

Discussion

Statutes and Administrative Procedures

Photis Nanopoulos¹

Abstract: Commenting on the Cecil and Jabine papers provides me with an opportunity to discuss statistical confidentiality in the European Community (EC) and to attempt a comparison between the U.S. and the EC approaches to these very sensitive questions. Cecil describes the U.S. system for data protection and access to confidential data for research and statistical purposes, regarding data collected by the various federal agencies either for administrative or statistical purposes.

Jabine deals with restricted data access as practiced by the statistical agencies to make data available to third parties.

Bringing together the various aspects of statistical confidentiality stressed by the two authors and the experience of the European Community, an effort is made to highlight a few important issues for official statistics in relation to the protection of the privacy of the respondents to statistical enquiries.

1. Statistical Confidentiality

1.1. The principle

The primary objective of a National Statistical System (NSS) is the collection of socio-economic information on the individuals of a population (firms, persons, households, administrations) and from that, the production of *aggregated information*, either on the whole population or on subgroups of the population. Consequently, to this objective comes the *principle of statistical confidentiality* which states that: *no information on an individual, even if in aggregated form, can be disclosed if there is any risk for direct or indirect identification of an individual respondent.*

Statistical confidentiality is fundamentally an ethical principle, common to all NSSs, founded on the declaration of

Human Rights, concerning the requisite secrecy and protection given to the privacy and the anonymity of the respondents.

A promise of confidentiality is usually given by statisticians to respondents and has to be considered paramount to any respectable statistical system.

Since its creation, most National Statistical Institutes' (NSI) legislation has been set up in order to protect the privacy of the respondents and to avoid any misuse of the data. Of course, the way that these legislations handle the problem of privacy protection and data utilisation depends on the way administrations are organised, and on the way and the purpose for which the primary data are collected.

1.2. The evolution

In the last two decades, the subject of statistical confidentiality has experienced increasing interest on both sides of the Atlantic and

¹ Eurostat, Statistical Office of the European Communities, Bâtiment Jean Monnet, L-2920 Luxembourg.

several conferences in the U.S. and Europe have been completely or partially devoted to this and related topics.

The reason for such attention is well known and not very recent. Since the 1960s, the development of computers has raised the problem of individual data protection, access and communication, because such developments have posed a potential threat to the privacy rights of individuals. The threat was perceived as coming from computerisation and other forms of office automation. Microcomputers store vast amounts of information with a corresponding capacity for search and identification.

Consequently, the legislators have focused their interest on themes such as *privacy protection, data protection, right to access, and data disclosure*.

Gradually, in most countries, the old legislation dealing with "professional secrecy" has been replaced with laws better adapted to current data processing capabilities.

1.3. Current developments

Council of Europe: The mandate of the Council of Europe's working party No. 13 is "to examine the legal and technical aspects of the protection of personal data collected and processed for statistical purposes, with a view to drafting a recommendation on that subject."

The starting point of the exercise is the Council of Europe's Convention (No. 108) for the "protection of individuals with regard to automatic data processing of personal data." This convention has led to a series of specific domain recommendations, such as, Recommendation No. R (83) 10 "relative à la protection des données à caractère personnel utilisées à des fins de recherche scientifique et de statistiques." That recommendation, dated 23 September 1983, is now being revised in the light of newer developments.

The definition of "personal data" will observe that statistical results can never totally preclude identification even given absolute security. Therefore, the limit is "reasonability," which means that an individual shall not be regarded as "identifiable" if the identification entails unreasonable delay, cost and manpower. One problem which merits further consideration is the general approach; the recommendations are intended for both private and official statistics.

Taking into consideration that official statistics have a legal basis, their rules are thus legally fixed and safeguarded and their aim is to endow the general public with the necessary information. This is not the case with private statistics, the legal basis of which is an agreement between the producer and the respondent. Here, the individual agreement defines the possible use, not public law. Thus, it is important that the recommendations strike a balance between being too general or too specific. If they are too general, they risk having no or little practical value. On the other hand, if they are too specific, they cannot be applied to both private and public statistics.

Proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data: The proposal, being at Council and Parliament discussion, seeks to ensure the free movement of personal data between Member States by providing for a harmonisation of national legislation. Thus, Member States shall protect the rights and freedoms of natural persons with respect to the processing of personal data and, in particular, their right of privacy.

