

europa



**Protection of privacy,
automatic data processing and
progress in statistical documentation**

**Eurostat News
Special edition**

1986

**Theme
Miscellaneous**

9

**Series
Accounts, surveys and statistics**

C



OFICINA ESTADÍSTICA DE LAS COMUNIDADES EUROPEAS
DE EUROPÆISKE FÆLLESSKABERS STATISTISKE KONTOR
STATISTISCHES AMT DER EUROPÄISCHEN GEMEINSCHAFTEN
ΣΤΑΤΙΣΤΙΚΗ ΥΠΗΡΕΣΙΑ ΤΩΝ ΕΥΡΩΠΑΙΚΩΝ ΚΟΙΝΟΤΗΤΩΝ
STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES
OFFICE STATISTIQUE DES COMMUNAUTÉS EUROPÉENNES
ISTITUTO STATISTICO DELLE COMUNITÀ EUROPEE
BUREAU VOOR DE STATISTIEK DER EUROPESE GEMEENSCHAPPEN
SERVIÇO DE ESTATÍSTICA DAS COMUNIDADES EUROPEIAS

L-2920 Luxembourg — Tél. 43011 — Téléx: Comeur Lu 3423
B-1049 Bruxelles, Bâtiment Berlaymont, Rue de la Loi 200 (Bureau de liaison) — Tél. 235 11 11

Las publicaciones de Eurostat están clasificadas por temas y por series. La clasificación se encuentra al final de la obra. Para mayor información sobre las publicaciones, rogamos se pongan en contacto con Eurostat.

Para los pedidos, diríjense a las oficinas de venta cuyas direcciones figuran en la pág. 3 de la cubierta.

Eurostat publikationer er klassificeret efter emne og serie. En oversigt herover findes bag i hæftet. Yderligere oplysninger om publikationerne kan fås ved henvendelse til Eurostat.

Bestilling kan afgives til de salgssteder, der er anført på omslagets side 3.

Die Veröffentlichungen von Eurostat sind nach Themen und Reihen gegliedert. Die Gliederung ist hinten in jedem Band aufgeführt. Genauere Auskunft über die Veröffentlichungen erteilt Eurostat.

Ihre Bestellungen richten Sie bitte an die Verkaufsbüros, deren Anschriften jeweils auf der dritten Umschlagseite der Veröffentlichungen zu finden sind.

Oi εκδόσεις της Eurostat ταξινομούνται κατά θέμα και κατά σειρά. Η ακριβής ταξινόμηση αναφέρεται στο τέλος κάθε έκδοσης. Για λεπτομερέστερες πληροφορίες σχετικά με τις εκδόσεις, μπορείτε να απευθύνεσθε στην Eurostat.

Για τις παραγγελίες μπορείτε να απευθύνεσθε στα γραφεία πώλησης, των οποίων οι διευθύνσεις αναγράφονται στη σελίδα 3 του εξωφύλλου.

Eurostat's publications are classified according to subject matter and series. This classification is indicated at the end of the publication. For more detailed information on publications, contact Eurostat.

Orders are obtainable from the sales offices mentioned on the inside back cover.

Les publications de l'Eurostat sont classées par thème et par série. Cette classification est précisée à la fin de l'ouvrage. Pour des informations plus détaillées sur les publications, contactez l'Eurostat.

Pour les commandes, adressez-vous aux bureaux de vente dont les adresses sont indiquées à la page 3 de la couverture.

Le pubblicazioni dell'Eurostat sono classificate per tema e per serie. Tale classificazione è precisata alla fine dell'opera. Per informazioni più dettagliate sulle pubblicazioni rivolgersi all'Eurostat.

Per eventuali ordinazioni rivolgersi a uno degli uffici di vendita i cui indirizzi figurano nella 3ª pagina della copertina.

De publikaties van Eurostat zijn ingedeeld naar onderwerp en reeks. Deze indeling is achter in het boek opgenomen. Voor nadere informatie over de publikaties kunt u zich wenden tot Eurostat.

Gelieve bestellingen op te geven bij de verkoopbureaus, waarvan de adressen op bladzijde 3 van de omslag zijn vermeld.

As publicações do Eurostat estão classificadas por tema e por série. Esta classificação está indicada no fim da obra. Para mais informações sobre as publicações é favor contactar o Eurostat.

Encomendas: serviços de venda cujos endereços estão indicados na contracapa.

Para obtener esta publicación, se ruega dirigirse a las oficinas de venta cuyas direcciones figuran en la tercera página de cubierta.

Denne publikation kan fås gennem de salgssteder, som er nævnt på omslagets tredje side.

Diese Veröffentlichung ist bei den auf der dritten Umschlagseite aufgeführten Vertriebsbüros erhältlich.

Την έκδοση αυτή μπορείτε να την προμηθευτείτε από τα γραφεία πωλήσεων, τα οποία αναφέρονται στην τρίτη σελίδα του εξωφύλλου.

This publication is obtainable from the sales offices mentioned on the inside back cover.

Pour obtenir cette publication, prière de s'adresser aux bureaux de vente dont les adresses sont indiquées à la page 3 de la couverture.

Per ottenere questa pubblicazione, si prega di rivolgersi agli uffici di vendita i cui indirizzi sono indicati nella 3ª pagina della copertina.

Deze publikatie is verkrijgbaar bij de verkoopkantoren waarvan de adressen op blz. 3 van het omslag vermeld zijn.

Para obter esta publicação, deve dirigir-se aos serviços de venda cuja direcção se indica no verso da contracapa.

**Protection of privacy,
automatic data processing and
progress in statistical documentation**

**Eurostat News
Special edition**

1986

This volume is also available in:
FR: N° de catalogue: CA-AB-85-006-FR-C

Cataloguing data can be found at the end of this publication

The views expressed in this publication are the personal views of their authors. They do not express opinions, or policies, either of the Commission, or of national governments.

As long as stocks last, copies in German and Italian of the papers presented at the seminar may be obtained by applying to Mrs M. Ottens, Eurostat, Bâtiment Jean Monnet, L-2920 Luxembourg (tel. 4301-2021).

Luxembourg: Office for Official Publications of the European Communities, 1986
Catalogue number: CA-AB-85-006-EN-C

© ECSC — EEC — EAEC, Brussels · Luxembourg, 1986

Printed in the FR of Germany

Table of Contents

	page
PREFACE	5
INTRODUCTION G. Als	7
PRIVACY, CONFIDENTIALITY AND DATA PROTECTION; A EUROSTAT VIEW C.W.D. Pearce	13
DATA CONFIDENTIALITY: CURRENT PRACTICE AND RESTRAINTS IN FRANCE J. Begué, C. Moreau	33
STATISTICS AND INDIVIDUAL LIBERTIES N. Lenoir	51
DATA PROTECTION IN A NATIONAL REGISTER-BASED STATISTICAL SYSTEM L. Thygesen	63
STATISTICS AND PRIVACY - THE OFFICIAL PRODUCTION OF STATISTICS AT A CROSSROADS E. Rapaport	79
CONFIDENTIALITY AND PRIVACY; CURRENT SITUATION AND PRACTICE IN IRELAND F.A. Embleton	93
PROTECTION OF PRIVACY, AUTOMATIC DATA PROCESSING AND PROGRESS IN STATISTICAL DOCUMENTATION IN THE FEDERAL REPUBLIC OF GERMANY J. Werner, E. Südfeld	99
STATISTICS AND ADMINISTRATION W. Begeer, W.F.M. de Vries, H.D. Dukker	133
THE N.I.S. OF BELGIUM AND PROTECTION OF PERSONAL PRIVACY E. van Langendonck	147
DATA PROTECTION ISSUES AS THEY AFFECT POPULATION CENSUSES AND SOCIAL SURVEYS IN THE UNITED KINGDOM R. Barnes	163
THE DEVELOPMENT OF DATA PROTECTION IN THE UK: ITS IMPLICATIONS FOR OFFICIAL STATISTICS T.J. Orchard	177

DO STATISTICIANS NEED A CODE OF CONDUCT? R. Jowell	197
DATA PROTECTION AND THE STATISTICAL COMMUNITY D.H. Flaherty	239
PROTECTION OF PRIVACY, AUTOMATIC DATA PROCESSING AND PROGRESS IN STATISTICAL DOCUMENTATION G. Cariani, G. Giacommo, P. La Cava	259
DRAFT LEGISLATION IN ITALY ON THE PROTECTION OF PERSONAL DATA M. Losano	275
THE CONFIDENTIALITY OF STATISTICAL DATA IN GREECE E. Stiliaris	305
RECORD LINKAGE vs. CONFIDENTIALITY FROM THE PERSPECTIVE OF THE U.S. BUREAU OF THE CENSUS D.D. Nelson	325
THE SITUATION IN SWITZERLAND H. Brüngger	337
CLOSING REMARKS G. Als	339
List of participants	345

PREFACE

The papers reproduced in this volume were first presented at an international seminar held under the auspices of the Statistical Office of the European Communities (EUROSTAT) at Luxembourg in December 1984. In preparing the papers for publication some latitude was given to authors to revise the original texts. In principle, however, there has been no substantial departure from the texts which were presented. Although there has not been space to reproduce in full the discussions on each paper, the value of such discussion to the seminar as a whole should not go without recognition.

In opening the seminar, Richard Burke, Vice President of the European Communities with responsibility for the Communities Statistical Office, said that during the past decade the Statistical Office of the European Communities has organized a series of seminars on topics of interest to the international statistical community. The 1984 seminar continued this series but, at the same time, differed from its predecessors in a number of significant respects. The most important difference was that it did not concentrate only on technical aspects of the statistician's craft but also dealt with matters of statistical ethics which had implications and interests going far beyond the confines of the statistical profession. The purpose of the seminar was not to attempt to reach definitive conclusions. EUROSTAT's aim was rather to provide a forum for an exchange of views and experience on the range of important problems currently faced by statisticians arising from considerations of privacy and confidentiality and to encourage an informed discussion on possible strategies for maintaining the flow of statistical information in the face of these problems.

The importance of the topic of this seminar was amply illustrated by the extensive public attention which it had received in recent years. This was reflected in the membership of the seminar which included participants not only from Member States but also from Canada, Norway, Sweden, Switzerland and the United States. Many countries across the world had either adopted or were actively considering legislation to control the collection and dissemination of information. Such legislation typically was applicable to a far wider range of information than that handled by statisticians but it frequently made special provisions for statistics which went some way to ease the burden on the official statistician. At the same time, however, it imposed responsibilities upon him which he could

not and should not evade. Of course, the problem did not affect only the statistician. It was also a matter of serious public concern which was expressed most commonly as the conflict between the decision maker's need to know and the individual's right to privacy. This conflict between the need for information and the genuine fears of the consequences of misuse of that information was critical. If the seminar could assist in the resolution of that conflict it would have made an important contribution to the solution of a major problem

EUROSTAT was particularly fortunate in securing Mr. George Als, the Director General of the "Service Central de la Statistique et des Etudes Economiques" (STATEC) of Luxembourg, as Chairman for the seminar. Mr. Als had been actively interested in the subject matter of the seminar for many years.

INTRODUCTION

G. ALS

Service Central de la Statistique et des Etudes Economiques,
Luxembourg

Government departments have long been regarded as unproductive. This somewhat unflattering reputation was elevated to the rank of an economic theory in the 1940s and '50s by Colin Clark and Jean Fourastié, who maintained that while in the secondary or industrial sector productivity showed rapid growth, in the tertiary sector - or the services as it is still known - it progressed extremely slowly.

Yet now we find that a branch of that same civil service which has been treated so condescendingly in the past is suddenly achieving such astonishing increases in productivity that entire populations are as frightened of it as they are of atomic energy.

What, then, has happened? Two things:

- firstly, the arrival of the computer, an instrument as mysterious as the nuclear reactor;
- secondly, the introduction of the national registration number for individuals, a kind of fissile material, capable of producing within the computer a chain reaction leading to an information explosion. A highly controlled reaction, say the statisticians; a threat, say the data protection experts.

This raises three questions:

- How is privacy threatened?
- How far does this trend affect statistics?
- What was this seminar expected to achieve?

A. The problem of the protection of privacy

An OECD report published in 1971 on "Computerized information and the protection of privacy" observed:

"It is as though the information contained in printed works or written manuscripts were in solid form and the information stored in computers in gaseous form. Information has acquired the necessary mobility, penetrating power and capacity to transform itself into new shapes and dimensions".

A number of potential dangers emerge from this, such as the ease with which information which is not strictly essential can be recorded, the ease with which unscheduled processing of information can take place, the ease with which information can be transmitted to non-authorized persons, the alteration or absolute erasure of data, and last but not least the linking of files.

In the report quoted above, the OECD concluded that the problems posed by the computer were of an entirely new order and that neither the technical safeguards nor the existing legislation, in the form of the various administrative confidentiality requirements and the right to legal remedy, offered in themselves a satisfactory solution to the problem of data confidentiality.

As early as 1970 the State of Hesse in the Federal Republic of Germany passed a law on data protection and in so doing became the first in the field, ahead even of the Swedish Law of 1974. Under the Hesse law a post of Commissioner responsible for data protection was created. During the 1970s many countries were to follow this example but the complexity of the problems involved was such that the need for discussion at international level began to be felt. Two organizations in particular were involved in this: the OECD and the Council of Europe, and two international instruments resulted from their work:

- the recommendations of the OECD Council meeting of 23 September 1980 concerning the guidelines governing the protection of privacy and transborder flows of personal data;
- the Council of Europe's Convention of 28 January 1981 concerning the protection of individuals against automatic processing of personal data.

This Convention contains a wealth of ideas common to all Community countries. Leaving aside the regulations concerning transborder flows, and given that the aim of the Convention was to guarantee every individual's right to privacy with regard to the automatic processing of personal data, the field can be reduced to five ideas.

1. As far as the status of data is concerned, the Convention established the principles of purpose, relevance and time-limit: personal data should be recorded for definite and legitimate purposes, the extent of such data should not be excessive in relation to these purposes, nor should such data be stored for a period exceeding that which is necessary for these purposes.

2. Certain categories of data - i.e. personal data which reveal an individual's political, religious or philosophical beliefs and data relating to a person's private and family life - may not be subjected to automatic processing unless, as the Convention says, the particular national legislation involved provides for appropriate confidentiality safeguards.
3. Appropriate safeguards must be taken for the protection of personal data.
4. As an additional safeguard for the individual concerned, the Convention lays down a right of access to files; all persons should have the right to know if files exist in a particular area, to confirm whether a particular file contains data which concern them and to ensure that these data are corrected or erased in all cases where they have been processed in violation of national legislation.
5. The use of this right of access can be restricted in the case of personal data used for statistical purposes or scientific research, provided it is clear that the privacy of the individuals concerned is not jeopardized. It is important to remember that these principles apply only when two conditions are met:
 - the data concerned must be personal data: depersonalized data are free from all restrictions;
 - the data must be processed automatically; the statistician retains his freedom of action with regard to data processed manually or with punch-card machines.

B. The protection of privacy, automatic data processing and progress in statistical documentation

The OECD guidelines and the Council of Europe Convention were useful in that they set limits beyond which there are breaches of privacy. They were thus taken into account in much of the developing national legislation. They did not, however, go very far in resolving the official statistician's dilemma, namely what other considerations should be taken into account in striking a balance between the individual's right to privacy and the community's information requirements. More and more statisticians felt that their legitimate requirements were being overlooked. Data protection laws were introduced in the 1970s, in some countries without statisticians being consulted, and this resulted in problems. The suggestion made in 1984 by the Conference of the Directors General of the National Statistical Institutes of the European Economic Community for a seminar to discuss the problems of the "protection of privacy, automatic data processing and progress in

"statistical documentation" was thus a timely one. That the Conference did not shy away from such an unattractive title was surely to draw attention to the need not only to consider problems of the protection of privacy but also to consider these problems in relation to progress in statistical documentation.

The fact was that statisticians felt themselves to be in something of a dilemma. Since the onset of the economic crisis, they have been obliged by budget austerity measures to make cuts and to rationalize, a necessity which is reinforced by the complaints of private individuals and firms who feel that the administrative burdens imposed on them are too great and who therefore urge statisticians to make greater use of administrative sources. If, however, statisticians attempt to do precisely that, they come up not just against the age-old obstacles of the various administrative confidentiality requirements such as confidentiality of tax data, but also against new obstacles arising out of legislation to protect privacy.

When automatic data processing was still in its infancy, the statisticians' dream was that the linking of administrative and statistical files by using the registration number for natural persons would ultimately lead to all statistics being based on administrative sources. Such hopes have more or less disappeared because such linking of files has been denounced as a threat to individual privacy.

In his opening address at the 43rd session of the International Statistical Institute in December 1981, the President of the Institute, Mr. Edmond Malinvaud, described the situation as follows:

"In recent years progress in the use of administrative data has often come up against legislation aimed at protecting the public against possible abuses in the use of these files. Naturally, statisticians welcome the existence of legal safeguards to protect individual privacy; in particular, such safeguards should help to create a climate likely to bring about a more positive public reaction to statistical investigations. Nevertheless, these safeguards must be modelled in such a way that they do not prohibit the performance of certain statistical operations which pose no threat to individual privacy. However, cases have arisen where sometimes the new laws themselves and sometimes the way in which they are implemented have restricted unnecessarily the scope of statistical operations using administrative data".

