

Confidentiality in the use of administrative data sources

Riitta Harala and Anna-Leena Reinikainen
Statistics Finland, FIN-00022 Statistics, Finland

According to the Statistics Act, Statistics Finland has had almost unlimited access to administrative data sources since 1970. The use of administrative data has increased systematically since then to the extent that the 1990 population and housing census was totally based on administrative data. Many years of experience in using administrative data has shown that the use of administrative records is the most effective way of rationalising data collection but also that it is essential to have a legislative base which governs their use.

The legal obstacles to the use of administrative data in many countries are mostly explained in terms of protection of the confidentiality of such data. These arguments, however, are not connected specifically with the use of administrative data any more than with traditional data collection as in population censuses. In fact the use of administrative data can offer distinct advantages over traditional questionnaires in terms of security as well as cost and rationalisation.

1. Confidentiality and rationalisation of data collection as starting points

The systematic use of administrative data sets for the compilation of statistics began in Finland almost 30 years ago. Extensive use was made of administrative registers for the first time in the census of 1970, some of the data for which were derived from administrative sources and some collected directly from the public. The use of administrative data has been increased systematically since that time, to the extent that the 1990 census, for example, was based entirely on this source.

The use of administrative data sets is the most effective way of rationalising data collection in the compilation of statistics. Among other things, data collected on one occasion can be reused, which substantially reduces data collection costs. Provided it is done properly, the use of such materials is also a better means of gathering data than direct acquisition as far as confidentiality is concerned, for one aspect of personal privacy is that respondents should not be harassed with unnecessary inquiries. The compilation of data from administrative sources also enables the number of persons engaged in collecting and processing the data to be reduced and the data to be better protected technically than in the case of collection directly from respondents. The most difficult confidentiality problem when using admin-

¹ This paper was first presented at the Third International Seminar on Statistical Confidentiality in Bled, Slovenia, 2-4 October 1996.

istrative data sets is that of informing the persons concerned of the fact that such data sets are being used for the compilation of statistics, but there are means available for doing this.

2. The role of legislation

The use of administrative data sets calls for the creation of the necessary legislation, as it would be difficult or even impossible to use such sources successfully if it is not viable in practice through legislation. Similarly, the legislation should give the officials compiling statistics sufficient powers to extract data from administrative sources. The legislation currently in force in Finland supports the statistical use of this material extremely well.

2.1. General data protection legislation

The general legislation governing the security of personal data in Finland enables information collected for other purposes to be handed over for the compilation of statistics and for research, a principle which is also contained in the general EU data protection directive. It also allows data derived from administrative data sets to be combined for statistical purposes, as this is not considered to endanger the privacy, interests or rights of individual persons.

The use of personal identification numbers in Finland has enabled administrative data sets to be used extensively and data derived from these to be combined. It is common, in fact, for some of the data used for compiling a certain set of statistics to be derived from administrative sources and some from personal interviews, in which case identification numbers are used to match the two. The data protection authorities have approved an interpretation of the law under which the use of identification numbers is recognized as necessary, and indeed essential, for combining data derived from different sources. Similarly, identification numbers are retained in other sets of data required for future use, such as census results, although in this case they are used in encrypted form.

It is laid down in the general law on data protection that individual citizens shall have no right of access to their personal details as contained in the records of the statistical authorities. Such access has not been considered necessary for the reason that the statistical authorities do not make decisions regarding the interests or rights of individual people. Although such a right would admittedly increase the opportunities for individuals to obtain information on the use of their personal details for the construction of statistics, it would at the same time imply a major confidentiality risk, as it would presuppose the creation of a system in which data on one person at a time could be recovered rapidly and easily. The principle of openness in government has been considered to be observed sufficiently well by the statistical authorities if they inform the general public of the purposes for which sets of personal data are being used and of the information content of the resulting statistics.

2.2. The Statistics Act

The new Statistics Act which came into force in Finland in 1994 marks a conscious attempt to rationalise data collection and to treat respondents more fairly. It also takes better account of the use of administrative data sets and the problems connected with this in its regulations governing the confidentiality of information.

2.3. Principles of data collection

It is laid down in the Statistics Act that use should be made primarily of data collected on previous occasions, i.e., measures should always be taken before commencing data collection to ascertain whether the information required can be obtained from existing administrative or other sources. The law also requires that the necessary information should be collected without personal identification codes whenever possible. This is regarded as impossible, however, if the aim is to combine the results with information obtained from other sources, a process which is regarded as requiring the use of social security numbers or the registration codes of businesses or corporate bodies.

The law lays special emphasis on an extensive preliminary examination of information needs and the eliciting from respondents of only such information as is essential for the compilation of the statistics. The data should also be collected in as economical a manner as possible, so as to cause the minimum inconvenience and costs to the respondents. The law requires in this connection that before any decision is made regarding an inquiry, the instance responsible for collecting the data should negotiate with the respondents over the type of data to be collected on the basis of general obligations to furnish the authorities with information, the times by which the data should be provided, the data collection procedures to be used and the feedback data to be given.