The Directive follows a general data protection approach for personal data as such. According to this, statistics are subject to

the same rules and principles as domain data.

The specifications of statistics – data about mass phenomenon with their intention to lose their individuality in statistical results – are discussed only in Article 2(a), where data, presented in statistical form, in which the persons concerned can no longer be identified, are – by definition – not considered as personal data, and in Article 14 paragraph 3, where Member States may limit the right of access of the respondents.

It will be a matter of interpretation, in practice, if these exceptions are sufficient to balance the necessity of public information and the personal rights of individuals in a reasonable way.

2. Statistical Systems and Confidentiality

2.1. *Centralised versus decentralised statistical systems*

The way the different countries are facing these problems depends on the administrative structures, on the way the National Statistical System is organised, on the way individual records are processed and finally on the established practices and the cultural environment in which statisticians work. To compare the European Community statistical system with the U.S. system, as far as confidentiality is concerned, is not an easy task, especially given limited space. The two systems have fundamental differences which do not favour an easy comparison.

The U.S. statistical system is a federal system, organised upon “horizontal or subject matter administrative agencies”, each one having its own statistical department which collects and processes statistical data relevant to the objectives of the agency. Geographical responsibility covers the whole country.

The EC system is centralised, as far as

statistical domains are concerned, but decentralised geographically for the collection of “individual data.” Each national statistical system is responsible for the primary data collected on its national territory. All data collected by Eurostat, the Statistical Office of the European Community, are transmitted by the national administrations of the Member States. Until 1990, Eurostat was considered a “third party” and no confidential information was transmitted by the Member States. Since 1990, regulation 1588/90 adopted by the Council has set the foundations for the transmission of confidential data for statistical purposes and thus opened a new era for statistical confidentiality, not only because it is a step towards a European statistical system, but also because the organisation of confidentiality at the EC level poses new theoretical and technical problems and challenges.

It seems that in a decentralised system, co-ordination of legislation and practices is needed in order to avoid “inconsistent at best and chaotic at worst” situations. At the same time, data sharing is also a problem because of restricted access and thus such systems tend to duplicate costs and increase respondent burden.

2.2. *U.S. and EC legislation*

The fundamental legislation in the U.S. is the “Privacy Act of 1974” whose requirements apply to records maintained by most federal agencies. Its objective is to guard against abuse of administrative records. Among the various aspects which are underlined in Cecil’s paper, those related to the Privacy Act and the U.S. Bureau of Census are more important for comparisons, as neither educational records nor health records are of interest at the EC level. I would like to focus on

those that lend themselves more readily to comparisons between the European and the U.S. legislation. Those aspects are:

- i. The Privacy Act has a general coverage of administrative records and does not distinguish between administrative and statistical uses of the information. The argument for such a distinction is the direct effect on individuals that the administrative uses may have and the consequences that eventual abuses may have which contrasts with the benign of the statistical uses.
- ii. The Privacy Act gives the right to individuals to access their identifiable information.
- iii. The need of a prior written consent of the individual for disclosure and the way it is possible to circumvent this requirement.
- iv. The existence of twelve categories of exceptions offering a large spectrum of possibilities for disclosure of identifiable information and especially the concept of "routine use."
- v. The special position of the Bureau of the Census and the protection of data collected under Title 13 recognised as statistical data.

The situation at the EC level is very similar to the position of the Bureau of the Census. Indeed, regulation 1588/90 poses three basic conditions to Eurostat: (1) use confidential information only for statistical purposes; (2) publish data only in a way that prevents the (direct or indirect) identification of individuals; and, (3) access to persons outside Eurostat is permitted only under special authorisation and written promise of secrecy.

The points (ii) and (iii) are not relevant at Eurostat's level because there is no primary data collection, but are very important at

the national level, as it severely affects the work within NSI and the burdens for the administrations they may generate. Legislators should seek a balance and an optimal way to address this issue.

Regarding point (i) it should be stressed that in all Member States' legislation and in the "EC regulation 1588/90", the statistical nature of the data and the distinction made between Statistical Services and the bodies of the administrations is explicitly recognised.