C. What was this seminar expected to achieve?

Statisticians, data protection specialists and researchers therefore urgently needed the opportunity to discuss together some of the practical problems created by the new legislation. The following are a few examples:

- Under what circumstances is interlinking of statistical files or linking of statistical files with administrative files permitted? This is a crucial question which raises virtually no difficulties in Denmark but for which no solution has been found in other countries.
- Can, as the Danish and Belgian reports suggest, the communication of administrative data to statisticians be considered justifiable provided it is certain that the flow is only one-way? Could the statistical office be given a general right of access to administrative data by means of the transfer of protocols between the office and the departments which have the information?
- Should files of statistical surveys be depersonalized after the validation phase? The Swedish report feels that this is unwise because to do so is to neglect the needs of researchers and, more particularly, of historians and future generations.
- The right of access to personal data in computerized files is restricted to the persons concerned. Should this apply to administrative files only and not to statistical files? Denmark, the United Kingdom and Belgium, for example, think that it should, since statistics are not likely to threaten privacy.
- Can public and private institutes, in particular researchers, be given samples of names and addresses obtained from a population census or another file? Opinions differ; some countries do not regard it as a problem (e.g. the United Kingdom and Belgium), while others (e.g. France) do not feel it should be done.
- Can a population census be used by municipal authorities (which supply census-takers) to update certain of their files of names and addresses? This seems a fairly harmless question, but could be very sensitive, as the example of Germany shows.
- National legislation forbids the automatic processing of data of a personal nature revealing political opinions and religious and philosophical beliefs or relating to the individual's private life, but the Strasbourg Convention enables this restriction to be waived if domestic law provides for suitable guarantees. Has any country created these guarantees?

- The scope of the seminar clearly went beyond the old familiar question of statistical confidentiality. However, it is essential to remember the fundamental importance and specific nature of this confidentiality, very different from any other type, as indeed the Greek, Belgian and Dutch reports emphasize. The Greek report raises the question of statistical ethics: if the Minister to which the Statistical Office is answerable asks the Director of the Office of one of his staff for personal data obtained from surveys and censuses, can this request be refused? It is the moral question *par excellence* of the government statistician. What value statistical confidentiality if it does not apply in the face of political authority, which certainly does not require the data for scientific purposes?

As those attending the seminar came from many different fields, it was possible for the data protection specialists and external research organizations, as well as the statisticians, both civil servants and others, to voice their opinions.

There was general agreement about the need for a continuing discussion and for an exchange of experience about ways on which the problems were being resolved. There was also a general feeling that the statisticians should work towards peaceful cooperation with the data protection departments.

This indeed is what we tried to do, keeping in mind Professor Flaherty's advice to statisticians: statisticians, says Professor Flaherty, - who is nothing if not sure of his ground - should develop 'a finely tuned political sense' and should aim towards peaceful coexistence with the data protection lobby, helped by summit conferences to bring about a reduction in the number of missiles deployed by either side.

PRIVACY, CONFIDENTIALITY AND DATA PROTECTION;
A EUROSTAT VIEW (1)

C.W.D. PEARE
Statistical Office of the European Communities (2)

SUMMARY

EUROSTAT has always been very conscious of its responsibilities for preserving the confidentiality of data but, until relatively recently, it has been largely insulated from the effects of the growing public concern with issues relating to privacy. This insulation has arisen from the fact that, in general, EUROSTAT is not involved in the direct collection of data but relies on the national services acting as its agents.

This paper describes some of the problems faced by EUROSTAT which arise from concerns with privacy and confidentiality.

0. INTRODUCTION

EUROSTAT has not been directly involved in the political discussion about privacy in Member States. The impact of considerations of privacy on Community statistics is, however, a matter of serious concern. This impact manifests itself in increasing difficulties for the national statistical offices acting on our behalf in the collection and compilation of statistics. There have indeed been instances where concerns with privacy in some countries have frustrated analyses required for Community purposes. On the other hand, EUROSTAT has always been directly involved and concerned with the problems of confidentiality, but in recent years it has been more and more necessary to reassure data providers that adequate measures are taken to protect material entrusted to us.

Work at EUROSTAT is normally one step removed from the data subjects whether individuals, or collectives of individuals such as families, businesses, farms etc. Our role places us as intermediaries between data users (whose demands for timely, relevant and reliable information are continually increasing) and data providers who are experiencing growing problems in meeting our requirements. The problems of providing data affect us in two respects. There are firstly, the increasing difficulties of our data providers (usually

-
- (1) EUROSTAT: Statistical Office of the European Communities
 - (2) This paper represents a personal view and should not be taken to reflect the position of the views either of the Commission of the European Communities or of EUROSTAT

national administrations) in providing the data we request and, secondly, even where the information can be collected there are questions of confidentiality facing national administrations. These latter may be constrained (legally or otherwise) from releasing, to EUROSTAT, information which may disclose aspects relating to individuals. In the case where such information may be provided to EUROSTAT the responsibility for ensuring its subsequent confidentiality is placed on EUROSTAT.

As may be seen above our direct responsibilities insulate us to some extent from the immediate front line problems of confidentiality and privacy. However, the indirect aspects cannot be ignored. This paper concerns itself with some of the present and potential future effects of concerns with privacy and confidentiality as seen from the viewpoint of EUROSTAT.

In view of our relatively limited direct experience in this area extensive use has been made of the work of others in preparing this document. In particular one should note the descriptions of problems experienced at national level which have been described in papers of the Conference of European Statisticians and also the attitudes and concerns which are manifested in the declarations of the Council of Europe and the OECD in respect to privacy and confidentiality of statistics. A selective bibliography is given in an annex. More extensive collections of reference may be found in several of the publications listed. (For example, in "Censuses, Surveys and Privacy", a collection of articles on problems related to privacy edited by Martin Bulmer, London, Macmillan 1979).

1. DEFINITIONS OF PRIVACY AND CONFIDENTIALITY

Various definitions of privacy and confidentiality have been proposed in the literature. One approach to the problem is to be found in "A Framework for Planning U.S. Federal Statistics for the 1980s (U.S. Department of Commerce, July 1978)" where the importance of a distinction between privacy and confidentiality is clearly drawn:

"Privacy, on the one hand, has been variously defined as: (1) the right to be left alone, to be spared from unauthorized oversight and observations, and from searching inquiries about oneself and one's business; (2) the ability to control the use of information about oneself, whether to give it free circulation, limited

circulation, or no circulation at all, and (3) the right to participate in a meaningful way in decisions about what information will be collected and how that information will be used. The concept of autonomy is also used to describe this right. In one view, these definitions imply that a data subject must be fully informed about all uses of data sought and be given the right to withhold consent from any or all such uses. In the extreme, of course, the Government should not collect any information at all.

Confidentiality, on the other hand, involves the conditions of use and disclosure of data once it is collected. The Government's needs for information about individuals, businesses, and institutions fall into many different categories including counting the population, as mandated in the Constitution; providing benefits such as welfare, student loans, or medical insurance; collecting taxes; regulating industry; enforcing laws; evaluating programs; and advancing the state of knowledge through statistics and research. Hence, the Government collects or causes to be collected great amounts of data, some of it highly personal or capable of inflicting great competitive injury if made public. The challenge posed by the dual concerns for privacy and the enhancement of knowledge is therefore to refrain from collecting unnecessary information and to maintain the necessary degree of confidentiality for that which is collected."

A similar distinction is made by C. Hakim in a paper "Census Confidentiality in Britian" (in M. Bulmer ed.)

"It is important that the two concepts of privacy and confidentiality be distinguished, for we are concerned mainly with the latter. Privacy can be defined as the preservation of personal (or business) information from public knowledge or inspection. The right to privacy implies the ability to control the release of information about oneself, whether to give it free circulation, limited circulation or no circulation at all. ... Confidentiality involves the conditions of use and disclosure of information about individuals or business once it is collected. In terms of the census, privacy concerns the type of question that is to be included or excluded from the census form, while confidentiality concerns the limits on the release of data and the use that is made of the information once it is collected. The two are related in that public

concern over confidentiality is likely to be greater where the information collected is considered to be private."

The essence of these, and other definitions, seems to be a distinction between "privacy" as "the right to be left alone" and "confidentiality" as "imposing conditions and responsibilities with respect to the presentation and use of data after its collection". For the statistician the distinction is perceived in the form that privacy affects his ability to obtain information, while concerns with confidentiality impose constraints upon the use which may be made of the information after its collection. This distinction between the two concepts is adopted in the present paper.

2. THE IMPACT OF CONCERNS WITH PRIVACY AND CONFIDENTIALITY

Concerns with privacy and confidentiality have had a number of very direct and important impacts on the ability of EUROSTAT to meet the continually increasing demand for harmonized information relating to Community Member States. Some of these effects have also been of major significance to individual Member States such, for example, as the legal and constitutional difficulties which have been encountered in the Federal Republic of Germany with respect to the census of population and the micro-census. Concern with confidentiality of data has had an impact where Member States have found themselves in a position to collect information but unable, because of concerns with confidentiality, to make it available to EUROSTAT. This latter problem has arisen with farm structure surveys where several Member States, for legal or policy reasons, found themselves unable to provide the Commission with sample (or complete enumeration) data relating to individual holdings despite the willingness of the Commission to offer stringent guarantees in respect of confidentiality (see for example Council Directive 75/108/EEC of 20 January 1975 which lays down the procedures for the farm structure survey conducted in 1975. Article 10 section 2 reads "the individual data ... shall be communicated to the Statistical Office of the European Communities in such a form that the holdings concerned cannot be identified. Moreover, they may be divulged only to those persons responsible at the Office for the application of this Directive").

The effects of such restrictions on the availability of information to EUROSTAT are felt most directly by our customers. In the main, these are policy makers in the Commission, members of the European Parliament, other bodies and individuals with an interest in Community affairs. The

effects may take the form of non-availability of results (which is particularly frustrating where information has been available in the past) or restrictions on the quantity, quality or flexibility of information available.

3. PRIVACY

The fundamental problem underlining the difficulties which are recognized in the associated fields of privacy, confidentiality and data processing appears to be that of privacy. Concerns with confidentiality and the implications of data processing are largely expressed in terms of preserving the privacy of the individual or group to which information relates.

It is almost universally recognized that, other things being equal, individuals (or collectives of individuals such as families, groups of workers, businesses, farms, etc.) have a right to privacy in matters which concern themselves alone. However, other things are not always equal, and it is accepted that, for specified purposes, this privacy may legitimately be invaded. Legitimacy is justified in terms of the striking of a balance between the right to privacy and a (society's) need to know. The individual will be required to make available information on his income to taxation authorities and will recognize the advantage of making available information on symptoms and medical history in the course of a medical consultation. In the former example, the breach of privacy will be accepted (with a greater or lesser reluctance) on the grounds that this is necessary for the running of society and in the latter example by the realization that there is a direct benefit in terms of improved diagnoses and treatment resulting from the revelation of personal medical information. Willingness to provide information which is regarded as economically or psychologically sensitive is very closely related to the acceptance that the breach of personal privacy is relevant and limited. Relevance arises in the sense that there is some direct or indirect advantage to the information provider and limitation in the sense that the information will be used only for the purposes specified. A breach of confidentiality whereby the information is divulged to other parties will generally be regarded as very damaging, and even further, a linkage of different elements of information to build up a more detailed profile of the individual is regarded as particularly threatening. It may be said that the importance of the sum of several pieces of information relating to an individual may be much greater than the total importance of the individual components. If relevance and limitation of diffusion are two aspects which may encourage the information provider to accept a breach of privacy, where does this leave the statistician in his search for information? The relevance of a statistical enquiry to the

individual will almost always be less direct than the medical example quoted above and will on many occasions appear to be vanishingly small to the respondent. Where response imposes a burden on a respondent which is significantly greater than the perceived benefit of the resulting statistics he may see his personal interest as being best served by refusing to respond or by seeking the discontinuation of the enquiry. Where information is provided, the respondent is generally seen as having a right to an assurance that diffusion of information provided will be limited in a closely prescribed fashion.

Informing and/or persuading potential respondents of the relevance of statistical enquiries seems, with a few exceptions, to be a relatively neglected activity on the part of statisticians seeking to improve the quality of their data. It does seem that it will be more and more necessary for public relations expertise to become a part of the statisticians' armoury in the future. Limitation of diffusion of information needs to be well defined in that the respondent should be made aware of the uses to which the information will or may be put and assured that it will not be utilized for other specified or unspecified purposes. The potential benefits of such an approach have been well described in Margaret Martin "Statistical Legislation and Confidentiality Issues" (International Statistical Review 1974 quoted in U.S. Department of Commerce op.cit. p. 255).

"Even when responses to requests for information are required by law, the success of a statistical program depends in large measure on the willing cooperation of respondents. Respondents who understand the purpose of the inquiry, who sympathize with the intended use of the information, and who believe that providing the government with the requested information will not harm them are much more likely to answer truthfully and with a minimum of effort on the part of the data collection agency. One element in enlisting such cooperation is assurance of harmlessness to the respondent, and one of the most common methods for making such assurance in statistical data collection is the provision for keeping the replies confidential."

As has been indicated above, from the EUROSTAT viewpoint, privacy is seen as an indirect problem, but this should not be taken to imply that it is not regarded as a potential source of serious difficulties. A deep concern is felt that, notwithstanding genuine concerns with privacy, the availability of information required for Community purposes should not be frustrated by artificial or exaggerated appeals

to the need to respect rights to privacy. As with national needs, a balance must be established between the right to privacy and the Community's need to know. At the same time it should be recognized that the Community's needs and priorities are not always identical to those of individual Member States.

4. CONFIDENTIALITY

The statistician cannot, and should not attempt to evade absolute responsibility for the confidentiality of data which have been provided to him directly or indirectly by respondents who have been assured that their privacy will be respected. Ethically this is a clearcut professional responsibility but, even on purely utilitarian grounds, the preservation of the relationship between the statistician or potential and actual respondents depends upon the latter being able to place trust on the statisticians' undertakings. (It might be emphasized that the ethical responsibility of the official statistician in this context is possibly greater than that of his non-official counterpart because of the availability of power to compel response to official enquiries.).

It has been said that no useful therapeutic drug is ever completely harmless in all circumstances; the same may be said of statistics. Any non-trivial information relating to a respondent has potential for harm and the statistician has the duty to ensure that, through its disclosure, the respondent is not harmed directly as a result of providing the information. The statistician can never give an absolute guarantee that indirect adverse effects from the provision of data can be avoided. For example, analysis may reveal the existence of particular groups, the members of which may be affected by subsequent policy or administrative decisions.

The discussion of technical procedures for preventing breaches of confidentiality lies beyond the scope of this paper. Potentially such breaches can occur in many different ways but two broad classes of problem may be distinguished: a) those resulting from direct "attack" on the data files relating to individual respondents and b) those resulting from "disclosure" of information, for example, from highly refined detailed tabulations which may indicate that one or a small number of subjects lie within a particular multi-dimensional classification.

Within EUROSTAT the first of these groups of problems is generally approached by requesting information only in anonymous form and by limiting access to data files to those who are directly responsible for the processing of results. Further information relating to individual respondents is

not extracted from the (usually machine-readable) files except for purposes of investigating apparent anomalies encountered in processing. The key linking records numbers with actual identities and the identities of subjects normally remains with the Member States. The limitation of access to the files inhibits any approach, by unauthorized users, to identifying individual information by matching known characteristics with survey responses.

The control of disclosure from tabulations may take several forms, the simplest of these being inspection and rounding of results. The most important technical difficulty in dealing with this type of danger to confidentiality results from the fact that a breach may be fully guarded against in terms of information available to the statistician but the existence of auxiliary information in the hands of others may lead to disclosure. (Take for example an industry where a very large proportion of those involved belong to an association and provide the association with information similar to that collected and published by an official agency. Unless account were to be taken of this external factor non-members of the association could find their information subject to disclosure by comparison of analyses derived from the two data sets.) Many of the technical devices and procedures which exist guard greater knowledge of the area of study and alternative sources of informations than is always available to the statistician. One of the consequences of this is likely, in the future, to imply an increasing need for the statistician to be aware of alternative (not necessarily published or available to him) sources of information in his field of study. In the EUROSTAT context this implies a continuation of the already close liaison with our colleagues in the Member States.

5. THE COUNCIL OF EUROPE CONVENTION AND OECD GUIDELINES

In view of EUROSTAT's particular concern with the international flow of information the provisions of the Council of Europe Convention for the protection of individuals with regard to automatic processing of personal data and the recommendations of the Council of the OECD concerning Guidelines governing the protection of privacy and transborder flows of personal data are of particular interest. The Council of Europe Convention, which was opened for signature in January 1981, is directed primarily at ensuring the right to privacy of the individual as laid down in Article 1.

"The purpose of this convention is to secure in the territory of each Party for every individual, whatever his nationality or residence, respect for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic processing of personal data relating to him ("data protection")."

The protection of privacy is seen as being safeguarded by a number of measures the most important of which, from the viewpoint of EUROSTAT, are those relating to transborder data flows as covered by Article 12.

- "1. The following provisions shall apply to the transfer across national borders, by whatever medium, of personal data undergoing automatic processing or collected with a view to their being automatically processed.
2. A Party shall not, for the sole purpose of the protection of privacy, prohibit or subject to special authorization transborder flows of personal data going to the territory of another Party.
3. Nevertheless, each Party shall be entitled to derogate from the provisions of paragraph 2:
 - a. insofar as its legislation includes specific regulations for certain categories of personal data or of automated personal data files, because of the nature of those data or those files, except where the regulations of the other Party provide an equivalent protection;
 - b. when the transfer is made from its territory to the territory of a non-Contracting State through the intermediary of the territory of another Party, in order to avoid such transfers resulting in circumvention of the legislation of the Party referred to at the beginning of this paragraph."