2.4. Provision of information for respondents

The law states that respondents should be informed of the purpose for which the data will be used, the procedures by which the statistics will be compiled, any particular principles governing dissemination of the resulting information and other factors of relevance. In the case of administrative data sources, this information should be communicated to the authority responsible for handing over the materials.

It should be noted, however, that the law does not require that each natural or juridical person should be informed of the use of administrative data sets for the statistical purposes, as this would often be impossible in practice. If administrative data are to be supplemented with interview data, the interviewer should inform the respondent of the other data sources from which his or her personal details will be taken for compilation of the statistics concerned. If it is sufficient to use data from administrative sources alone, information can be conveyed to the subjects as a group through the media. Statistics Finland distributes information on its main

surveys to be based on administrative data sets (e.g., population censuses) through the press, TV and radio. Attempts have also been made to ensure that the legislation governing the administrative authorities that hand over data for statistical purposes also makes specific mention of this fact, so that information on the use of such sources will spread in this way. It can already be said that the Finns are quite well aware of the statistical use being made of administrative data and have largely accepted this practice.

The obligation to keep respondents informed of the use of their data also includes furnishing them with feedback data on the statistics compiled in this way, although this is usually relatively narrow in scope and concentrates on those items of greatest relevance to them.

2.5. Obligation to provide data

The use of administrative data sources has the practical consequence of limiting the direct obligations upon citizens to disclose information for statistical purposes, so that nowadays this obligation applies to government authorities, employers, self-employed persons and unincorporated public enterprises, local government authorities, non-profit institutions and pension funds, but all information collected from natural persons is provided on a voluntary basis, i.e., they are no longer under any obligation in this respect, as all the data required, even for population censuses, can be obtained from administrative sources.

The obligations upon government authorities to disclose information to Statistics Finland are extensive, and apply irrespective of the confidentiality regulations which are otherwise binding upon them. They apply to all information that is regarded as essential for statistical purposes (including information of a sensitive nature) and data on the operating of these departments, agencies, etc., their finance and their activities, other than information which is classified by reasons of national security or defense. Extensive data collection rights of this kind are essential if efficient use is to be made of administrative data for the compilation of statistics.

2.6. Confidentiality

The confidentiality of the information possessed by administrative authorities in Finland is likewise regulated by law. It is an essential part of the Finnish and Scandinavian notion of democracy that such information should be as open and public as possible in order to guarantee citizens the right to supervise the operation of the public administration. This means that practically anyone has the right to examine the public information possessed by government offices. A revision of the law is currently being prepared in Finland which will further increase public accountability, particularly as regards the preparation of matters brought before decision-making bodies.

It is often claimed that information derived from public sources should continue to be of a public nature when in the possession of the statistical authorities. Experiences gained from the use of extensive bodies of such data nevertheless

suggest that it is impossible to transfer such a far-reaching principle of openness to a statistical organisation even though that organisation may be making use of public administrative data. Confidentiality is an integral part of the image of a statistical office and is a precondition for the acquisition of sufficiently comprehensive and reliable data for statistical purposes. The individual citizens and companies that serve as respondents evaluate the office's 'confidentiality image' as a whole without distinguishing between public and confidential data on the grounds of the type of source used. Thus an excessively open attitude towards information derived from public sources could be detrimental to this image and seriously hamper the collection of data for statistical purposes. Consequently the Statistics Act lays down that administrative data obtained by Statistics Finland for the compilation of statistical tables shall primarily be deemed confidential upon passing into that organization's possession. Exceptions to this concern data describing the activities of central and local government authorities and data contained in the Business Register, which remain public even in the possession of Statistics Finland. This has been found in practice to represent a generally acceptable level of public accountability which the image of the statistical authorities in Finland can tolerate without it detracting from the acquisition of data.

3. Confidentiality advantages and drawbacks arising from the use of administrative data

The legal obstacles to the use of administrative data in many countries are often explained in terms of protection of the confidentiality of such data, and prejudices still persist as regards the use of personal identification numbers and the opportunities for combining extensive data sets, often stemming from the abuse of statistical information during the Second World War. These ideas are often based on false assumptions and are in many cases not connected specifically with the use of administrative data any more than with traditional data collection.

In fact, the use of administrative data can offer distinct advantages over acquisition by means of traditional questionnaires in terms of security as well as cost and rationalisation.

3.1. Data security advantages

The use of administrative data has a number of advantages with respect to security and confidentiality (see below).