This is a very important advantage of a specialised statistical system, like the system formed by Eurostat and the National Statistical Institutes of the Member States, as opposed to a decentralised system as in the U.S. where the double mission of collecting statistical data and policy making or surveillance creates an uncomfortable confusion when dealing with statistical confidentiality. Seen from the outside, there is no distinction between the administrative and the statistical utilisation of the individual information. This certainly affects the way legislators have applied the fundamental principle for privacy protection.

2.3. The notion of confidential individual data

In accordance with Article 2.1 of the Regulation 1588/90, confidential statistical data are defined by national legislation. The same applies to information on the private lives of natural persons (Article 2.3).

It is obviously very important to be clear about what "property" is being protected so that persons violating such protection can subsequently be prosecuted. Legislation in most Member States does not define precisely individual data or confidential data. Nevertheless, it is commonly understood

that these concepts are linked to the notion of the direct or indirect identification of the natural or legal person who has supplied the information.

It can be seen that all items of individual data are considered confidential and that, in principle, the type of data concerned are those that relate to the private lives of natural persons; it can also be seen that so-called "auxiliary" data relating to legal persons, i.e., data which allow these to be identified, are deemed to be confidential.

A few typical examples of national approaches are:

- France: all individual data relating to personal and family life and more generally to facts and behaviour of a private nature are considered confidential; furthermore, data of an economic or financial nature are protected against fiscal control or economic repression.

- Federal Republic of Germany: individual data are confidential and are defined as data supplied by a person for federal statistics and which relate to personal or material circumstances.

- Belgium: the law does not specify what confidential data are. In practice, all data that can be legally obtained only via the party concerned are deemed to be confidential.

- Italy: data are deemed to be confidential if they are personal in nature and relate to race or ethnic origin, political or ideological opinions, religious beliefs, as are data on an individual's state of health, sex-life, criminal convictions, and data classified as such under national laws or in international agreements ratified by Italy.

- Netherlands: all data are considered confidential if a person, enterprise or institution is identifiable, irrespective of the law on which the collection of data is based.

- Spain: all individual data obtained by the statistical services which relate to natural or

legal persons and allow the party concerned to be identified directly or indirectly are confidential.

Data are also deemed confidential if they are likely to reveal ethnic origin, political opinions, religious or ideological beliefs and, more generally, circumstances relating to an individual's private and family life, and if they may not be collected without the consent of the parties concerned.

- Greece: the law does not define confidential data; in practice, the National Statistical Service of Greece (NSSG) considers all individual data confidential.

- Luxembourg: no data supplied by individuals may be used for any other than statistical purposes; no individual item of data may be communicated to a third party, and under no circumstances to the tax authorities. Personal data, i.e., data relating to an identified or identifiable person, are also deemed confidential.

- Portugal: all individual statistical data collected by the Instituto Nacional de Estatística (INE) are deemed confidential. The INE may not request data relating to political, religious or other beliefs, as well as data of a highly personal nature.

- Denmark: all data are considered confidential if they can be attributed to an identifiable person, even if attribution presupposes knowledge of an individual's personal identification number, or use of a similar method. In general, the same applies to all data relating to individual enterprises.

- United Kingdom: we were unable to locate any statutory provisions on the notion of confidential data; however, the Statistics of Trade Act prohibits the dissemination of data relating to an individual "enterprise" without its written consent.

– Ireland: all individual information collected during a statistical survey is deemed confidential, even if it is public knowledge or recorded elsewhere.

2.4. *Protection of and access to confidential data*

In protection matters, legislations confuse the distinction between the protection of individual rights and protection of the data. In many countries the absolute approach has been chosen for both notions: “any data referenced to one individual is by definition confidential.” The situation in the EEC countries is not unique and the diversification of the NSS raises different approaches and ethics in relation to confidentiality matters. However, because they still all have a unique central statistical service even if, not in rare cases, delegation and decentralisation has transferred the activity of data collection to different departments or regional offices.

The protection of individual statistical data relating to natural persons should, in the view of users, differ from that of business data. Although it may well be possible to apply stricter rules to protect individuals, the same cannot be said of some types of business data where these are of great importance, sometimes at the national level.

All legislation stipulates that data collected by statistical services may not be used for any other than statistical purposes. In spite of the fact that all legislation prohibits the transmission of confidential data, there are numerous exceptions to this rule; these link the transmission of such data to protection of the statistical confidentiality arising from there.