The OECD Guidelines (1981) lay down a number of basic principles on the international flow of information and these are reproduced here:

- "15. Member countries should take into consideration the implications for other member countries of domestic processing and re-export of personal data.
16. Member countries should take all reasonable and appropriate steps to ensure that transborder flows of

personal data, including transit through a member country, are uninterrupted and secure.

17. A member country should refrain from restricting transborder flows of personal data between itself and another member country except where the latter does not yet substantially observe these Guidelines or where the re-export of such data would circumvent its domestic privacy legislation. A member country may also impose restrictions in respect of certain categories of personal data for which its domestic privacy legislation includes specific regulations in view of the nature of those data and for which the other member countries provide no equivalent protection.
18. Member countries should avoid developing laws, policies and practices in the name of the protection of privacy and individual liberties, which would create obstacles to transborder flows or personal data that would exceed requirements for such protection."

Both the Council of Europe Convention and the OECD guidelines are concerned with the transfer of data in general. The specific problems of statistical data are not treated in any detail. For example Article 5 of the Convention lays down conditions relating to the quality of data including a specification that "personal data undergoing automatic processing shall be preserved in a form which permits identification of the data subject for no longer than is required for the purpose for which these data are stored". Article 6 defines specific sensitive categories of data while Article 8, specifying safeguards for the data subject, lays down rights of access, rectification, erasure, etc. which may be availed of by any person covered by a data set. The only exceptions which are permitted to these general rules are laid down in Article 9 and these include the statement that restrictions on the exercise of rights specified in Article 8 may be provided only with respect to automated personal data files used for statistics where there is obviously no risk of an infringement of the privacy of the data subjects.

The OECD guidelines appear to be rather less restrictive; in particular, paragraph 3 lays down that the guidelines should not be interpreted as preventing the exclusion from their

application of data which "obviously do not contain any risk to privacy and individual liberties".

We do not yet have sufficient experience of the interpretation to be placed by Member States on these two documents to assess their implications for Community statistics. It does, however, seem possible that a restrictive interpretation of the terms of the Convention and the guidelines could have further adverse effects on the availability of statistical data to EUROSTAT.

6. SOME PARTICULAR CONCERNS

One area likely to be particularly affected by restrictions imposed to safeguard privacy and/or confidentiality is that relating to studies of the dynamics of populations. In a number of areas longitudinal studies are becoming an important source of information. From the viewpoint of privacy repeated questioning of the same subjects may be regarded as an excessive intrusion. From the viewpoint of confidentiality the preservation of identifiers or the linkage of information from successive surveys may be regarded as undesirable. This is not a problem which is specific to EUROSTAT but is seen as one area of potential development which may be inhibited by concerns with privacy and confidentiality.

An area which frequently gives rise to concern is that of linkage of data. This has been touched on in paragraph 3 above and is not a problem which directly affects EUROSTAT. The non-identifiable nature of almost all the data received by EUROSTAT precludes subsequent linkage of data files and any linkage, e.g. for panel studies, must be undertaken prior to transmission to us. Thus, any linkage of data-files by EUROSTAT without the full knowledge and cooperation of national offices is excluded.

The proposals to give data subjects a right of access to data relating to themselves has been touched upon in the previous section. Most commentators on privacy and confidentiality in relation to statistical data have suggested that the need for this right of access is reduced, or even eliminated, where one is dealing purely with statistical records. The administrative difficulties which would face EUROSTAT in the event of such a right of access being granted appear to be enormous while the value of such a right to the data subject appears to be minimal. Almost all data received from Member States are in a form in which the individual respondent cannot readily be identified and copies of these data files are (presumably) retained by Member States. It, therefore, appears reasonable to suggest that where national legislation may give such a right of

access the requirements to fulfil this right should be satisfied by access only to the national data sets, possibly with provision for amendments to be forwarded to Luxembourg where such appear to be justified.

7. SOME CONCLUDING QUESTIONS

It has been suggested that publicity of steps taken to ensure confidentiality will help to reassure respondents. Is it not also possible that such publicity will inspire or increase the doubts of respondents?

Confidentiality and statistical reliability of sample results are quite closely associated. If the numbers of subjects in a particular group are few enough to endanger the confidentiality of information it is probable that they are also insignificant statistically. Should more use be made of this aspect in reassuring respondents of the statistician's disinterest in information in a form which may breach confidentiality requirements?

Despite the many concerns which have been expressed, official agencies have a very good record of preserving confidentiality in the past. Should this not be stressed more by statisticians?

It has been suggested that respondents to statistical surveys should always be clearly informed as to whether or not response is voluntary. This would be particularly important where the respondent is faced with an official request for information (implying some form of compulsion to respond). Should this proposal be regarded as a matter of absolute statistical ethics or should it be treated as a matter of practical judgement which may yield differing solutions from survey to survey or from country to country?

Has the concern, justifiable though it may be, for privacy and confidentiality gone too far? Or at least, has the time come for statisticians to be less defensive about their role in discussions and should they not stress more the benefits to be obtained from the availability of reliable, relevant and timely statistics? The availability of such statistics should place a minimum burden on respondents and imply the smallest possible invasion of their privacy. If these requirements are achieved, and if confidentiality of the information relating to individuals is respected, the statistician appears to have no need to make any apologies for practising his profession.

ANNEX

A selective bibliography on problems
relating to privacy and confidentiality

- AMERICAN STATISTICAL ASSOCIATION, "Report of the *ad hoc* Committee on Privacy and Confidentiality", *American Statistician*, 31, 1977, pp. 59-78.
- M. BULMER (ed.), *Censuses, surveys and privacy*, London, Macmillan 1979.
- Computers and Privacy*, Cmnd.6353, London, HMSO, 1975.
- Computers: Safeguards for Privacy*, Cmnd.6354, London, HMSO, 1975.
- T. DALENIUS, *Information Privacy and Statistics: a topical bibliography*, Washington DC, US Department of Commerce, Bureau of the Census Working Paper 41, 1978.
- T. DALENIUS AND A. KLEVEMARKEN (eds.), *Personal Integrity and the Need for Data in the Social Sciences* (proceedings of the Hassleby Slott symposium), Stockholm, Swedish Council for Social Science Research, 1976.
- R. DOLL, "Public benefit and personal privacy, the problems of medical investigation in the Community", *Proceedings of the Royal Medical Society*, 67, December 1974.
- Federal Statistics: Report of the President's Commission*, 2 vols, Washington DC, US Government Printing Office, 1971.
- D.H. FLAHERTY, E.H. HANIS and S.P. MITCHELL, *Privacy and Access to Government Data for Research: an international bibliography*, London, Mansell Scientific, 1979.
- A Framework for planning US Federal Statistics in the 1980s*, Washington DC, US Department of Commerce.
- International Commission of Jurists, "The Legal Protection of Privacy: a comparative study of ten countries", *International Social Science Journal*, 24, 1972, pp. 417-583.
- I.S.I. Code of Ethics. Text and discussion, *Proceedings Volume L Book 3*, pp. 527-561, Madrid, 44th Session of the International Statistical Institute, 1983.

- R. JOWELL, A professional code for statisticians? Some ethical and technical conflicts, Buenos Aires, 43rd Session of the International Statistical Institute, 1981.
- Sir CLAUS MOSER, "The role of the Central Statistical Office in assisting public policy makers", American Statistician, 30, 1976, pp. 59-67.
- Sir CLAUS MOSER, "The environment in which statistics offices will work in ten years time", Statistical News, 38, 1977, pp. 1-6.
- Policy Issues in Data Protection and Privacy: concepts and perspectives, OECD Informatics Studies 10, Paris, OECD, 1976.
- Privacy and Computers, Ottawa, Information Canada, 1972.
- Privacy Protection Study Commission, Personal Privacy in an Information Society, Washington DC, US Government Printing Office, 1977.
- Records, Computers and the Rights of Citizens, Cambridge, Mass., MIT Press, 1973.
- Report of the Committee on Data Protection (Chairman, Sir Norman Lindop), Cmnd.7341, London, HMSO, 1978.
- Report of the Committee on Privacy (Chairman, Kenneth Younger), Cmnd.5012, London, HMSO, 1972.
- D. RHIND (ed.) A Census user's handbook, London and New York, Methuen, 1983.
- Security of the Census of Population, Cmnd.5365, London, HMSO, 1973.
- A. SIMMEL, "Privacy", in International Encyclopaedia of the Social Sciences, New York, Macmillan, 1968.
- "The legal Protection of Privacy: a comparative study of ten countries", International Social Science Journal, 24, 1972, pp. 417-583.
- US NATIONAL BUREAU OF STANDARDS, Computers, Health Records and Citizens Rights (by A.F. Westin), Washington DC, NBS Monograph 157, 1976.
- H.P. WYNN. "Freedom of Statistical Information", Journal of the Royal Statistical Society, Series A, 141, 1978, pp. 1-13.

DISCUSSION

Mr. JENSEN (*Danmarks Statistik*) opened the discussion:

The concepts of privacy, confidentiality and data protection were extremely important in the production of statistics, and thus for the activities of central statistical offices and the overall level of information in any country. General statistics which threw light on society formed a significant part of such information. Although there were many differences in this field, there were also many points of similarity from country to country. It was therefore natural that the European Community countries cooperated. Cooperation was also, obviously, in the interest of EUROSTAT. Against this background, *Danmarks Statistik* welcomed most warmly the fact that EUROSTAT organized this conference. It was hoped it could be used to create a common basis for the policies of the individual countries without diminishing the role of statistics in any of them. This presupposed however, that the basic concepts were clearly stated and that agreement was reached on fundamental points of view. Mr. Peare's paper provided a very good introduction to this debate.

This paper strengthened his conviction that the concepts of privacy and confidentiality ought to be precisely defined, otherwise the basic points of view would become all too easily fudged and a muddled discussion would result. It gave briefly more detailed reasons for this view. Firstly, as far as 'privacy' was concerned, Mr. Peare concluded, after a painstaking review, that the concept could best be defined as "the right to be left alone". Of course, many people took the concept to mean this, but in the real world this idea would be practicable only on desert islands - where, obviously, many of the other good things in life were missing. In other words, the concept thus expressed was unrealistic in modern society, which was based to a large extent on the interaction of society and the individual - both with respect to information and otherwise. If the concept were to be used as a basis for meaningful reflections on statistics in general - the aim of which was, precisely, to throw light on social activities - then it would have to be defined more concisely. Everyone knew that it was unpleasant to have to put up with controls and checks, but everyone also knew that such arrangements were to a certain extent necessary and would remain so as long as everyone was not one hundred per cent honest. He proposed the following definitions:

'Privacy is the right of the individual not to be exposed to controls other than those which are the direct result of concrete considerations', or, to put it briefly, 'the right to be free from unnecessary controls'.

The real problem with regard to privacy was that - in the abstract it was difficult to maintain a proper balance between these two important considerations. The problem was first and foremost administrative and political. Less obvious were the reasons why, in the public debate of the issue, the problem had to a certain extent been associated with statistics, since statistics had nothing to do with checks on the individual. When it came to the social functions of statistics, the opposite was rather the case, in that they provided the individual with a basis for checking up on the activities of society.

The question then was what could be done to counter the effects of such criticism.

A main step would be consistent enforcement of the principle of 'confidentiality', but this would require a definition of the concept that would exclude the possibility of passing on basic statistical information, not only to the public but also to all administrative bodies. If exceptions were unavoidable, they should be well-defined and have a firm legal basis.

On the one hand, the concept of confidentiality should, in an international context, apply only to properly identified information on individuals and the economic activities of such individuals. There may be many reasons why information on minority or special groups of one kind or another should be kept secret, but these reasons differ so much from country to country and from one branch of statistics to another that it was not possible to lay down regulations which would be sufficiently comprehensible or well-defined for use at international level. One decisive point here was that considerations of privacy can, by the very nature of things, be of relevance only in connection with identified information, since without it no controls would be possible.

Mr. BAUMANN (*der Bundesbeauftragte für den Datenschutz*) commented as follows:

The first of a new series of meetings to discuss economic, political and administrative questions bearing on statistics had recently taken place in Wiesbaden. He had delivered the opening address, entitled "Data protection and statistics - adversaries or allies?" The meeting concluded that neither

option properly described the data protector's function. In fact, both activities had a duty to ensure maximum data protection. He then reviewed very briefly the main points emerging from the judgment of the Federal Constitutional Court concerning the Population Census Law 1983.

1. The principles were established that:

- the individual had a fundamental right to determine how his personal data should be divulged and used;
- exceptions to this right - e.g. for a compulsory census - were permissible only if it was predominantly in the public interest and there should be a constitutional legal basis for this;
- the use to which these data were to be put should be specifically stated in the legal documentation and their use should be restricted to this purpose;

it was therefore no longer possible to regard data simply as an administrative aid;

- the citizen had a right to an explanation and should be informed about the recording and use of his data;
- no data were 'unimportant', because data were relevant only in the context in which they were used.

One of the main points as far as the Federal Constitutional Court was concerned was that the citizen who was obliged to provide information should be assured that the data he provided would be protected right along the line. The Court therefore argued that any activity on the part of the State which did not help to build up his trust - by making the data processing procedure public and strictly partitioning off the areas of use - would, in the long term, serve only to make people less ready to cooperate because of the doubts they would have.

2. These principles affected statistics in the following ways:

- data collected for statistical purposes should not in principle be passed on for administrative purposes;
- strict personnel and organizational precautions should be taken to ensure that statistical data were not, despite this basic principle, used for administrative purposes;

- the persons obliged to provide information should be given a comprehensive description of the type, scope and purpose of the statistical surveys and their right to information should be guaranteed.
- before any decision was made about the compilation of statistics, the legislator should ascertain whether a system which was less sensitive from the citizen's point of view would be possible.

3. A number of questions inevitably resulted from this, in connection with both data protection and statistics, and these would affect not only the Federal Republic of Germany but also Eurostat and the Community as a whole.

3.1. How did Eurostat and other Community institutions guarantee statistical confidentiality? What legal provision was there for this and on what legal grounds were any unjustified demands for the transfer of data rejected?

3.2. There were rules on the transborder transfer of data (Chapter 3) and the essential principles of data protection (Chapter 2) in the Council of Europe Convention.

An effort should be made - and this would be very reassuring - to ensure that these principles applied not only to the Member States but also to the EC institutions themselves.

3.3. Section 1.3.b. of the OECD Guidelines expressly stated that the guidelines should not be interpreted as preventing the exclusion from their application of data which 'obviously do not contain any risk to privacy and individual liberties'. This rule could hardly be reconciled with the abovementioned principle of the Federal Constitutional Court that there were no 'unimportant' data.

3.4. The German legislator should look into every statistical survey where the provision of information was compulsory in order to find out whether there was an alternative. It would be useful for the Member States to exchange experiences in this field: in this context it was pointed out that in the case of the EC Labour Force Sample Survey it was left to the Member States to decide whether the statistics should be collected on a voluntary basis; this gave rise to problems, because any compulsory information should have a legal basis.

Mr. FORECAST (*CSO London*) *made the point that:*

In the U.K. all inquiries to individuals, with the exception of the census of population, were voluntary. This, he thought, was the best guarantee of privacy since those questioned could reply or not as they wished. The second

technique for ensuring public cooperation was to explain fully why the questions were being asked and how the information would be used. With regard to the census, the questionnaires were fully tested in pilot surveys to ensure that they met with public acceptance. As well as this they had to obtain the agreement of interested Ministers.

Mr. EMBLETON (*CSO Dublin*) commented:

The problem of the publication of data which, of itself, did not reveal confidential information but which, when taken in conjunction with other available information, could do so, was becoming increasingly important. There now existed many non-governmental sources of statistics, some good, some bad and it was impossible for the official statistician to be aware of all their activities. Furthermore agencies connected with the European Communities sometimes had information relating to enterprises, not available to the national statistical services. On the question that concern for privacy and confidentiality had gone too far, he thought that concern for confidentiality could never go too far. From the statistician's viewpoint, the emphasis on privacy might have gone too far, since it was presenting an obstacle to the collection of the required material. The main problem nowadays was one of non-response, due not to a concern over confidentiality, but simply to annoyance at the volume of information being asked of the public.

Mr. DAMANN (*Referatsleiter, Bundesbeauftragter für den Datenschutz, Bonn*) made the following comment:

It was stated in point 2 of the paper that difficulties had arisen when statistical offices in the various Member States had either wanted, or had been required, to make available to EUROSTAT data capable of yielding information about individual holdings or persons. Such difficulties had arisen, it seemed, despite the Commission's allegedly stringent confidentiality safeguards. As an example of such safeguards Council Directive 75/108/EEC of 20 January 1975, Article 10, point 2 was quoted. From the point of view of data protection one was led to wonder whether the existing safeguards were adequate to provide effective confidentiality of statistical data, and in particular whether the protection they afforded was as effective and comprehensive as that providing statistical confidentiality within the Member States themselves. He asked the following quite specific questions: Was there embodied within Community law a principle of statistical confidentiality which automatically applied whenever information about individuals could be obtained from data collected for statistical purposes?

Did such a principle of statistical confidentiality provide a safeguard against any and every disclosure of personal data to departments outside the statistical office? If no such principle of statistical confidentiality existed, were there any moves afoot to remedy this? Was it generally accepted that comprehensive legal protection of this kind was nowadays one of the indispensable bases for statistical operations? As far as the Federal Republic of Germany was concerned, the inclusion of EUROSTAT in the data network would be inconceivable if this were to entail a lower level of data protection for persons obliged to supply information.

