supply of very detailed information gives rise to other problems apart from those relating only to data security. There is a risk that users might publish figures which are not statistically reliable or are difficult to compare with the official statistics.

An example of how the supply of detailed statistical information is dealt with in Denmark is provided by the so-called "lovmodel" (law model) which Danmarks Statistik has developed in collaboration with the Ministry of Economic Affairs.

The "law model" is a joint EDP system which is designed to provide rapid and reliable statements about the immediate effect of legislation on income distribution and budgetary matters, especially as regards taxes and subsidies. The term "immediate consequences" is used here because the model does not take account in its calculations of the secondary effects of amendments to legislation arising from the fact that citizens base their behaviour on existing conditions (e.g. changes in the system of welfare benefits affect the recipient's consumption patterns), which in turn has consequences for other citizens and, hence, for other areas covered by legislation.

An important task for a central statistical office is to supply information which enables Parliament, the government and the administrative authorities to calculate the

consequences of a particular piece of legislation. This applies not least to calculations made when preparing new legislation or amending existing laws. In such situations it is necessary to examine the ways in which legislation will affect both individuals, the family or groups of persons as well as society as a whole, where, for example, it is useful to know the overall effects of legislation on budgetary matters. As a rule, the calculations have to be made in a hurry and require a highly flexible statistical basis. At the same time it is desirable to be able to analyse the way in which one law interacts with others.

Danmarks Statistik has established the statistical basis of the law model using the statistical registers to form "model populations". The information exists in the form of information on non-identified persons or families forming a representative cross-section of the population. Each model population is intended to be used within one or more legislative areas and thus there is no universal model population which can be used for calculations relating to all the possible consequences of legislation. At present there are 15 model populations.

The system also contains a computerized model of existing and contemplated legislation on taxes and subsidies.

The users (at present 10 Ministries and administrative departments) have access via terminals to Danmarks Statistik computer installation where the law model is run. The users carry out their model calculations themselves but can draw on the systems' joint facilities, including the model populations.

Access to the model is given only to central government and the model cannot be used to produce statistics.

2.3. Data collection

Since this paper is mainly concerned with statistics based on administrative registers, the question of information on respondents obtained in interviews or postal inquiries will not be discussed here. However, problems of data policy also arise when data is collected from administrative authorities.

Danmarks Statistik has a general legal right to ask for information from the authorities. However, the question as to the form in which the information can be requested has been a subject of some debate. Thus, at the end of the

1970s, certain municipalities claimed that Danmarks Statistik was not entitled to ask for information about individual recipients of social assistance on the grounds that this information was protected by confidentiality provisions. The information in question was to be used for a new set of social statistics and the method of collection was devised so as to minimize the amount of reporting required of the municipalities. The case, which led to a great deal of public debate, was settled by the Danish Supreme Court, which clearly upheld Danmarks Statistik's contention that the information should be provided.

There is also the question as to what types of personal data Danmarks Statistik should collect. Of course, Danmarks Statistik's position on this question must depend in part on the type of statistics which users would like to see produced. In addition, however, the law on registers lays down direct restrictions on the information that can be recorded on a person's political and religious beliefs, etc.

2.4. Data processing

One of the main purposes of data policy is to prevent the basic data over which the statistical service has control from falling into the hands of unauthorized persons or being used for purposes other than those originally intended. The dangers arising here range from various types of accidental leaks of personal data due to careless gossip or negligent handling of information on the part of an official to attempts at industrial espionage or intelligence gathering. The probability that an attempt will be made to gather intelligence must be regarded as negligible since Danmarks Statistik does not possess information of great interest to intelligence services. On the other hand, there is a real risk of accidental leaks. Even though such a leak, if it occurs, will not normally harm the individuals covered by the information, it is important to take steps to prevent this happening. This also applies to the risk of a "political" attempt aimed at proving that the statistical service does not take sufficient care of its personal data; if an attempt to steal personal data were successful, the harmful effects would be incalculable. To provide effective protection against these risks, various data security measures have to be taken. However, the most important thing is to recognize that there is a need for vigilance in every part of the service and that the problems may appear to be trivial or absurd to individual members of staff if they cannot be convinced of their importance.

As regards electronic data processing, the following special security precautions have now become standard practice:

- restrictions on and control of physical access to certain areas, e.g. to the magnetic tape library;
- the authorization of persons who will have access to files with confidential information;
- systematic checks on the use of files;
- rules regarding the dispatch of magnetic tapes.

At Danmarks Statistik these rules are supplemented by very stringent provisions on the handling of files containing particularly sensitive personal data. Thus, all personal identifiers in a register of diagnosed cancer cases which is being used for an epidemiological analysis must be encoded and may be retransformed to obtain the person numbers only for those few individuals whose personal data have to be corrected.