All Member States’ legislation allows the transmission or publication of individual

data if written authorisation from the party concerned is obtained.

The following are the general rules governing the transmission of individual data areas for each Member State:

– Denmark: in principle, the transmission of individual data is prohibited, unless exceptions exist.

– Italy: individual data may be communicated only if they have been aggregated to prevent identification, or if they have been made anonymous. Decree-Law No. 322 does not make explicit provision for the transmission of individual data between the statistical services that form part of the national statistical system (Article 2); however, under certain circumstances, it may be possible to do this in order to ensure the smooth functioning of the government (Article 97 of the Constitution).

– Netherlands: several laws impose the obligation to communicate data and to treat as confidential data that which are collected for statistical purposes. The Netherlands Central Bureau of Statistics (CBS) may disclose data only on the condition that these are used for statistical purposes. However, the CBS may allow individual statistical data to be transmitted subject to the following conditions: (a) that the data have been collected on a voluntary basis, and (b) that a contract has been signed.

This contract is based on the following conditions: (i) that the individual data in question are used for statistical purposes only; (ii) that access to these data is restricted to research institutes; (iii) that they are published in aggregate form only, and (iv) that a fee is paid.

– Belgium: individual data collected during statistical investigations may be used solely by the Institut National de Statistique (INS) for statistical purposes only.

– Spain: the statistical services are obliged not to disclose individual data, whatever

their origin. The transmission of confidential data between government statistical services is authorised, subject to conditions guaranteeing the non-violation of statistical confidentiality.

– F.R. of Germany: the transmission of individual data is prohibited; data may be transmitted upon application to the Statistisches Bundesamt (SBA) by the Länder statistical offices where this is necessary for the methodological and technical preparation of federal statistics. German legislation contains a wide range of exceptions.

– France: individual data may be communicated only after a joint decision has been taken by the Minister in charge of the service conducting the survey and the Minister in charge of the Institut National de Statistique et des Etudes Economiques (INSEE) after the Committee on Statistical Confidentiality for enterprises has given its opinion.

– Greece: confidential data may under no circumstances be communicated or used for any other than statistical purposes.

– Luxembourg: individual data may not be communicated to third parties, and under no circumstances to the tax authorities. Individual data may under no circumstances be disclosed.

– Portugal: individual data concerning legal entities may not, in principle, be disclosed. Data on cooperatives, public and private enterprises, credit institutions and other economic operators may not be disclosed unless written authorisation has been obtained from the parties concerned, or unless authorisation is obtained from the Supervisory Council for Statistics, this being granted if the information is required for planning or economic co-ordination requirements, or if external economic relations are at stake.

– United Kingdom: the Statistics of Trade Act prohibits the dissemination of data relating to an individual enterprise without

its written consent; the possibility of transmitting individual data between statistical and administrative services varies according to the law under which the data have been collected.

– Ireland: items of individual data or verbal responses relating to an individual, an enterprise or a group and collected under the law may not be disclosed without the consent of the respondent. Individual data may be transmitted to a state authority if the Prime Minister is satisfied that such transmission is necessary for the proper discharge of its functions. In practice, confidentiality is guaranteed by the procedure followed by the Central Statistical Office (CSO) and by its code of conduct.

2.5. Other provisions

In general, the methods used to protect data concerning legal persons are the same in all Member States, i.e., active and passive confidentiality, although there are some variations. All Member States recognise the right of the respondent to be informed of his/her rights and obligations arising from the compilation of statistics. Several Member States grant respondents the right to correct data held on them.

Most Member States do not publish or disseminate statistical data relating to fewer than three economic entities (minimum number rule), and apply other rules such as that of the dominant enterprise or the anonymity rule. Other methods for protecting active or passive confidentiality exist, such as primary and secondary disclosure (France, United Kingdom), or “disclosure by difference” (United Kingdom), whereby individual data as defined under primary and secondary disclosure are reconstituted by comparing items contained in different tables.

3. Current Issues

3.1. *The risks of the “global approach”*

The many faceted developments in human rights and privacy protection and the subsequent legislations had, and continue to have, serious consequences for the organisation and the activities of National Statistical Institutes. Data collection rights and data treatment and dissemination are increasingly subject to restrictions by authorities that have been set up in many countries. This makes the circulation of information more difficult and subject to rules which are difficult to comply with.