The author then replied:

He did not disagree with Mr. Jensen's view that the definition of privacy as "the right to be left alone" was unrealistic as an absolute guide to action in a modern society. However, he felt that it was important for statisticians and others to recognize that the collection of information frequently involved an invasion of privacy which should be openly acknowledged and justified. Public confidence could be created and retained only if it was made clear that statisticians were sensitive to concerns about privacy and that they took all reasonable steps to minimize intrusion into the privacy of others.

In answer to the queries of the German delegates about the steps taken to ensure confidentiality of data at EUROSTAT it was pointed out that, not only did the economy legislation relating to particular surveys include stringent safeguards similar to those quoted in respect of the Farm Structure Survey, but the treaties establishing the European Communities contained specific provisions to ensure the confidentiality of data required by the Community institutions in the course of their work. Reference should be made to Article 47 of the Treaty establishing the European Coal and Steel Community and to Articles 213 and 214 of the Treaty establishing the European Economic Community.

**DATA CONFIDENTIALITY:
CURRENT PRACTICE AND RESTRAINTS IN FRANCE**

J. BEGUE and C. MOREAU
Institut National de la Statistique
et des Etudes Economiques

SUMMARY

In France, the protection of private information collected by means of compulsory statistical surveys is guaranteed by a law passed in 1951 on the obligation to make returns, the coordination of surveys and statistical confidentiality. Statistical offices have always been careful to ensure that the principles laid down by this law are rigorously applied and have quite naturally extended these principles to cover other sources of data which they have been called upon to process, such as administrative files to which they have been given access and the results of optional surveys.

The 1978 law on automatic data processing, data files and civil liberties, which was aimed at protecting the public against the dangers arising from a proliferation of statistical offices without, however, fundamentally altering their application of the principles of confidentiality.

In this paper the presentation of the legal framework established by the two laws already mentioned is followed by a description of how that framework is applied in practice to statistical operations. This entails a review of the following aspects: the physical security of data, the dissemination of statistics from administrative files, the utilization of user identifications, the risks of indirect identification in depersonalized files or statistical tables and the linking up of files.

0. INTRODUCTION

The principle of statistical confidentiality in France is embodied in a law of 1951 introducing the system of compulsory public surveys. In this law a distinction is established between private information, which may not be disclosed under any circumstances, and personal data of an economic or financial nature whose disclosure is not, under certain circumstances, entirely prohibited.

In theory, the data protection policies which statistical offices have implemented in order to stay within the law apply only to compulsory surveys but they have been extended quite naturally to cover the other sources of data on which statistical operations depend, namely optional surveys and administrative files, both of which may moreover be subject to other secrecy requirements.

In strengthening the legislation relating to the protection of personal data, the 1978 law on automatic data processing, data files and civil liberties also imposed new restrictions on statistical offices, though it hardly entailed any significant changes regarding the application of secrecy requirements.

After describing the legislation pertaining to public statistics work in France, this paper goes on to describe some practical aspects of the implementation of these requirements.

1. THE LAW ON STATISTICAL CONFIDENTIALITY

The system of compulsory statistical surveys was introduced by Law No. 51/711 of 7 June 1951 (amended by several later versions) on the obligation to make returns, the coordination of surveys and statistical confidentiality.

Alongside the obligation to make returns this law provides for the coordination of public statistical surveys with the aim of keeping their number within reasonable bounds. Above all the law provides a guarantee for respondents to these surveys that any data they supply will be protected.

The provisions to this effect contained in Article 8 of the Law vary depending on whether the data involved are "personal data concerning the private and family life of individuals including, in general, the circumstances and behaviour of such individuals in the private sphere" or whether they are personal data of an economic or financial nature. Unless countervailing provisions exist in the code of criminal procedure, the former "may not under any circumstances be disclosed by the institution to which they have been entrusted". As far as personal data of an economic or financial nature are concerned, the Law simply states that such data "may not under any circumstances be used for purposes of tax inspection or for the imposition of economic sanctions on an individual" and goes on to impose certain conditions on the disclosure of data of this nature.

In the section concerned with the need to ensure confidentiality, the Law of 7 June 1951 invokes an article of the code of criminal procedure (Article 378) which deals in broad terms with the question of professional secrecy. This article provides that: "(...) all (...) persons who, as

a result either of their station in life, their profession or any temporary or permanent position they may hold, are entrusted with information of a confidential nature and who disclose that information to third parties, shall, unless obliged or authorized by the law to act as an informant, be liable to one to six months imprisonment and a fine of between 500 francs and 8000 francs".

Although in theory the provisions of the Law of 7 June 1951 apply only to compulsory statistical surveys, the public statistical offices afford the same confidential status to data collected from the optional surveys which they are sometimes called upon to carry out (regionally based surveys, opinion polls, short-term economic surveys). These surveys are optional because of the personal nature of the questions they contain.

Moreover, if in the 1950s most statistics were compiled on the basis of data collected by means of surveys (population censuses, polls) and thus were subject in principle to the 1951 Law, it is nevertheless true that data contained in the administrative files of various government departments were already being used to produce statistics in certain areas, a practice which has increased steadily since then. For example, the use of civil status documents has long provided a means of compiling up-to-date demographic statistics, while from 1950 onwards the annual declarations of wages and salaries which employers are required to submit to the tax and welfare authorities (and which we shall return to later) have been a major source of data for the compilation of statistics on wages and salaries.

Strictly speaking, this use of administrative data does not fall under the confidentiality requirement of the 1951 Law. What does apply, however, are the general requirements for the protection of personal data deriving from the obligation to respect professional secrecy. Furthermore, such use of administrative data is subject to any confidentiality requirements which a particular government department may have. Clearly, such requirements, while not always amounting to an absolute prohibition, nevertheless place severe restrictions on the use of administrative data. However, while such requirements must be respected by statistical offices whenever they are given access to certain administrative files, they are not generally any more restrictive than the requirements which would result from a strict application of the 1951 Law.

For more than 30 years, then, this law has been the point of reference for virtually all statistical operations regardless of the origin of the data being processed. Needless to say, the law applies to both indirect and direct disclosure of data, even though this is not explicitly stated. Indirect disclosure of data may result from the

publication of statistical tables which, by virtue of certain of the criteria they use or as a result of cross-checking, could lead to the identification of the statistical units concerned; it could also result from the dissemination of personal data files which have in theory been depersonalized but which could in practice still lead to certain of the individuals concerned being recognized.

As far as direct disclosure of data is concerned, we have already seen that the 1951 Law rules out entirely the dissemination of private information (the period of application of this prohibition now being 100 years, since the passing of the 1979 Law on public records) and only admits the disclosure of personal data of an economic or financial nature provided the conditions imposed by the law are respected. This is a distinctive feature of the secrecy requirements for business statistics which we will not discuss any further here. Apart from this, there is no significant difference as regards data secrecy requirements between the two spheres (private information and personal data of an economic or financial nature), though it should be mentioned that in the case of individual entrepreneurs and members of their families who are employed in the business, no clear demarcation line between the two exists.

2. THE 1978 LAW ON AUTOMATIC DATA PROCESSING, DATA FILES AND CIVIL LIBERTIES

The strict application of the principles established or assimilated by the 1951 Law never gave rise to any real problem as regards the compilation of social or demographic statistics. There is no known case of violation of statistical confidentiality and the statistical offices, who have always taken for granted that data relating to respondents to surveys or employees must be kept confidential, did not feel that the progress of their work was impeded in any serious way.

Thanks in large part to the progress made in the 1960s and above all in the 1970s in the field of automatic data processing, it seemed to the statistical offices that their own work could expand without this posing additional risks for the individuals seen as statistical units. Indeed, whatever the techniques of data file administration and statistical compilation, they were confident that the existing legislation allowed them to guarantee the confidentiality of the data which they held.

Be that as it may, this legislation was extended by Law No. 78-17 of 6 January 1978 which, in France as in other countries, was passed in order to protect the public against the dangers arising from the growing computerization of society. This law, which applies to any and every collection, storage and processing of personal data (in

particular by computer) regardless of the origin of such data (administrative or otherwise) and regardless of its purpose, provides for no exception whatever for statistical offices or, more generally, research institutions.

It is interesting to note that one of the reasons - though not the only one - for introducing this law was the public anxiety aroused by a media-orchestrated campaign opposing the planned computerization of the national identification register for natural persons. This register is administered by INSEE (the National Economic Studies and Statistical Institute) as part of its responsibility for managing the large-scale files which cut across several government departments, and entries are restricted to details of civil status (surname, first names, data and place of birth) and an identification number corresponding to each person on the register. The planned computerization measures gave rise to fears that a kind of super-file might be created containing all the data held by government departments, thereby constituting a threat to civil liberties.

A detailed discussion of the 1978 Law is provided in the paper presented by Mme. Lenoir, who is Director of Regulations at the National Commission for Data Processing and Civil Liberties (CNIL), an organization called into being by the 1978 Law. Nevertheless it is necessary to mention here the new restriction which this law imposes on statistical offices.

No automatic processing of personal data, i.e. data which, as defined by the law, may lead to direct or indirect identification of the person or persons concerned, may be undertaken without:

- the prior passing of a regulation (and sometimes even of a law), after consideration of the reasoned opinion of the CNIL, when the data to be processed are for use by the State, by a public institution or by a regional or local authority,
- a prior declaration to the CNIL in all other cases.

Nevertheless statistical operations which quite obviously pose no threat to either privacy or civil liberties, may, depending on the standards set by the CNIL, be subject to what is known as a "simplified declaration" procedure. Thus one category coming under this simplified procedure is the automatic processing for statistical purposes of personal data relating to natural persons and concerning their status as individual entrepreneurs or members of their families working in the business, provided that this processing is undertaken by the public services and bodies called into being by the Law of 7 June 1951. Fitting into the same category is the processing of data on natural persons collected by means of surveys carried out by the State and

by public institutions of an administrative nature, provided these surveys use relatively low sampling fractions (i.e. a sample comprising less than 5% of the population under survey) and provided the questions asked in the survey pertain to routine categories of information which are not "sensitive".

For automatic data processing carried out by public bodies and requiring an act of legitimation, that act of legitimation must in particular specify:

- the purpose for which the data are being processed,
- the categories of personal data which are to be recorded and the categories of authorized recipients who are to have access to these data.

When data are to be processed using the national identification register for natural persons, a decree of the Council of State, passed after considering the opinion of the CNIL, is required. Certainly, the existence of a national identification number makes the linking up of files far more of a technical feasibility.

On the technical level, the law on automatic data processing, data files and civil liberties stipulates that all persons who process or cause to be processed personal data must "take all necessary precautions to maintain the security of... (those) data and in particular to ensure that the data are not distorted, damaged or disclosed to unauthorized third parties".

Besides these provisions concerning the automatic processing of personal data, the Law of 6 January 1978 provides a certain number of safeguards for persons supplying government departments or any other bodies with information concerning themselves regardless of whether that information is to be automatically processed or not.

To this effect the Law of 6 January 1978 provides for a "right of access" allowing persons to acquaint themselves with the data which concern them in personal files, and establishes the principle of the right to correction if these data should prove to be wrong.

The law provides that (article 27):

"persons requested to supply personal data must be informed of

- the compulsory or optional nature of replies
- the consequences which the withholding of replies might have for them

- the natural or legal persons who will receive the data which they supply
- their right of access and their right to correction

When personal data are collected by means of questionnaire, these requirements must be stated on that questionnaire (...)"

The law also prohibits the collection and storage of "personal data which reveal, either directly or indirectly, a person's racial origin, political, philosophical or religious beliefs or trade union membership", except in cases where this is felt to be in the public interest or where a decree of the Council of State, based on the proposal or assent of the CNIL, has been passed.

Although the Law of 6 January 1978 certainly provides individuals with a considerable measure of protection as regards the use of personal data concerning them, it has nevertheless not altered in any radical way the policies applied prior to the law by statistical offices with regard to the processing of personal data for statistical purposes.

With the exception of the right of access and the right to correction, the protection already afforded under the Law of 7 June 1951 to those supplying information in compulsory public statistical surveys was essentially the same as that introduced on a general level by the Law of 6 January 1978.

The Law of 6 January 1978 does indeed provide legal protection for the data collected in optional surveys and ensures observance of certain rules with precise provisions, but these rules were already an integral part of the professional ethics of statistical offices. The main impact of the law on statistical offices was the requirement that a certain number of administrative operations be carried out prior to the surveys and processing planned.

Only the principle of end use may be a source of difficulties for the statistical offices. This principle has already had negative repercussions during the 1982 population census, since the CNIL refused to grant authorization for the use of data from the Directorate-General for Taxes' file on domestic rates for checking that the collection of housing returns was complete, on the grounds that this was not a valid end use for the fiscal file.

3. PRACTICAL ASPECTS OF OBSERVING DATA CONFIDENTIALITY

Having outlined the legislation on the protection of personal data concerning the circumstances and behaviour of individuals in the private sphere, it is appropriate at this

stage to see how such legislation is applied in practice in the collection and dissemination of INSEE data. What follows is not a comprehensive and detailed review, but a brief description of some of the principal types of measures adopted.

(a) First, as a matter of record only, since a full description would be too long and wearisome, precautions are taken at the various premises of the institution to guarantee the physical security of data: surveillance of entrances and exits, controlled access to computer rooms and especially tape libraries etc. Generally speaking, the following approach is adopted for protecting confidential data stored on magnetic files:

- except in special cases, all INSEE files can be consulted by all INSEE users and only by INSEE users;
- only users responsible for a group of files are authorized to amend or destroy them;
- files of general interest may be consulted by everyone and software of general interest is open to all users;
- exceptions may be made to these rules, but they must satisfy definite procedural requirements;
- lastly, access to certain highly confidential files or groups of files is strictly reserved for certain users.

At the data base level, protection of access to the different elements is normally taken into account when designing the management tools. In addition, an external non-INSEE system for protecting all magnetic tape or disk files has been placed on the IBM 3033 computer owned by the Centre National d'Informatique de Paris I.

(b) Processing by INSEE of the data collected by government departments in the course of their management tasks is generally subject to a protocol which lays down the conditions under which the institution may disseminate statistical results obtained from personal data supplied. This applies, for example, to the processing by INSEE of the annual income returns (D.A.S) supplied by the Directorate-General for Taxes which are used to prepare statistics on annual income by activity of the firm, location and characteristics of the income earner. Very often the protocols contain the same rules governing the dissemination of statistics prepared from these data as those applied to the data from statistical

surveys. Disclosure of personal (non-identifiable) information is always subject to the prior agreement of the originating government department.

- (c) During the processing of these administrative files the identifications of the individuals concerned are removed, as in the case of survey files, as soon as preparation of the statistical operating files has been completed. This is the case, for example, with the information system on civil servants operated jointly by the Directorate-General for the Administration of civil servants and the INSEE. The system involves the centralized processing of the computer salary files of government departments which makes it possible to collect statistical data on civil servant employment and salaries in government departments. The information collected is always depersonalized. Each individual is identified by his number in the national identification register for natural persons (NIR), but in no case does this provide a clue to the individual's identity. It is used solely for processing the information collected. The personal information held centrally by the INSEE cannot, whatever its origin, be disclosed or even consulted for other INSEE work or for the work of other departments. It is not until all the preliminary processing has been completed that copies of the files, without the identifications of course, can be made available to users.

In certain cases, however, for example when it is necessary to monitor a sample of individuals over a period or to regroup data on an individual contained in several files of the same type, the identifications are kept at INSEE for a longer period. In the case of the permanent demographic sample, for example, whose primary aim is to provide statistics on population movements and changes of occupation on the basis of data from successive general population censuses, civil statistics documents and the electoral register, the NIR is used as a permanent identification in the computer file produced. The names and first names of the persons in question are not kept in the computer file. And, naturally, the INSEE discloses only data which precludes direct or indirect identification of the persons in the sample and which therefore do not contain the NIR.

Similar provisions also exist for the study of mortality by socio-professional category.

It should be noted that these two operations involving the use of the national identification register for

natural persons were the subject of a decree of the Council of State, passed after considering the opinion of the CNIL, pursuant to Article 18 of the Law of 6 January 1978 on automatic data processing, data files and civil liberties referred to above.

Similarly, during the processing already described of annual returns made by employers to the tax administration, the use of the NIR (which is authorized for income files) is extended in order to permit grouping of salaries paid to the same person by one or more employers during a given year which appear on different income returns, as well as to allow the career of individuals in a sample of income earners to be followed over successive years. Here too, the information produced of course no longer contains any factor which might allow identification of the persons in the sample. However, simply removing an identification number is not always sufficient, as we shall see below.

- (d) The legislation laying down rules for the protection of personal data collected for a statistical purpose does not prohibit subsequent dissemination, particularly for research needs, but stipulates that such personal data must be strictly depersonalized, i.e. must not permit identification, even indirect identification, of the individuals concerned.

The best example in this respect is the general population census whose results are of interest to a great number of external bodies (local authorities, regional public bodies and study organizations). The INSEE laid down precise arrangements for the disclosure of the detailed files of the 1982 census on the basis of the recommendations of the National Commission on Data Processing and Civil Liberties in a resolution adopting a standard agreement protocol between the INSEE and the communes laying down rules governing the use by communes of the data from the 1982 census. Two types of processing are available for the files, depending on the precision of the geographical detail in question. In all cases, the number of the building, the number of the dwelling and the day and month of birth of the inhabitants are removed. In addition, in files where identification by district is retained:

- the code showing the relationship with the head of household is replaced by a two-category code (head of household/other):
- only a two-digit "profession" code is included, i.e. the professional activity is indicated only at the socioprofessional category code level, which is highly aggregated.

In files prepared from the one-in-four survey of the census returns in which no indication of geographical location at a fine level (district) is retained, and in the one-in-twenty survey files, the district identification codes are erased. In the case of rural communes, the commune and locality codes are erased. Finally, as regards the individual, the identification codes of the area of the place of work are erased if they appear on the return.