3.2. Reduction in the number of external parties

The statistical institute is capable of operating without having to employ external subcontractors for tasks such as postage and the recording of information from questionnaires, which always entails a certain security risk. The use of outside services has almost always been necessary in the past in connection with popu-

lation censuses, for example, where millions of forms had to be posted and the information on them recorded within a short period of time.

3.3. Reduction in staff handling the data

Even if the staff handling the data are all employed by the statistical office itself, any increase in their number or their dispersal among centres in different parts of the country is liable to compound the security risk. By comparison with the last full-scale census to be conducted on the basis of forms filled in by individual citizens, for which a staff of about 2200 persons was employed in local offices all over the country, the current census methods based on administrative registers can be operated by two planning officers and a project team of 15 persons all working in the same locality. This restriction of the number of people involved implies a substantial improvement in confidentiality.

3.4. Data seen only by the computer

A third advantage with the use of administrative data is that no one person sees all the census data on a given individual in the course of processing, for unlike the contents of a questionnaire form, which will in principle be read and recorded by a single operative, it is only the computer that sees the material in its entirety. In addition, access to data on individual persons can be restricted to a very small proportion of the computing staff, only those who need full rights in order to process the material.

Above all, we should start out from the principle that the statistician is not interested in the personal details of individual citizens but in descriptions of whole population groups and their characteristics. In any case, the data will have been obtained specifically for statistical use, and cannot be put to any form of administrative use in spite of having originally been supplied by an administrative body. Statistics Finland insists very strictly that the passage of unit-level data, whether on forms or in administrative registers, shall take place in one direction only, and that once received by the statistical office, no data will be released in a recognizable form at the individual level. Micro data can be released for research and statistical purposes under certain carefully defined conditions (defined in the Statistics Act and its implementation instructions), provided that no actual individual can be identified from them.

3.5. Data security disadvantages

The disadvantages attached to the use of administrative data sets compared with the collection on questionnaire forms are relatively few in number, for the same potential difficulties that can be perceived in this approach have already existed in the extensive use of questionnaires bearing the respondent's identification number (e.g., census forms). Likewise, the outcome of the data collection process is in both cases a vast body of data which would be equally easy or difficult to misuse if one so wished. Administrative data sets naturally offer the opportunity

to extend the scope of the information collected, and in this sense increase its sensitivity. Statistics Finland takes care to encrypt all social security numbers in both its census data and its annual register-based regional employment statistics in order to ensure confidentiality.

Perhaps the greatest danger of all to the use of administrative data for statistical purposes is posed by the attitudes of the general public. The use of such material and the possibilities for combining the contents of extensive registers of this kind can arouse a fear of infringement of personal privacy in some sectors of the population, as the authorities are in possession of the 'whole picture' of the life history of individual citizens (cf. longitudinal data file from population censuses). Frequently, however, these sentiments are linked chiefly with a fear of administrative action of the 'big brother' type against private persons, which would be utterly impossible using data of the kind that a statistical office possesses. Such speculations have scarcely ever been put forward in Finland, not even in the context of national population censuses, and when Statistics Finland has distributed information on its working methods to the press at the time of performing a register-based population census, the publicity obtained has almost always been favorable. The savings in terms of costs and the convenience resulting from not having to ask the public to fill in forms with information that has already been given to some government office have been regarded as such obvious advantages that no serious discussion has arisen regarding questions of confidentiality. One factor contributing to this has undoubtedly been the good image that Statistics Finland possesses as an office that pays attention to the confidentiality of its data and can therefore be trusted. The preservation of this image is an essential condition for its continued use of administrative data. It must be remembered, of course, that it would take only a small incident to shake this public atmosphere of trust, and that rebuilding it could require a great deal of time. For this reason it is worth committing considerable resources to the maintenance of confidentiality and of a public image of this kind.

4. Conclusion

In summary, it may be stated that the Nordic approach to the use of administrative data for statistical purposes sets out from the notion that it is unreasonable to restrict the statistical use of such data on the grounds of confidentiality. Alongside the many other advantages of the method, it incorporates certain improvements in security, in spite of the fact that the increased amounts of information obtained render the material more sensitive than ever. One essential requirement, however, is the existence of a legislative base that defines the principles that govern both the acquisition and handling of data, the legal obligations upon certain instances to divulge information and the principles governing confidentiality. The implementation of these confidentiality principles at a practical level and the maintenance of appropriate working methods are matters of everyday procedure that should concern everyone working in a statistical office, as they are crucial to the success of the whole operation.

Riitta Harala is Director in Statistics Finland (the central statistical office of Finland) with the responsibility for population statistics. She joined Statistics Finland in 1975. She has worked with population censuses and especially with the register-based population census system and the annual production of register-based employment statistics.

Anna-Leena Reinikainen is Chief Administrative Officer in Statistics Finland. She is the head of the unit which is responsible for administrative and legal services. She joined Statistics Finland in 1971 and has been working with legal and personnel affairs since then.