Usually there is little notice of the distinction between the purpose of collection of the official statistics and other administrative purposes and even the commercial purposes of private firms. This “global approach” to privacy rights and individual data protection can have serious consequences for social objectives of a national statistical system.

In this context it becomes important for the NSS to identify the pertinent issues in order to preserve the basic objectives and the social role of official statistics. This also has repercussions for the users of official statistics, mostly administrations but also researchers and academics as access to valuable information becomes increasingly difficult.

Both papers point out several crucial issues for statistical agencies.

3.2. *What kind of information can be requested?*

How is the need for information of a developed society reconciled with the right for privacy of an individual? What are the limitations on the questions statistical agencies (as public authorities) may ask to a survey subject?

Traditionally, common sense has been the rule. No systematic legal approach has been defined and the development of legislation on the rights of a person to privacy devoted little attention to these questions. Here again, the blind global approach is applied and general provisions are established. A typical example is the Luxembourg law putting a ban on the collection and recording in data bases of information relating to an individual's political, trade union, philosophical or religious opinions or activities, together with data on the private lives of individuals (Law of 31 March 1979).

As legislation on individual rights progresses more and more, the danger for statistics and science is that very important and valuable information, e.g., health, nutrition, behaviour, etc., may be lost or at least lose its reliability.

Padiou (1992) develops the notion of “information sociétale”, as being the information an individual has to produce for society needs and proposes the distinction of two types of “individual” data depending on the purpose of the collection (1)-personal data when the aim of the collection is the person and (2) elementary data when the aim of the collection is a social need and the use has little to do with the person.

The famous Karlsruhe Court, while establishing the right of auto-determination of an individual on his/her personal data, underlined that this right is not absolute and that one has to consider that an individual belongs to a social community and in this sense “individual information” is not the property of the sole individual.

3.3. *Disclosure of directly identifiable individual information*

This is probably the most central contro-

versy. The rights for privacy of individuals are opposed to the right for information of the society. This is a difficult and lengthy debate which for the time being has been strictly prohibited to official statisticians, except, as we have seen, in some specific exceptions, covered by law. This is not the case for the private sector, where lists of persons or firms with names and addresses and all kinds of information are sold as commercial products. Some NSIs, like INSEE in France, are selling files of firms with addresses and some information and this seems to be a interesting income, harmless to the confidence that the firms have in INSEE's secrecy rigour.

3.4. *Linking different sources of information*

Mainly, the linkage of administrative records containing information on the same individual, coming from different sources and collected for different purposes is also a very important issue. It juxtaposes the right to be informed and determine the uses of one's own information with the optimisation and the efficiency of public administrations. How can we conciliate the concern for respondent burden with the non-disclosure principles? Although many countries have solved this efficiently and without subsequent problems, it remains that, in many countries, this is still forbidden because of strong theoretical opinions and fears.

3.5. *Conservation of individual records and re-utilisation*

This is the controversy between protecting the individual information from potential misuses in a political and social context different from the one the data were collected

in and the need for historic investigations for socio-economic analysis of very long periods. Most of the NSIs keep the records (raw data) for many decades (up to 100 years in UK). It is not excluded that intermediate solutions may be found in order to conciliate both needs.

3.6. *Access to individual records by the information subjects*

This is becoming more and more an issue for official statistics because of the global approach procedure.

An important illustration of this is the decision made by the constitutional Court of Karlsruhe (Germany) in 1983, where opposition was made to the German Population Census. Briefly, this decision gives priority to the person's rights versus the society needs and recognises the right of a person to auto-determination, that is, the right to determine the uses that can be made of his/her own data, which implies the right to access one's own data and to know who is using them, and how they are used.

This is certainly not very welcomed by statisticians. Not because they are opposed to the principle of auto-determination but because the use of this right may become cumbersome for the statistical services. Numerous records accumulate on a given person and any changes of the content of a record may have great consequences on all the derived and disseminated information.

4. References

- Padieu, R. (1992). *Statistique et Intimité*. International Seminar on Statistical Confidentiality, Dublin. Publication Eurostat (to appear).