Furthermore, when publishing statistical results relating to natural persons, precautions are taken to avoid the risk of indirect identification arising from the small number of persons appearing in certain sections of the tables. In general, one applies by analogy a rule on the dissemination of results concerning undertakings established following an agreement in the National Statistics Council (a consultation body on statistical programmes set up by the 1951 Law) according to which no result relating to a group of less than three units may be disseminated.

And finally, one should note that samples of persons or households taken from the population survey cannot be disclosed to external bodies.

- (e) If one wishes to guarantee strict depersonalization of the data processed one must use somewhat complicated but proven methods when comparing data from different sources. One of the most "classical" is the "double-blind" method. It was used to compare the file relating to the 1980 survey on French health expenditure with the files of the national employees' sickness fund (CNAMTS) to check the number of cases of hospitalization declared in the survey. The procedure developed for this purpose by the INSEE and approved by the CNIL was as follows:

1. INSEE provided the CNAMTS, on a manual data medium, with a list of the names of persons who had declared in the survey that they had been hospitalized under the general social security system as insured persons or as their dependants.

In addition to the number of the survey questionnaire and serial number of the person in question, the list contained the name, first name, sex, month and date of birth of the person hospitalized during the three-month period of the survey or during the previous year. If the person hospitalized was a dependant, the name, first name,

sex and age of the insured person on whom they were dependent were also supplied to the CNAMTS. The list also contained the place and the starting and finishing dates of hospitalization if known.

It must be stressed that these data were supplied with the aim of helping the CNAMTS in the search for data which it already holds on the persons in its field of responsibility.

2. CNAMTS undertook to provide CREDOC (Research Centre for the Study and Observation of Standards of Living), a research body conducting the survey jointly with the INSEE and responsible for processing this part of the data, with a list showing the information supplied by the primary sickness funds on hospitalizations recorded in the health survey with, as the sole link-up, the number of the questionnaires supplied previously by the INSEE, but without the name and address. As a result, CREDOC received only non-nominative data.
3. INSEE provided CREDOC with depersonalized data from the health survey, the only reference being the questionnaire number without the name and address. Only a commune category code and ZEAT code, a highly aggregated area code containing several administrative regions, were included.
4. CREDOC linked up the data from the health survey supplied by the INSEE and the sickness fund data on hospitalizations supplied by the CNAMTS. The only key for the link-up was the questionnaire number, and the linked data were therefore completely depersonalized, as were the constituent data.
5. In order to avoid any subsequent individual identification, CREDOC was not able to return to CNAMTS the detailed tape which it produced, but only the statistical tables.
6. CREDOC supplied INSEE with the detailed linked tape so that the Institution would have all the data available on the sample which it itself had compiled, with a view to further studies.

Earlier, the INSEE had destroyed the individual identification codes in its possession (address files) with the same aim of ensuring the anonymity of the files in circulation.

Law No. 51-711 of June 7, 1951 (*) on Legal Obligation,
Coordination, and Secrecy in the Field of Statistics

The National Assembly and the Council of the Republic have deliberated,
The National Assembly has passed,
The President of the Republic has enacted the following law:

Paragraph 1 - (Amended by Decree No. 72-1103 of December 8 1972)

A National Board of Statistics attached to the National Institute of Statistics and Economic Studies shall be established. It shall be responsible for the coordination of the statistical surveys of government departments and institutions, excepting internal statistical work which does not involve the assistance of persons from outside government service. This Council shall establish yearly a programme covering the totality of surveys planned for the year and shall set their approximate dates and the time-limits to be given to the natural and legal persons for response. The programme and its implementation procedures shall be determined by the Minister having supervisory authority over the National Institute of Statistics and Economic Studies.

The membership of the National Board of Statistics and its operating procedures shall be established by a decree which shall specify in particular the manner in which the representation of the appropriate natural and legal persons and of the Parliament and the Economic Council will be ensured.

(*) Amended by the Appropriation Bill for 1969 No 68-1172 of December 27 1968 (Official Gazette of December 29; page 12349), the Law of Public Records No. 79-18 of January 3 1979 (Official Gazette of January 5 1979; page 43), Decrees No. 59-1350 of November 16 1959 (Official Gazette of November 29 1959; page 11451), No. 62-275 of March 12 1962 (Official Gazette of March 14, 1962; page 2662), No. 72-1103 of December 8 1972 (Official Gazette of December 13 1972; page 12830).

The National Board of Statistics shall be headed by the Minister of Economic Affairs, who shall delegate his powers to the Chairman of the Board.

Paragraph 2 - Every statistical survey conducted by the public authorities, excluding internal statistical work not involving the assistance of persons from outside government service, shall be submitted for prior approval by the Minister having supervisory authority over the National Institute of Statistics and Economic Studies and by the Minister responsible for the area in which the survey is to be taken.

Approval may only be granted if the survey fits into the framework of the program provided for in the preceding paragraph, if it is provided for by a specific law, or if there is a situation of imperative need or emergency.

Paragraph 3 - The natural and legal persons shall be required to respond accurately and within the set time-limits to the statistical surveys having received the approval defined in Paragraph 2.

Paragraph 4 - (Amended by the Appropriation Bill for 1969 No. 68-1172 of December 27 1968).

Professional or inter-professional organizations may be duly appointed by the public authorities to act as intermediaries in the conducting of statistical surveys. Authorization shall be granted or withdrawn by a joint order of the Minister having supervisory authority over the National Institute of Statistics and the Minister responsible for the sector concerned.

When a questionnaire having obtained approval is disseminated in this way by a duly appointed organization, the concerned parties shall have the option of responding through the said organization or directly to the public authority in charge of the survey.

Duly appointed organizations shall give the information they have gathered to the public authority in charge of the survey within the time-limit of the contract. ("arrêté d'agrément" - order of appointment).

Paragraph 5 - (Repealed by Decree No.62-275 of March 12 1962).

Paragraph 6 - (Amended by Law No. 79-18 of January 3 1979).

Subject to the provisions of Articles 29 and 89 of the Code of Criminal Procedure, personal information contained on questionnaires having obtained the approval provided for in Paragraph 2 and which concerns personal and family life and,

in general, facts or behaviour of a private nature, shall under no circumstances be released by the depositary department before a period of one hundred years since the carrying out of the census or survey has elapsed.

Personal information of an economic or financial nature contained in questionnaires having obtained the approval provided for in Paragraph 2 shall under no circumstances be used for internal revenue investigations or surveillance of illegal economic activities. Departments in possession of information of this type are not bound by the obligations provided for, in particular, by Paragraph 31 of the Law of July 31 1920, on the Establishing of the General Budget for the 1920 Fiscal Year, amended by Paragraph 30 of Law No. 45-0195 of December 31 1945, and Paragraph 15, second Subparagraph, of Enactment No. 45-1483 of July 30 1945.

Employees of public authorities or organizations called upon to act as intermediaries for surveys according to the terms established in Paragraph 4 are subject to the obligation of professional secrecy provided for in Article 378 of the Penal Code.

The censuses and statistical surveys conducted in accordance with the provisions of the present law shall be considered as public records.

Paragraph 7 - (Amended by Decree No. 59-1350 of November 16 1959)*

In the event of failure to respond, subsequent to formal notice, within the time-limit set by the said notice, or in the event of a knowingly inaccurate response, natural or legal persons may be subject to an administrative fine imposed by the Minister having supervisory authority over the National Institute of Statistics and Economic Studies as advised by the National Board of Statistics.

The amount of the first fine thus incurred by a natural or legal person shall not exceed 10 000 francs.

* The fines provided for by this Paragraph are expressed in francs of 1959, that is, in centimes.

In the event of a repetition of the offence within a period of three years, the amount of the fine shall be increased to a minimum of 20 000 F and a maximum of 100 000 F for each infraction. For firms of over one hundred salaried employees, in the event of a repetition of the offence within a three-year period, the fine shall be 200 F minimum and 1 000 F maximum per employee, not exceeding a total of 200 000 F.

These fines shall be collected according to the terms provided for by the provisionally applicable Law of March 13, 1942 concerning the collection of public claims other than those of internal revenue and State-owned properties.

However, every failure to respond, subsequent to formal notice and within the time-limit set by the said notice, and every knowingly inaccurate response to questions concerning personal and family life, shall be punished by a fine of 100 to 600 F and, in the event of a relapse, of 200 to 12 000 F. This fine shall be imposed according to the procedure provided by the Enactment of November 2 1945, regarding the collection of cumulative fines.

Paragraph 8 - All provisions laid down by law or regulations to the contrary of the present law are hereby repealed.

Paragraph 9 - The present law shall be applicable in the Overseas and Associated Territories.

The rules of application of the present law shall be prescribed by decrees of the Council of State based on the Report of the Minister for Economic Affairs or on the Joint Report of the Minister for Economic Affairs and the Minister for Overseas France.

The present law shall hereby be enforced as the law of the State.

DISCUSSION

M. WEIDES (*Service central de la Statistique et des Etudes économiques, Luxembourg*) opened the discussion:

He remarked on the parallelism between the situations in France and Luxembourg as regards ensuring the confidentiality of data obtained for the purposes of official statistics as well as the new problems raised by the proliferation of data banks containing personal data. This position arose naturally from the fact that Luxembourg drew largely upon the French and Belgian experience in drafting the relevant legislation. He noted that in France a higher degree of confidentiality was stipulated for personal data relating to an individual's private life than for economic or financial matters. In Luxembourg the levels of confidentiality of the two classes of information were treated as equal and he asked whether there was any value in ascribing different degrees of confidentiality to different classes of data. The proliferation of data banks, both public and private, and the fear of their misuse and the cross-linking of information had led the authorities in Luxembourg, as in France, to introduce very restrictive rules governing their operation. These concerned the initial collection of data, its purpose, its storage, its processing, its transmission to third parties and its ultimate use. These rules tended to make the task of the statistician very difficult. Thus persons conducting inquiries were obliged to inform their subjects of the exact uses to which the information would be put, of their right of access and rectification and whether the inquiry was statutory or voluntary. All this introduced a quite unnecessary atmosphere of mistrust into the inquiry. He asked whether provision could be made in future legislation to take account of the special position of the statisticians and remove these artificial barriers to their work.

During the course of the discussion other questions were put to Mr. Begué concerning the use of tax files as a source of statistics on individuals and the degree of cooperation between the statistical services and the Commission nationale de l'informatique et des libertés.

The author replied to the discussion:

He said that the different rules regarding confidentiality of personal and economic data provided for statistical purposes arose from the legislation under which such data were collected. Data on individuals could in no circumstances be released. The legislation relating to the collection of statistical data on enterprises did not, however, specify a similar complete embargo on their release. Tax files were not used as a source of statistical data on individuals. Good relations had been consciously developed between the statistical services and the Commission nationale de l'informatique et des libertés. The Commission recognised that the use made by the statisticians on information relating to individuals was not such as to threaten liberty or privacy. It therefore did everything possible to facilitate the work of the statisticians. The collection of statistics in France was not greatly inconvenienced by the legislation on the protection of individualized data.

STATISTICS AND INDIVIDUAL LIBERTIES

Mme. N. LENOIR

Commission Nationale de l'Informatique et des Libertés

0. INTRODUCTION: SCOPE OF THE PROBLEM

It appears odd at first sight that there should be any link between statistics and individual liberties.

Indeed:

- statistics deal not with individuals but with groups or social categories;
- statistical data are the product of anonymous aggregates and therefore preclude identification of individuals;
- the aim of statistical data is to provide more knowledge, in contrast to management data which serve as a basis for decisions on individuals (wages and salaries, granting of social security benefits, levying of taxes etc.).

Nevertheless, this state science *par excellence* (if only by etymology) has always been linked with the individuals who provide the information on which statistics are based.

- Ever since statistical surveys have existed, it has been necessary to overcome the reluctance of individuals to complete survey questionnaires. As a result, there have sometimes been incidents during the general population census, although the latter is now a well established part of French life.
- It is true that the growth in the number of surveys and of statistical application of different types of data, facilitated by the advent of data processing, increases the risk of the violation of privacy by the disclosure, whether intentional or not, of confidential data.
- Furthermore, statistics on production or services, insofar as they introduce an element of standardization, are often seen as a particularly burdensome means of monitoring the work of individuals.
- Finally, like any science, statistics may be abused or contain serious errors of methodology and thus serve to lend legitimacy to actions which are pernicious or inopportune.

These risks of abuse, as well as the recognition of the wish of private individuals to preserve their privacy, have led to the establishment of protection for statistics, particularly public statistics, either by means of legislation or the introduction of practices on the initiative of researchers and statisticians, particularly those of the National Economic Studies and Statistical Institute (INSEE)(1) .

The legal framework for statistics in France, which forms the subject of this report, is relatively restrictive (1 and 2). One should even say that it sometimes acts as a brake on the development of research, and consideration should therefore be given to ways of relaxing this framework (3 and 4).

1. LEGAL FRAMEWORK: LAW OF 7 JUNE 1951 ON THE OBLIGATION TO MAKE RETURNS, THE COORDINATION OF SURVEYS AND STATISTICAL CONFIDENTIALITY AND ITS IMPLEMENTING DECREE

The years following the Second World War, which were a period of reconstruction and growth, saw the introduction of modern planning techniques and the concomitant development of the government statistical apparatus, notably with the creation of the INSEE in 1946 as a directorate of the Ministry of Economics and Finance. This established statistics as an aid to governments in shaping their policies on the basis of prior information on the social and economic fabric.

In the absence of electronic data processing, the favourite method of gathering data at the time was the use of surveys. But since any survey was likely to be viewed by individuals as an investigation if not an intrusion into their private lives, legislation was introduced in 1951 permitting public authorities to conduct obligatory surveys on condition that a guarantee was given to those concerned that the information collected would remain confidential. Refusal to complete the questionnaires for surveys declared obligatory by the Minister of Economics and Finance carried a maximum fine of FF 10 000 imposed by the Minister. In order to arrive at a social consensus on the survey programmes to be conducted each year, the 1951 law set up a Statistical Survey Coordination Committee, which was replaced in 1972 by the National Council on Statistics. Under a Decree of 17 July 1984, the latter was itself replaced by the National Council on Statistical Information (CNIS). This consultative body, whose secretariat is

(1) These practices are described in Mr. Begué's report.

provided by the INSEE has the same composition as its predecessors: it brings together the "vital elements of the nation" with members of Parliament, members of the Economic and Social Council, representatives of diverse public authorities, trade unions, local authorities, researchers, the association movement etc. The most significant change introduced by the 1984 reform is that it widens the field of action of the new Council, which now no longer simply delivers its opinion on the annual survey programme of the public services but also discusses plans for processing management-file data for statistical purposes, as well as the computerized processing of statistical surveys or management data. Indeed, for reasons of cost, reliability and availability, statistics are increasingly prepared from existing files, which have therefore become information repositories of national importance.

2. GENERAL PROTECTION OF DATA: LAW OF 6 JANUARY 1978 ON AUTOMATIC DATA PROCESSING, DATA FILES AND CIVIL LIBERTIES

But in addition to public statistics, and even surveys and data processing subject to the opinion of the CNIS, there are numerous statistical applications and studies for commercial, research and other purposes. Statistics are not the exclusive domain of government statisticians, but are used by all types of bodies with the widest variety of statutes and objectives. Until recently, one had to rely on the voluntary observance of professional ethics, such as the code of ethics laid down by the survey institutes belonging to the SYNTEC professional organization.

The Law of 6 January 1978 on automatic data processing, data files and civil liberties, which deals with the protection of data in general, applies primarily to statistics and makes no provision for exceptions or derogations. One might have thought that since statistics are anonymous, the law would not apply to them. This is not the case, since automated data processing systems, and even manual files, are subject to the law if the data collected are name-linked, i.e. allow direct or indirect identification of a natural person. Because of the principles which it sets out, the 1978 Law provided the opportunity for laying down a code of statistical ethics, and it was up to the National

Commission for Data Processing and Civil Liberties (CNIL) (1), an independent administrative authority, to work out precise rules.

These rules, which stem from the 1978 Law or the CNIL, apply in addition to those laid down by the Law of 7 June 1951. They are all the more important when the latter law does not apply, in which case they go some way to plugging a legal loophole.

The constraints placed on statisticians by the law on automatic data processing and civil liberties are of various types.

Firstly, the law imposes the requirement of transparency. As in the case of all automated processing of name-linked data, processing of statistics prepared from personal data from surveys, or from existing files, must be declared to the CNIL. The CNIL examines the measures taken to preserve the confidentiality of the data processed and ensures that the processing is consonant with the tasks of the body submitting the declaration.

The 1978 Law, on the other hand, imposes requirements concerning the methods used. Needless to say, it prohibits the collection of data by any fraudulent, dishonest or illicit means. But above all it establishes the principle of "informed consent" for respondents, particularly when questionnaires are used. Respondents must be informed whether replies are compulsory or optional, and of the consequences of failure to reply. They must also be told for whom the information collected is destined, and that they have a right to access and correction. This right subsists for as long as the data are kept in a name-linked form. This was pointed out by the CNIL, for example, in the case of census questionnaires retained by the administration which

(1) Created by the 1978 Law, the CNIL has 17 members (members of parliament, members of the superior courts, members of the Economic and Social Council, and appropriately qualified persons). Its statutes give it total independence *vis-à-vis* the executive, and its members cannot be bound by instructions from any authority. The CNIL ensures that the procedures and principles laid down in the law are respected. In particular, it:

- records all automatic processing of data and delivers opinions on the processing of data by the public sector;
- lays down "simplified" standards governing the most common types of automatic data processing;
- ensures that individuals are able to exercise their right of access to data held on them in magnetic or manual files.

bear a reference number to a file containing the names and addresses of the persons covered by the census.