Introducing rules of this kind inevitably involves a great deal of extra work and, hence, an additional strain on resources. However, the problems involved in safeguarding lists of errors and other transcripts containing personal data which are used by the staff in various statistical departments are much greater. Such material may be made accessible only to persons who need to examine it during the course of their work. The material must therefore be locked up when it is not being used and the doors of the offices concerned must be locked when no one is inside.

Contrary to what one might expect at first sight, the greatest security risk is associated with manual data processing. Here, maximum care must be taken when handling confidential information.

3. CONCLUDING REMARKS

The way in which official statistics are organized varies from one country to another. This is due, *inter alia*, to differences in legislation or the way in which society is generally organized and this includes the mode of operation of the public administrative authorities. There are also differences, determined by past history, in the public's attitude towards statistics and the registration of individuals.

The data policy of each statistical service must, of course, adapt itself to the conditions existing in the country concerned. The experience gained in Denmark is therefore not quite the same as that gained in other countries. Readers will doubtless be familiar with many of the thoughts on data policy which have been presented in this paper and which

involve the need to reconcile conflicting considerations. There is a common interest in ensuring that personal data are protected and that the interests of individual citizens are not put at risk by statistical activities, although there is no clear answer as to how this should be done.

* * *

DISCUSSION

Mme. LENOIR (*Commission Nationale de l'Informatique et des Libertés*) opened the discussion:

French statisticians were dazzled by the possibilities for statistical analysis offered by the Danish system. In Denmark the manipulation of the various registers for the compilation of statistics was concentrated in Danmarks Statistik while there were, presumably as in France, many requests from local administrations and private bodies for statistical data. It would be interesting to know whether all such requests had to be channelled through Danmarks Statistik as well as whether the linking of the registers was done solely for statistical purposes or whether there were cases in which this might have use for administrative purposes. Finally she asked for some information of a historical nature on the steps involved in setting up the Danish system.

Mr. RAPAPORT (*Statistics Sweden*) remarked as follows:

In Sweden they received between 20 000 and 30 000 requests per annum for extracts from the central register from people wishing to check the information about themselves. Initially the authorities were apprehensive about this feature in the operation of the register but later were very much in favour. People were able to check for themselves that there was nothing sinister in the register and their usual comment was "But I knew that already". There was no quick way of gaining the public acceptance required to operate a central register system effectively. This had to be built up by a policy of creating an informed public opinion over a long time.

Professor FLAHERTY (*University of Western Ontario*) raised the point:

There had been no big debate about the 1981 census in Denmark. In view of the intense debates in the adjoining countries was the 1981 census carried out so discreetly that it aroused no public attention?

Dr. BAUMANN (*der Bundesbeauftragte für den Datenschutz*)
remarked:

It was clear that the most important source of population statistics in Denmark was the administrative register system, taken in conjunction with a unique identification number for each citizen. In the case of the Federal Republic of Germany this form of statistical inquiry was ruled out by the judgment of the Constitutional Court concerning the census. The Federal Constitutional Court dealt with this question in connection with the examination of whether or not there was at present any simpler alternative to the full enumerations of the population. Its rejection of the Danish solution was based on the fact that the use of data from different registers required that technical, organizational and legal measures be put into effect before these data could be collated. This would, according to the Court, be a decisive step in that the individual citizen's entire personality would be registered and catalogued. The collating of available data was, therefore, in no way a simpler method.

Mr. CARIANI (*Istituto Centrale di Statistica*) asked:

Could further information be provided about the problem of updating certain items of information in the register notably those relating to occupation.

The author replied to the debate:

The local administration was not entitled to obtain information about identifiable individuals from the register, while applications from the courts had also been successfully resisted by Danmarks Statistik. As regards the history of the Danish system, this had its origin largely in ideas developed in other countries, in particular those of the Norwegian statistician, Professor Norbotten in the 1960s. When Danmarks Statistik was founded, Parliament proposed that it should make use of administrative data from other bodies, so that the lines of work were established. Individuals were not entitled to know what was registered about them as in Sweden. Less than 100 requests for such information were received in a year. The cost of instituting such a system would be very considerable and it would make the risk of leakage of personalized data somewhat greater. Considerable efforts were made to inform the public about the census so as to illustrate the use made of the register. However the news media showed little interest. The data Surveillance Authority was neither particularly in favour nor particularly against a census being carried out. There

had been some public debate but this occurred in the 1970s when the law on public registers was being passed. Some mention of this was made in the paper.

The question on updating occupation was one of the big problems of the register system. However, they did have information on place of work and the activity carried on there. There was also information on occupation from tax registers and unemployment registers. With a view to improving the occupational data possibilities based on these various sources were being examined.