Lastly, when delivering opinions of the processing of data from the public sector, the CNIL works on the principle of end use, according to which a body is allowed to process only those data relevant to its work. The processing of statistics, like other operations, is subject to the observance of this principle which in effect helps strengthen professional secrecy. The principle is applied both to sampling and the use of management files. Both are allowed only if the "source" files are connected with the final purpose of the research or statistics in order to avoid the use of the final purpose as justification for the exchange of confidential data between different statistical bodies which would not normally have the right of access to such data.

3. THE CNIL'S "CASE LAW": BALANCING THE RIGHTS OF INDIVIDUALS SUPPLYING INFORMATION FOR RESEARCH OR STATISTICS AGAINST THE INTEREST OF STATISTICS

The need to defend the interest of statistics as a scientific tool for obtaining knowledge is a constant preoccupation of the CNIL. However, the CNIL cannot go as far as to obviate the application of the 1978 Law, which does not allow any exemptions for the processing of statistical data, not even for the right of access on which the Council of Europe Convention (1) allows restrictions to be imposed when files are used for statistical purposes or scientific research. Nevertheless, the Commission has sought to simplify the procedures for declaring statistical files and to facilitate the use of certain administrative data.

The first step taken by the CNIL was to issue "simplified standards" laying down the conditions to be met by statistical processing operations in order for them to qualify for a simple "declaration of existence". Processing operations outside the bounds of the standards must be declared in the normal way or, in the case of processing operations by the administration, submitted to the CNIL for an opinion. Two standards were adopted in 1981, and a third was in the process of being adopted in November 1983. The first standard, covering individual enterprises, applies to sectors. The second is concerned with sample surveys carried

(1) Convention for the protection of individuals from the dangers inherent in the automatic processing of personal data, opened for signature in January 1981 and ratified by France at the end of 1983.

out by the government and public administrative bodies, but is restricted to surveys using samples of less than 5% of the population covered by the survey. The third standard has a much broader scope and, in addition, has the advantage of providing for changes in statistical methods. It allows for several possibilities: the use of an administrative source for survey samples or for obtaining statistical results; the coupling of data from one sample survey with information taken from an administrative file; and the comparison of administrative files on the condition, of course, that the secrecy regulations applying to each administrative unit and its particular field are observed.

The CNIL then allowed survey institutes to submit a single annual declaration per series of surveys instead of one declaration per survey, which would have been too onerous.

But the Commission, in praetorian style, also felt it appropriate to relax certain conditions laid down by the law, whose strict application would have impeded research. Firstly, it exempted certain social science and medical researchers for the requirement of obtaining the prior consent of handicapped patients on whom certain data were collected, on the grounds of the psychological reactions that certain questions to such persons might produce. On the other hand, with political surveys, the Commission showed itself to be much more strict and requires any bodies carrying out such surveys to obtain the written agreement - as they are required to by law - of the respondents. The survey institute argued that this procedure would be certain to produce biased survey results, since only those with deep political convictions were likely to agree to sign survey questionnaires. As a result, they asked for an exception to be made.

Secondly, the CNIL developed the concept of "broader end use" on the basis of which it approved the use for statistical purposes of data taken from administrative files. But this concept nevertheless remains of limited application, since it extends only to internal statistics produced as a by-product or from surveys conducted by bodies with the same aims as the service managing the processing of source data.

With the exception of the possibilities offered by the concept of "broader end use", access by researchers and statisticians to files, and especially administrative files, is limited by the principle of end use, and especially the older principle of professional secrecy.

4. MOVES TOWARDS ADAPTING DATA PROTECTION LAW TO RESEARCH AND STATISTICAL NEEDS

The work of researchers and statisticians raises two sets of ethical questions.

- The first relates to the end use of statistical work, which may prove incompatible with the spirit of the research work and pose a threat both to individuals and to the general interest. This is the case when statistics are falsified in order to manipulate public opinion. It also applies when personal data collected or prepared in the course of a research project are used as a basis for decisions relating to individuals prejudicial to the individuals concerned.
- The second set of questions concerns the methods used, both as regards the facilities provided for statisticians for gaining access to information and particularly the guarantees of anonymity given to research subjects.

But here, as we have seen, French law provides only incomplete answers. Given the substantial growth in statistical needs, it is understandable that there are those who ask whether it would not now be appropriate to update the Law of 7 June 1951 (see above), or even to prepare special legislation governing research and statistics. In return for certain new possibilities granted to them, such legislation should require researchers to observe a common code of ethics whatever the field in question and whatever their personal status might be, since one of the major difficulties in regulating statistics is the extremely varied nature of the research and statistical field.

The new regulatory instrument, if it materializes, should at all events achieve a balance between the interests of research and statistics and the preservation of individual liberties, including the respect of privacy.

Regarding the first set of questions:

Under certain conditions it might be appropriate to provide researchers with access to data held by public administrations. This would require a change in current legislation on professional secrecy. The CNIL, which is studying the problems posed by the creation of statistical records on cancer by bringing together information from the most diverse sources (analysis laboratories, hospitals, anti-cancer centres etc.), has virtually arrived at the conclusion that these databanks will be legal only if there is a change in the secrecy regulations. Should this change be restricted to the medical field? Should it be extended to all public statistics by, for example, granting the INSEE a general right of access to administrative data subject to the conclusion of protocols between the INSEE and the public administrations in possession of the information?

Should one not also seek to introduce a certain flexibility as regards the collection of personal data, at least in certain circumstances? Cancer records are a good example of instances in which it is difficult to tell those being researched, who often have little or no knowledge of the reality of the illness from which they are suffering, of what is behind the questions they are being asked.

It should also be possible to make it easier to use the general population census for obtaining samples. Under current legislation, only the INSEE has this possibility and it is forbidden by the statistical secrecy principle to pass on the information to any other body conducting surveys.

As regards the second set of questions:

It would appear advisable to accept the principle of a "functional separation" of data managers and researchers set out in a Recommendation of the Council of Europe of April 1982. According to this principle, "personal data collected for research must not be used for other purposes. In particular, they must not be employed for taking decisions or measures directly affecting the person concerned, except

in the context of the research project or with the consent of the person concerned". This principle is, in effect, aimed as much at protecting the persons being researched as at safeguarding researchers and statisticians from pressure and suspicion.

The penalties for the violation of the privacy of respondents have already been laid down in various penal statutes. Article 378 of the Penal Code, invoked by the 1951 Law on statistical secrecy, provides for penalties of up to FF 8 000 and imprisonment for periods of up to 6 months. The law on automatic data processing and civil liberties also contains numerous provisions for the imposition of severe penalties, notably for the fraudulent collection of data, failure to take security measures to preserve the integrity and confidentiality of data (Article 42), the disclosure of name-linked information, even as a result of imprudence or negligence, when such disclosure is prejudicial to the reputation or esteem of the person concerned or constitutes a violation of that person's privacy (Article 43), as well as a change in the end use of the data processed (Article 44).

These penalties could equally well be applied to research and statistics. It is in order to make researchers more accountable that the CNIL ensures that all processing of statistical data is declared to it, not only by the requesting body but also by the head of the research project. This principle of joint responsibility could be incorporated into law.

The CNIL has a decisive influence on shaping the security measures to be taken to guarantee the anonymity of research subjects. One must also recognize that professional secrecy, which, in particular, forbids the exchange of information between administrations, even for statistical purposes, provides ample justification for the introduction of cryptography or depersonalization techniques such as the "double blind" method. In the event of a relaxation of the rules governing the access of statisticians to information, it would be absolutely essential to tighten the security regulations. The CNIL is therefore working on a standard security regulation within the framework of its regulatory

powers. This standard regulation could be supplemented by a special regulation applying to the processing of statistical data.

5. CONCLUSION

Whatever their justification, whatever their - for the most part - disinterested aims, there can be no question of granting research and statistics any immunity whatever as regards the protection of personal data. A legal framework, supplemented by guidelines on public statistical programmes drawn up by the National Council on Statistical Information and by CNIL guidelines on automatic data processing, should make it possible to satisfy the social need for more detailed statistical information, while at the same time respecting the rights of the individual to a "human identity", to quote the phrase used in the first Article of the Law of 6 January 1978 on automatic data processing, data files and civil liberties.

*

* *

DISCUSSION

PROFESSOR KEIDING (*University of Copenhagen*) opened the discussion:

Since he worked as a biostatistician with social science applications, he could look at the question of data protection from the outside. He doubted whether the general public was aware of the distinction between administrative and statistical uses of the many registers now being compiled, and strongly supported Mme. Lenoir's recommendation that the principle of "functional separation" should be strengthened. Nowadays statistical data could no longer be defined simply as tables since the individual record was the basic statistical item of data, particularly where longitudinal records were concerned. The consequences of incorrect data could therefore be very serious since the effects of even slight errors might accumulate. The measures necessary to ensure correctness were rather close in

substance to administrative uses of the registers which made functional separation more difficult to accomplish for the authorities and less easy for the public to believe. He saw no simple solution to these problems.

The data protection agencies represented another aspect of the question of forming public opinion. Professor Keiding asked for some further precision as to how current aspects of public opinion were represented in the CNIL and how the CNIL ensured that the public was aware of and appreciated its work.

Mr. JOHNSTON (*United Nations Statistical Office*) posed a number of questions concerning the mandate of the CNIL:

(a) Was this confined to data processing within government departments or did it extend to data on individuals collected in the private sector, e.g. banks, insurance companies, credit-card companies? (b) was it concerned with data collection plans, data processing, the compilation of statistics and the dissemination of the data? He noted that the concept of "final use of data" was emphasised in Mme. Lenoir's paper and asked whether this was important for the CNIL's work in the private as well as in the public sector.

The author replied to the discussion:

She gave further information about the Commission Nationale de l'Informatique et des Libertés. This was an independent administrative authority. It had seventeen members drawn from the legislature, the judiciary, the executive and the "forces vives" of the nation represented by the Economic and Social Committee. The Commission was concerned with all data-registers whether public or private.

Anyone wishing to set up a register must make a declaration to the Commission giving full details of its nature and purpose and must receive the authorization of the Commission. The functioning of the register was then supervised by the Commission.

Public opinion was not represented on the Commission as such, nor could it be defined for this purpose. However, the work of the Commission received wide publicity. An annual

report of the deliberations and decisions was published and was given to the press. Important cases were widely reported in the press and might be the subject of press conferences. Decisions of the Commission were available to the public in municipal offices, while the general public had access to the register of registers maintained by the Commission. No special dispensations or derogations applied to statistics: if a statistician wished to extract data from a file, his activities concerned the Commission in the same way as those of any other user.

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. In the system, these three types of objects are linked by means of the unique identifiers for each subject: person number, the exact address of a dwelling and a single code for the place of employment.

Each object has a number of characteristics, such as age, employment, and number of rooms in the dwelling, which can constitute useful statistical information regarding the situation of the individuals concerned.

The aim of the system is to create a statistical description of the person's social and demographic situation and of changes in this situation. The description is intended to enable many types of statistical analysis to be made. There will often be a need for analyses which cut across traditional statistical branches and which require data from many different types of register.

The basic information on the three objects mentioned (persons, dwellings and places of employment) and on the relationship between them is contained in a number of independent statistical registers. Each of these registers is designed to be used for a single statistical area (e.g. education statistics) and contains the data required for this purpose. At present there are 35 statistical registers. Two or more registers are combined to extract anonymous statistical data only if this is required for an actual survey.

Most of the registers are updated once a year although some are updated more frequently. The updating work normally consists of a statistical revision of extracts from one or more administrative registers. The basic information is compared, checked and amended so as to obtain the best possible overall view of the characteristics which form part of these statistics. The sources for a statistical register often consist of many administrative registers because the combinations of data which occur in the administrative registers are not relevant to those who use the statistics.

Since the coordination of data using person numbers is an essential part of the updating procedure, information extracted from the source registers must contain personal identifiers.

The time lag between the date on which the data are collected and the updating of the various registers ranges from two months to one year, depending on the updating procedures used for the source registers.

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.

2.2. Dissemination of statistics

Both the central statistical offices and users of statistics have good reasons for wishing statistics to be presented in a form which makes them immediately available to users for further processing and analysis.

However, it is difficult or impossible for the central office to find the resources to study user's problems in sufficient depth or to assist directly with any analytical work which users would like to carry out.

On the other hand, the central office has an obligation to provide users with the best possible conditions for using the basic data collected; this is especially important in a country like Denmark where statistics are produced centrally and where, in many areas, users have little possibility of themselves collecting basic information of the same quality as that provided by the central office.

For the major users of statistics, publications are not sufficient. From them it is important to have access to analytical facilities which can test new methods or give answers to new questions in a rapid and flexible manner. It would be inconvenient if this work required communicating with third parties.

Many users might therefore wish to receive copies of or extracts from Danmarks Statistik's registers. But, as mentioned earlier, there is a well-established tradition at Danmarks Statistik, which is now backed up by legal authority, that personal data are not passed on.

The central office must find solutions to this problem which satisfy users and which also enable it to maintain its central role in the production of statistics and its access to information from registers. One technical solution is to supply machine-readable information, usually on magnetic tape, which must have a high level of breakdown so that it can be used in a flexible way. This type of information would not be suitable for publication. A borderline case is the system which has become increasingly popular in several countries, including the USA, namely the supply of "model data", where each record deals with one of the "objects" covered by the statistics (a person, family, undertaking, etc.). The objects, of course, must not be identified.

Supplying very detailed information may raise a number of problems which have to be considered before deciding whether this form of statistical service should be provided. The danger that the statistics might reveal information about individuals or undertakings must be taken into account in

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 of 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.

STATISTICS AND PRIVACY - THE OFFICIAL PRODUCTION
OF STATISTICS AT A CROSSROADS

EDMUND RAPAPORT, POL.MAG., JUR.KAND.
Statistics Sweden

SUMMARY

From about the beginning of the 1970s those concerned with the production of statistics have been grappling with certain problems relating to the use of primary data. This problem is derived fundamentally from the apprehensions of the general public and of legislators about the detrimental effects of the collection and storage of data for those whom the latter concern. In this paper, this and connected problems are described and discussed on the basis of certain experience that has been gained, chiefly in Sweden. Questions concerning legal safeguards for confidentiality and privacy, data security and record linkage in the production of statistics are taken up. Finally, certain hypotheses relating to the situation concerning the production of statistics are put forward in the light of current opinion trends and also some thoughts of relevance to the latter regarding a desirable line of policy.

0. INTRODUCTION

The well-known problem with which the official statistical agencies are grappling today can be summarized briefly in this way. They now find themselves in a field of force whose opposite poles consist, on the one hand, of an intensified and more clearly articulated demand for statistical information, a markedly improved production potential - chiefly due to automatic processing techniques but also due to advances in methodology - and reduced financial resources. On the other hand, they are opposed by the apprehension felt by the general public about statistical activities in respect of the preservation of confidentiality and personal privacy. The first three factors - demand, techniques and methods, finance - can be dealt with by rational and logical methods. The fourth factor - fears about imperfect confidentiality safeguards and encroachment on the personal sphere - can only to some degree be met in the same rational manner. Such problems are elusive, are largely to be found at the psychological level and, despite the volume of material which has by now been written in this field, must be said to be still insufficiently investigated

and, to an even lesser extent, to be solved. These fears, also standing diametrically opposed to the three other factors, constitute the issue that gives rise to the tension in this field.

It is probably no exaggeration to say that in this situation the official producers of statistics find themselves at a crossroads; before choosing a path to follow, they need to obtain a better understanding of the pre-conditions for various lines of action, they must obtain a better perception of the effects of different measures and projects - in short to obtain a better foundation on which to base their policy. This does not mean that today they are unaware of this discussion and are unprepared. Quite the contrary, people are acutely aware of this and related problems, and this discussion is continuing both internally, and between statisticians and non-statisticians. In particular, the serious difficulties encountered by the official statistical agencies in various countries have contributed to this state of affairs. However, it seems essential that this discussion should be made more profound, and should not abate, so that we can have guidance about the attitude to be taken in these matters and about what is to be done in the future.

The intention of this paper is to contribute to the underlying basis of the discussion at the EUROSTAT seminar from the viewpoint of certain developments and certain experience gained mainly in Sweden of the aspects in question. I shall be chiefly focusing attention on questions relating to 1) legal safeguards for confidentiality and privacy, 2) data security and 3) problems relating to record linkage.

1. DEVELOPMENTS IN SWEDEN

Sweden, the first country in the world to do so, passed a statute (the Data Act) in 1974 which regulates the automatic processing of personal data. A special authority, the Data Inspection Board both issues the licenses or permits needed before an automated register containing individual data can be set up, and ensures that the Act is complied with. No exception has been made in the Act for data for statistical purposes: on the other hand, in the explanatory introduction to the Act, it is recommended that special regard should be paid to the needs of statistics and research. Since practically all the compilation of statistics at Statistics Sweden - our official statistical agency - is computerized, this Act and its accompanying regulations mean that a permit is required for most of the statistics produced - for the setting up of new files, the linkage of files, and for the supplementing of files with the inclusion of fresh data, etc. Unlike what is usual in the data legislation of other countries, it is laid down in the Swedish legislation

that extracts from statistical registers must be supplied if requested by those to whom the data relate. In accordance with the provisions of this legislation, in the 1983/84 fiscal year some 24 000 people asked Statistics Sweden to supply them with some 145 000 register extracts.

The legal safeguard for primary data for statistics is mainly formulated in a clause in our special Secrecy Act; the protection given is regarded as being substantial. Statistics Sweden has however called for certain improvements in this safeguard and we have good prospects of having our request met. Our most immediate requirement is a better safeguard against the risk of being forced to supply particulars to a court or some other public authority. I must emphasize here that our call for increased protection has not been put forward because we have had any experience in practice of such a case. On the contrary, confidentiality safeguards have functioned satisfactorily in Sweden - and this also seems to be the view of statistical agencies in other countries.

Since 1947 all individuals in Sweden have been allotted a personal identity number which is unique for them (it consists of a person's date of birth and of an additional four digits. Nowadays all undertakings and other juridical persons also have a unique registration number. Naturally this system is an excellent aid for identification in contexts involving automatic processing technology. It is also used by Statistics Sweden throughout in its processing of statistics and is today regarded as an absolutely necessary instrument in its work. Were the personal identity number system to be abolished, the result, for Statistics Sweden, would be the widespread disruption of the work carried out. I must also mention that since the 1960s Statistics Sweden has been made responsible for about 4/5 of the whole of Sweden's extensive statistical activities and that the agency keeps a large number of registers, including two total ones - one of the population and one of enterprises with several institutional units.

As I mentioned above, special permission must be obtained from the Data Inspection Board - or from the Government - before there is any linkage of records in different files. In 1979 the Government commissioned Statistics Sweden to draw up and present proposals for a new Census of Population and Housing in 1985 which was to be based entirely on the linkage of existing registers at Statistics Sweden and other authorities; the general public were to be relieved completely of their obligation to fill in and send in household returns forms, as they had been required to do previously. In January 1983 Statistics Sweden presented proposals in line with the Government's requirements. Our agency proposed, nevertheless, that some particulars -

though very few - should be collected from the general public since the linkage of existing registers alone would not prove sufficient. Six months after the presentation of our proposals and their circulation to some 80 authorities and organizations for their comments, a major daily newspaper "discovered" our proposals and raised the alarm. Our proposals were claimed to be a threat to the privacy of the individual. This led to a widespread debate, its consequence being that the Government ultimately rejected our proposals and instead submitted to Parliament a statute proposing a "conventional" Census of Population and Housing in which particulars would be collected from the general public on forms. Notwithstanding, this new statute is also based on a considerable interlinkage of existing registers; this method has in fact been increasingly used in censuses of population ever since the 1960 Census. However, this statute was not passed unanimously by Parliament either, which had been the case with all previous censuses of population. Although the measure received the support of a majority in Parliament, all the political parties to the right of the Social Democrats voted against it.

In the past the scope of data collection for statistical purposes and linkage issues have from time to time given rise to a public debate. Especially noteworthy was the controversy with the Data Inspection Board, in particular, during the middle of the 1970s, in regard to the introductory interview survey in the statistical series called "Living Conditions in Sweden". The data to be collected were held to be too comprehensive and too sensitive.

A question that has not been solved is the attitude to the preservation of identifiable statistical units for future research purposes. There are two conflicting opinions; on the one hand, it is held that the safeguarding of personal interests ought to necessitate the anonymising of primary data or their complete destruction and, on the other, there is the view that identifiable units, just as the material as a whole, should be preserved at any rate in the case of material regarded as valuable from the viewpoint of research. In Sweden, above all because we have our system with personal identity numbers, it has been considered that the prospects are good for the carrying out of longitudinal research and of research that is based on record linkage;

this method has therefore been widely used. But in Sweden, too, as in other Western countries, this remains a matter for discussion: the opinion that individual data are the property of the respondent is gaining increasing ground as also are the apprehensions about the storage of individual units. The Data Act, such as it is interpreted and applied at the present time, is based on the view that registers containing individual units, after being used for the purpose for which they were intended, should be anonymised or destroyed.

2. LEGAL DATA SAFEGUARDS FOR STATISTICAL UNITS

In all countries with established arrangements for the production of statistics, it has been considered for a long time that statutory safeguards for primary statistical units are a necessity. In brief, the arguments for these safeguards are these. Seen from the viewpoint of the respondent, there is the requirement that individualized information, which is often detailed and sensitive, should be protected from access by unauthorized persons and from being used for purposes other than those intended; from the viewpoint of the producer of statistics, regard must be paid to the endeavour to avoid non-response in respect of data and create trust in reliable particulars. In addition, it is usually maintained that the use of the statistics results in compilations that do not reveal who is the individual respondent.

The techniques vary in the prescribing of legal confidentiality safeguards. Variations also occur between and within different countries in the level of safeguards that statistical activities are assured. In the latter respect, special attention has been focused on the conceivable risk that the body responsible for the statistics may be forced to supply individual particulars to a court of law for specific administrative purposes. No example of this happening in reality seems to be available. However such a contingency has in itself been regarded as being a threat, in respect of public opinion, to statistical production.

In the 1970s and the 1980s an increasing number of statutes regulating data activities have been passed. These statutes also affect, to a varying degree, the compilation of statistics. Above all, attention should be focused especially in this context on the demand for the control of statistical activities, too, in view of possible risk for individuals (or juridical bodies), the demand for information on existing registration and its content and the demand for the storage for a limited time of the original particulars.

There often seems to be general agreement that this new type of legislation has had a detrimental effect on statistical production. The objections raised against this legislation, just as in Sweden, are usually that statutory registrations hamper the rational carrying out of the work and that the restrictions act as a restraint on the free production of statistical information for the public debate and for research. However, favourable aspects may possibly be ignored; the data legislation - in the changed atmosphere in society - may also help to improve the working climate in which statistics are compiled by calming people's fears about adverse effects.

3. TECHNICAL SECURITY FOR STATISTICAL DATA

The nature of the conventional protection for the facilities devoted to statistical production varies considerably according to local conditions, such as the nature of the building, external and internal communication lines and similar factors. The statistical agencies have long experience and practice in the framing of this security system. This also includes the protection of the premises used for computer centres. Modern technology today offers various ways of providing this protection. As far as I know, these security arrangements have functioned faultlessly everywhere.

A problem that is harder to overcome is that of protecting the transmission of data in an automatic processing system. In the statistical field these difficulties are mainly to be found in the electronic transmission of data between premises that are geographically at a distance in the production facilities. The method of security which is generally used is enciphering or transformation. Even though it is not entirely impossible to overcome security devices of this kind, the level of security is usually sufficient in practice.

A problem which is much more troublesome seems to be the protection of confidential information in a data system to which there is external connection in the form of direct access to the sections of the statistical information that are public. Statistical data bases with terminal connections for users are already in use, or are under development, in several countries. The apprehensions expressed by experts in the field are generally framed in the form of a statement to the effect that it is not possible to safeguard confidentiality if both public and confidential units are stored in the same computer. By manipulations from the terminal, skilled programmers can break in to the part of the information store which is secret. It has been

maintained that the only way to create security protection against this form of unauthorized access is to have separate computers for public and for confidential information.

I should perhaps mention here one method that aims at creating further protection for primary statistical units which is at present being studied at Statistics Sweden. This method requires that all general identifiable data which are processed or stored at Statistics Sweden should be enciphered.

A security problem in statistics which is well known is the danger that, in certain circumstances, the tabulations are of such a nature that it is possible for the individual units that are behind the tables to be disclosed. This phenomenon generally appears in cases of a distorted distribution of the population recorded or when particulars are recorded at a low geographic level. There are various rules of thumb and various methods whereby disclosure can be guarded against. In recent years rather advanced dataprocessing techniques have been evolved to enable checks of disclosure hazards to be made and protection against disclosure to be provided. Absolutely watertight protection can rarely be achieved, but nowadays there seems to be a general consensus that the protection is adequate if the costs of eliminating it seem unreasonably high in comparison with the information gain.

The anonymization of the primary statistical units can also be regarded as a technical method for the safeguarding of confidentiality. In itself it is a simple technical method to carry out. In particular it must be borne in mind that anonymization is not a reliable security method because - depending on the size of the population or the sample, the existence of extreme values, access to other register data etc - there is a danger that at least some of the units in the population can be re-identified. I shall be referring in the next section (record linkage) to special disadvantages from the statistical viewpoint of anonymization.

4. RECORD LINKAGE

Record linkage as a valuable statistical tool is well known nowadays and well established. This method can be used to expand the variable content, to improve statistical quality and to check on accuracy. It is a cheap method both in regard to the resources for statistical production and to the burden put on respondents. Frequently, especially in longitudinal studies, record linkage is the only way by which the required information can be extracted.

The obstacles and difficulties involved in regard to statistical techniques are numerous and often substantial. Where the actual linking of statistical items is concerned, there are, in principle, two possible procedures. One procedure - practically the only one used in Sweden - is to use a unique concept for identification. The main difficulties encountered here are the defects and errors in the identification concept which can give rise to "a miss". The other procedure involves matching with the aid of some other non-unique information e.g. name, address, variable values. The difficulties here are naturally far greater but a large number of statistical works devoted to this type of matching method are now being written. Among the special problems connected with record linkage to which I wish to draw particular attention are the discrepancies between the target population and the register population, dissimilarities between definitions, lack of agreement between the aim and the available worksheet data and similar differences.

Where the record-linkage method is concerned, there is one special and important difficulty - that is the opposition from large groups of respondents, from political parties and from organizations. The opposition takes the form of both statutory regulations that govern record linkage, and of reactions from public opinion. The adverse effects for statistical work are well known both in Sweden and elsewhere.

It is hard to analyse and discuss rationally and in greater detail the apprehensions expressed about statistical record linkage. From the debate one can discern distrust of record-linked statistical units being used for what are called "administrative purposes" - even though the statisticians concerned have given assurances to the contrary - and that intentional or unintentional leakage from statistical data compilations may arise despite assurances by statisticians of high-level technical safeguards and the non-existence of examples of such leakage. The fear has also been expressed that statistical material might prove a hazard for private individuals in the event of war or a *coup d'état*. However, one may well ask whether the opposition is not mainly derived from a profound emotional aversion to there being somewhere - that is also in statistical archives - all too much information about the individual, about the household, about the enterprise. This is indicated by, for instance, the phrase that is rather frequently used: record linkage creates a total and revealing picture of the individual, even though every

informed and reflecting person must see that it must be absurd to think of record linkage as giving, in the psychological or literary sense, anything but a fraction of the true picture of a human being.

If it is to be possible to carry out longitudinal studies, it is of course vital for the data material that is needed to be preserved. A pre-condition that is equally self-evident in countries where there are unique identification concepts is, too, that these concepts must not be abolished. However, both in Sweden and at the international level, there is a tendency today for rules and attitudes to be characterized by the reverse: statistical data compilations must be destroyed or at least anonymized after being used for the primary purpose for which they were intended. This new attitude which has emerged in public opinion and among legislators was scarcely known previously. The sorting of data is now mentioned as the principal method whereby privacy can be safeguarded, for instance in the Council of Europe's Convention for Protection of Individuals with Regard to Automatic Processing of Personal Data (1980). The Council of Europe's Recommendation on the Protection of Personal Data Used for Statistics or Scientific Research (1983), the OECD's "Guidelines Governing the Protection of Privacy, etc." (1980) and the European Science Foundation's "Statement concerning the Protection of Privacy and the Use of Personal Data for Research" giving anonymization as the security method but opening the door to the possibility of complete preservation to meet the needs of future research. On the other hand, nothing is said about future research needs in the very detailed ruling concerning statistical issues by the German Federal Constitutional Court in December 1983 in regard to the Census of Population Act: in this ruling anonymization was indicated as a necessary security method. As I mentioned above, in Sweden the life of each automated register containing personal data is now limited in time also in the case of statistical data files with anonymization or destruction as the security method.

5. SOME PARTICULAR ISSUES

In this concluding section I shall put forward as a basis for discussion certain hypotheses and certain questions ensuing from them. Here I feel it is important to underline my declared aim - these are topics for discussion which is why I shall formulate them so incisively.

a. In their attempts to calm the fears of the general public, statistical agencies have principally concentrated on assurances of a high degree of confidentiality for primary statistical units. They have not been particularly successful. One may well wonder whether our efforts should not be focused more on showing the value and importance of statistics in easily read, convincing and concrete examples from various fields such as communications, education and medical research. Of course it is important not to neglect confidentiality safeguards and data security and to inform about them, especially as an "accident" could do enormous damage to public opinion - but the general willingness to take part, and to let statistical production function undisturbed, depends more on the general public's assessment of its positive value.

b. Those who produce statistics have not become sufficiently involved in questions concerning legislation pertinent to statistical work and have not actively sought to influence what happens in that respect. The consequence has been that insufficient regard has been paid to statistical needs in Acts of Parliament and other statutes. There are probably grounds for paying, in the future, more regard to questions relating to legislation in order to prevent legislation which impedes the rational production of meaningful statistics.

c. It appears that there has been an important change in the general attitude to data collection and data storage in the Western world. This is expressed more plainly in the above-mentioned ruling by the Federal Republic of Germany's Federal Constitutional Court in regard to the proposed 1983 Census of Population: the Court laid down the right to "informational self-determination" - a new principle and right for the person to whom the data relate, a kind of new or extended right of ownership in regard to the information about oneself. For the producers of statistics this point of view should enhance the need to open and strengthen the dialogue with the respondents in order to convince them - on the basis of a recognition of the changed attitude - that statistics are of use to society. To put it another way: the change in the manner of making contact with public opinion is said to mean a shift from "deciding" to "negotiating". What I have said here also has a bearing on the use of administrative data in statistical production.

d. The trend towards the destruction, entirely or partly (anonymization), of primary data is a hazard for society. Such destruction makes any broad and prolonged statistical studies difficult or impossible. Inherent in this trend is an unhistorical view of the development of society, one which does not pay regard to coming generations. Certainly one may well ask where historical research would stand today if this view had been the one subscribed to by our forefathers. Very little of the discussion of these questions has been seen up to now. Nor have the voices of statisticians been heard on this question to any appreciable extent. Should we not act?

*

* *

DISCUSSION

MR. BARNES (*Office of Population Censuses and Surveys, London*) opened the discussion:

He drew attention to the dissimilarities and similarities between the situations in Sweden and the United Kingdom. The changes in public opinion in regard to the census of population were less marked in the UK than in many other countries, perhaps because the 1981 census was less controversial than elsewhere or perhaps because most inquiries for social data in the UK were on a voluntary basis. A similarity with Sweden was the irrationality behind the fears of supplying information which should not be countered by solely rational argument. It was most important that the case for the census be believed and it was insufficient to speak of the harmlessness of the census and of the guarantees of secrecy. A public-relations and not simply a publicity approach was needed.

Statisticians needed to keep in touch with draft legislation to ensure that no legislation was enacted which might give the courts the right to demand, as evidence, information given under a guarantee of statistical secrecy.

As in Sweden, there was considerable pressure in the UK from users for better access to statistical data bases. However, public-use tapes were not provided from the census, being prohibited under census legislation. He agreed on the value of data protection legislation in calming people's fears in furnishing statistics. Mr. Barnes supported the view in the paper that positive information on the use of statistical data was of greater value in securing public cooperation than arguments seeking to calm fears about disclosure. He agreed also that statisticians should make their voice heard when data protection legislation was being drafted. The principle of "information self-determination" which emerged in the Federal Republic of Germany had not, as yet, made its appearance in the United Kingdom. With regard to the stocking of statistical data for historical studies, the position in the UK was that census records remained confidential for 100 years.

Mr. EMBLETON (*Central Statistics Office*) enquired:

When it was envisaged that the 1985 Swedish census might be taken on a record-linkage basis, were any soundings of public opinion made? Similarly, when it was decided to revert to the traditional style census, were any soundings of public opinion made?

Dr. BURGIN (*Statistisches Bundesamt*) pointed out:

Most of the inquiries conducted by the Statistisches Bundesamt were multi-purpose and had many users in government, science and the economy. It would be impossible to inform the respondents of all the uses. What was needed was for the administration and politicians to emphasize the essential nature of statistical data. As regards statisticians influencing discussion of relevant legislation, he recalled that the president of the Statistisches Bundesamt had protested, without success, against precisely those features of the 1983 Census Act concerning the passing on of data which were subsequently judged unconstitutional. There was no possibility of

retaining statistical records relating to identifiable individuals for historical purposes. As things stood, such records must have personal identification removed as soon as possible.

Dr. DAMMAN (Bundesbeauftragte für Datenschutz) stated:

He did not agree that emphasizing the value of a survey was more important than emphasizing the confidentiality aspects: both were equally important and much of the failure of the 1983 Census could be attributed to a neglect of both aspects. As regards different climates of opinion in the different countries mentioned by Mr. Barnes, these could change very quickly - witness the contrast between the Federal Republic and the Netherlands in 1970 and the similarity in the 1980s. Resentment against a number of unfavourable factors, not connected with statistics, were projected by the public on the census of population. Census preparations coincided with the electoral campaign, there was tension between the public and the administration, fear by the public of computers, awareness of the computer related problems of unemployment and deteriorating employment conditions.

Mr. REDFERN noted:

It was difficult to give the general public convincing reasons for the taking of a census. The value of the census in calculating "rate support grants" was much emphasized in the UK but this did not seem to him to be a very convincing reason.

Mr. HARRIS remarked:

He agreed with Dr. Damman that, viewed from outside Germany, the 1983 census of population project had not been sufficiently presented or defended. He suspected also that there was a "domino effect" involved, starting with the Netherlands, passing next to Germany and now showing signs of appearing in Sweden.

Mr. JOWELL (Social and Community Planning Research) was of the opinion:

He thought that there was little real evidence for the thesis that the public was growing tired of the burden of answering surveyors. This thesis should not, therefore, be

produced as a justification for replacing traditional survey methods by record linkage.

The author replied to the discussion:

He was surprised at the large measure of agreement for the views in his paper which he had thought were quite controversial. On the question of the changes in 1985 census plans, the original proposal for a record-linked census had had government approval and Statistics Sweden had been instructed to draw up a detailed scheme. The questions of the effects on privacy emerged suddenly and was taken up by newspapers with the result that plans were modified and it was decided to revert to the traditional style census. A disquieting feature of the 1985 census was that the census law was not accepted unanimously in Parliament but only with a majority. He felt that there was a big public relations problem here. Statisticians were not normally very skilled in public relations and it would be useful if experts in this field could be employed.

CONFIDENTIALITY AND PRIVACY
CURRENT SITUATION AND PRACTICE IN IRELAND

F.A. EMBLETON*
Central Statistics Office, Dublin

CENTRALIZED STATISTICAL SYSTEM

Through the Central Statistics Office, Ireland operates what is primarily a centralized statistical system. When the Office was set up in 1949 it was placed under the aegis of the Department of the Taoiseach (i.e. Prime Minister) primarily to underline its independence and central role in collecting and coordinating statistics. As in other countries, of course, many statistics result from the activities of other Departments of the State as a by-product of the administration of various Government programmes and schemes. However, the Central Statistics Office alone conducts all the official nation-wide statistical inquiries.

Confidentiality

Fundamental to the standing and public image of the Office are the guarantees of confidentiality which the Office extends to the suppliers of information. Confidentiality from the viewpoint of the supplier is the set of constraints and conditions governing the use and disclosure of information while it is being collected and once it is collected. "Confidentiality" however is extremely difficult to define in a form which will be universally unambiguous to the suppliers of the basic information, to the collecting agency which becomes the custodian of the information, to the legal authority which may be called on to interpret the guarantees and to potential users. In Ireland, the guarantees are supplier-orientated and enshrined in basic legislation - the Statistics Act 1926.

Statistics Act 1926

The act covers the collection, compilation and publication of statistics. It empowers the relevant Minister of Government, which since 1949 is the Prime Minister, to specify by statutory order the subject matter and respondent group for statistical inquiries. Under the Act the Prime Minister "may collect, compile, abstract and (subject to the provisions of this Act) publish statistics relating to any matter affecting the general economic and other activities and conditions" in the State. There is also provision for other Government Departments, should it be necessary, to

* The views expressed are personal and should not be taken as representing or reflecting the position or views of the Central Statistics Office.

collect statistics under the Act but subject to prior consultation with and coordination by the Prime Minister.

There are penalties laid down for failure to supply information formally requisitioned under the Act and, equally on the other hand, for revealing individual return without the consent of the supplier. Happily nobody has ever had to be prosecuted for disclosing confidential information collected as part of statistical inquiry. While a number of inquiries are carried out by formal requisition under the Act, for example the annual Census of Industrial Production and the periodical Census of Population, there is a large volume of information collected by means of voluntary inquiries, that is without recourse to a statutory order. However, equal guarantees of confidentiality are given by the Office to respondents to such voluntary inquiries. In other words, these inquiries are also conducted as if fully statutory in nature.

In the case of voluntary inquiries, there is in any event the protection afforded by the "Official Secrets Act 1963". This Act specifically prohibits a civil servant from communicating official information, which includes information collected for statistical purposes, to any person unless he or she is duly authorized to do so in the course of and in accordance with his/her duties. All Central Statistics Office staff, on appointment, sign a notice which sets out the principal provisions of the Act.

Specific Confidentiality Provisions of the Statistics Act

Returning to the 1926 Statistics Act, the particular provisions relating to disclosure and use of statistical information which, as already stated are applied even if the inquiry is a voluntary one, are set down in Article 13.

Article 13 stipulates that:

- (i) no individual schedule, form, questionnaire or other document or any part thereof, verbal information or answer relating to any individual person, business or concern shall be disclosed without the consent of such person, business or concern;
- (ii) so far as is reasonably practicable, no report, abstract, summary or other publication shall contain information so arranged as to enable the identification of particulars relating to any individual person, business or concern without the written consent of that person, business or concern.

These two conditions extend also to records or documents which are not open to public inspection but which have been

inspected, examined or copied for statistical purposes by an Office of Statistics - the latter being defined also by the Act.

In practice, Article 13 imposes constraints on the detail which may be published or otherwise made available. Where appropriate, rules are applied to avoid disclosure of information which may be attributable to an individual responding unit. For example in the case of business inquiries, no information is generally published where there are less than 3 reporting units. Where there are at least 3 but 1 or 2 are dominant, figures are also not published. However where the information is considered to be important and of general use, a pragmatic approach is adopted. In such instances, the Office seeks the consent of the unit or units involved, as provided for in Article 13, to publish the information. This approach has met with success on many occasions; of course if no consent, then no publication.

Great care must also be exercised to avoid the inadvertent disclosure of individual or sensitive information by way of detailed cross-classifications or provisions of detailed small area statistics. In recent years there have been growing demands from planners, researchers and other users for more detailed information with concomitant risks of disclosure by isolating certain characteristics with readily identifiable qualities or through linkages with other known details. Modern data processing techniques have greatly promoted the ease with which such information may be made available. Every effort is made to avoid possible disclosures or perceived breaches of confidentiality and where necessary steps are taken to suppress or otherwise disguise the offending figures.

The above are examples of the so-called "active" approach to indirect disclosure. We do not apply this rigorously in publishing Foreign Trade Statistics where the "passive" approach is adopted and where if a complaint is made, we attempt to rectify the position.

One further aspect merits mention. We do not take into account any additional sources of information which users may have at their disposal and which, in conjunction with our published data can unlock the puzzle.

Restrictions on Access

Needless to say the constraints imposed by the guarantees of confidentiality extended to respondents restrict access to statistics and are not therefore universally appreciated. Official collection agencies consider it essential that the suppliers of the basic data are fully protected by

law - indeed we consider that the respondents' interests take precedence over those of all others and in general we would oppose any moves which could be interpreted as weakening existing guarantees. Most of us are justifiably proud of and wish to maintain the traditions of confidentiality which have for so long been a basic principle of our collection mechanisms. However, as indicated already, confidentiality cannot be defined unambiguously to have the same meaning for all. The axiom that one cannot please everybody all of the time is all too apparent and in the modern statistical environment becoming even more so. Dissatisfaction with existing arrangements exists in Ireland, as undoubtedly elsewhere, and questions have been raised with regard to the rigid application and interpretation of these arrangements. Indeed, it has to be admitted that what may be confidential to the supplier, may well not be viewed in a similar light by others, particularly potential users and more particularly if some of the data in question or a very good approximation to it can be obtained from other sources.

"Public-use" Samples

An example of the type of access which gives rise to differing views of confidentiality pertains to the provision of so-called "public-use" samples. These are anonymous samples, generally of Census of Population returns, made available to users in some, but by no means all, countries for the conduct of their own private analyses and compilation of tables. The availability of such a sample of the Irish Census of Population returns has been raised on numerous occasions. However the legal authority of the State has confirmed again recently that such samples cannot be provided under the statistics Act, even though the identities of the individuals to whom the records relate are concealed. This decision results from the interpretation that rendering the data anonymous without some form of aggregation (in order to remove the disclosure of any part of an individual return) is in itself not sufficient. To potential users of "public-use" samples this decision may appear to be a somewhat rigid interpretation of the guarantees of confidentiality under the provisions of the Act. The Central Statistics Office, however, is sympathetic to the needs of the users in question and to overcome the legal constraint while fully preserving confidentiality, will be extracting a suitably structured and stratified random sample of about five per cent of the Census returns - this sample will be held on magnetic tape and will be used to produce the special tabulations and analyses required by users either by running the users' own programs or using appropriate packages. In all cases the output will be vetted to ensure the maintenance of full confidentiality.

Public access/inspection

There is another dimension to access which highlights another restriction in the existing legislation. This is the matter of opening to the public individual records after some lengthy period has elapsed - say 100 years. Again the main interest centres on Census of Population records which to future generations can be a rich store of historical, social and cultural information on their ancestors. The Irish Statistics Act under which the Censuses of Population are taken, does not presently cater for such public access in the future - the absence of any provision for it is legally interpreted as excluding any possibility of it. For some, this is a matter of regret. Indeed one can readily appreciate the usefulness of Census records in 100 years from now and the question of accommodating this type of need would appear to merit sympathetic consideration. In so doing, however, existing Census records which may be undergoing processing or are stored on computer tapes (even if names and addresses are omitted) would have to be excluded; that is to say that such records could not retrospectively be made available. To do so would be a serious breach of contract in respect of the existing guarantees of confidentiality given to householders and others making Census returns heretofore.

Mention is made of this and other access restrictions to demonstrate that in Ireland, as elsewhere, there are changing demands, often legitimate and meritorious, which legislation needs to take stock of periodically. This is not put forward to suggest or imply that the present Irish Statistics Act is in anyway deficient - indeed the contrary applies where guarantees of confidentiality are concerned - but to highlight the changed circumstances, principally brought about by modern statistical processing techniques and under which present day data demands are far in excess of those experienced as recently as 20 to 30 years ago. Statistical legislation, no less than other laws, need to keep pace with developments and thus provide the means of, on the one hand, protecting the individual respondent while, on the other, allowing for the maximum use of the information to facilitate full study and analyses of the social and economic conditions of the nation.

Record linkages

A further issue pertaining to confidentiality and to the protection of privacy to which we in Ireland must direct some attention in the not too distant future is the linkage of individual records whether derived from administrative sources or direct statistical inquiries. In the case of administrative sources, the increased use of computers in public administration has resulted in the establishment of data files covering transactions involving the individual

and the State, relating to social transfer payments, tax collection and other activities. The potential use and linkage of these records or statistical purposes, if the required designatory and other basic characteristics are included on the records, have not gone unnoticed although in our case they have not been developed. The statistical benefits to be derived are obviously enormous and given the costs of collecting primary data (i.e. directly from individuals) it is incumbent on Government Agencies, including the Statistical Office, to maximize the use of any available information. This applies equally to the linkage of existing statistical files which may contain diverse information on the same reporting unit.

The approach raises fundamental questions with regard to confidentiality and the right to privacy. Here the right to privacy is interpreted as the "right to know" and, more importantly, the right to limit or control the release and use of information relating to oneself. Privacy of course extends beyond statistical aspects - it applies to information in the broadest sense pertaining to an individual or indeed to groups of individuals. The right to privacy is a particularly sensitive issue and in to-day's society, not to mention in "1984", is one which is likely to excite passions and arouse hostility. Indeed it has been noted that protests about this issue become contagious in that they migrate quickly from one country to another. This of course is greatly accelerated by the present day range of news media which tend to give such protests instant and considerable prominence whether or not they are valid. The public attitude of course is often contradictory in that, on the one hand, the demand is made that public authorities use existing information rather than collect it anew, while on the other hand, objections are voiced if attempts are made to collate the required information from existing sources.

Presently, in Ireland, there is no legislation relating to the privacy aspects as raised here. Neither is there a debate, or what could be described as the beginning of a debate, on the issues in question. The issues however have been commented on and more frequently so in recent times. They are issues to which we must become increasingly alert and, in my view, there is a rapidly emerging need to enact, sooner rather than later, appropriate legislation to protect the privacy of, and provide full confidentiality to, the suppliers of the basic information.

PROTECTION OF PRIVACY, AUTOMATIC DATA PROCESSING
AND PROGRESS IN STATISTICAL DOCUMENTATION
IN THE FEDERAL REPUBLIC OF GERMANY

J. WERNER and E. SUDFELD
Federal Statistical Office, Wiesbaden(*)

SUMMARY

Part 1 of this paper contains a description of the legal situation regarding the confidentiality of statistical information, data protection and the individual's control over the information collected on him. Part 2 examines the effects on Federal statistics of the recognition, contained in the Federal Constitutional Court's judgement regarding the Census of Population Act, that the individual has a right to control the use of information relating to himself. Possible ways and means of meeting the new requirements imposed on statistical surveys in the Federal Republic of Germany are discussed in Part 3.

1. CONFIDENTIALITY OF STATISTICAL INFORMATION, DATA PROTECTION AND THE INDIVIDUAL'S RIGHT TO CONTROL THE USE OF DATA RELATING TO HIMSELF

1.1 Background: the confidentiality of statistical information

1.1.1 The Federal Statistics Act (FeSA)

1.1.1.1 Constitutional proviso and confidentiality of statistical information

Under the Basic Law (Constitution) of the Federal Republic of Germany, any State intervention affecting the fundamental rights of citizens must have a legal basis. This is known as a "Gesetzesvorbehalt", or constitutional proviso. Statistical surveys represent an intervention of this kind and, as such, may be carried out only under a public statute.

The Federal Statistics Act (FeSA) (in German, Das Gesetz über die Statistik für Bundeszwecke) of 3 September 1953, which was the first piece of legislation to codify the law relating to Federal statistics, laid down the framework within which statistics for Federal purposes were to be compiled. With a few exceptions, individual surveys were conducted in accordance with specific laws. These laws also contained provisions regarding the confidentiality of the data collected and thus formed the basis for the provision in the FeSA regarding the duty to observe confidentiality, which applies to all statistics used for federal purposes. Under this provision, details of individual personal or material circumstances collected for the purposes of

(*) Mr. J. Werner wrote Section 1 and 2.1, while Mr. E. Südfeld prepared Sections 2.2. and 3.

compiling Federal statistics were not to be divulged by the persons entitled to receive information unless there was legislation on statistical matters which provided otherwise, such as the 1970 Census of Population Act.

The legislators regarded the abovementioned provision, also referred to as the "statistical secrecy" provision, as a necessary correlate of the general duty, laid down at the same time, to provide information in statistical surveys. If those questioned were obliged by the State to provide true and complete information, they had to be sure that the details which they provided would be kept secret.

Given status as a *lex specialis* the provision regarding the duty to preserve the confidentiality of statistical information took precedence over existing duties to observe secrecy, such as the obligation to observe official secrecy placed on all civil servants. It therefore already met the requirement which is nowadays laid down by commissioners for data protection, everywhere, namely that data protection should, wherever possible, be governed by concrete provisions on specific areas rather than by general laws.

The provision on the confidentiality of statistical information was also a forerunner of later legislation on data protection in other respects. Thus, the legal definition of the property requiring protection given in the Federal Data Protection Act (FDPA)(Bundesdatenschutzgesetz) adopted 24 years later, namely "personal data", is partly derived from the references to the confidentiality of statistical information contained in the FeSA.

1.1.1.2 Meaning of the confidentiality of statistical information

The section in the FeSA regarding the duty to observe confidentiality basically provides that "persons entitled to receive information" should maintain absolute secrecy regarding "individual information on personal or material circumstances" collected for the purpose of compiling federal statistics.

The "individual statistical information" requiring protection includes all the information provided by an individual to enable a set of federal statistics to be compiled. This applies to all data, regardless of whether they relate to natural or legal persons. In addition, information of this kind may not appear in publications. On the other hand, summary information based on data collected from several persons obliged to provide information need not be kept confidential, since it does not relate to individual cases.

The duty to observe confidentiality is placed on "persons entitled to receive information". The latter comprise the persons and agencies officially entrusted with the task of compiling federal statistics. The main agencies and persons

concerned are the Federal Statistical Office and its staff, and - by virtue of the federal structure of the Federal Republic whereby the Länder are in general responsible for enforcing laws - the Statistical Offices in the Länder and their employees.

In order to clarify the extent of the duty to maintain confidentiality and to ensure its observance, the duty laid down in the Revenue Code to disclose information to tax offices is waived in the case of individual statistical information and failure to observe the confidentiality of such information is a punishable offence.

Detailed statistical information may be passed on, exceptionally, to a strictly limited number of persons and agencies, subject to special conditions.

Potential recipients of individual statistical information include the relevant supreme Federal and Land authorities and the agencies designated by them. The bodies in question normally consist of those ministries whose area of work is covered by the statistics concerned and agencies acting on their behalf.

In addition, if one of these potential recipients requests the passing on of individual information, this must have been expressly authorized in the regulation governing the set of statistics concerned and the survey documents must have made it clear to the persons questioned that this would happen.

If all these conditions are met, the various statistical offices as well as the other authorities and agencies involved in surveys have not only the right but also a duty to pass on these data.

1.1.2 The Federal Statistics Act (FSA) (Das Bundesstatistikgesetz)

1.1.2.1 The FSA as a further development of the FeSA.

The experience gained in implementing the FeSA and the new requirements laid down for statistical surveys were taken into account when drawing up the Federal Statistics Act (Das Bundesstatistikgesetz) of 14 March 1980, which replaced the FeSA and is currently in force.

The revised definition of the confidentiality of statistical information contained in the FSA reflects in particular the developments in the area of data protection and criminal law.

1.1.2.2 Meaning of the confidentiality of statistical information as laid down in the FSA

The principle of the absolute confidentiality of individual statistical information, which was at the heart of the

provisions in the FeSA governing the confidentiality of statistical information, was also incorporated into the FSA. Apart from the inclusion of more detailed definitions, the area of application of the principle of the confidentiality of statistical information was extended in some respects and restricted in others. The FSA also laid down additional conditions regarding the passing on of data; these, however, cannot be regarded as extending or limiting the confidentiality of statistical information since the scope for passing on information has changed.

The more precise definitions concern, firstly, those who have a duty to observe confidentiality. The term "persons entitled to receive information" was replaced by the phrase "officials and persons specially employed by the public authorities to compile Federal statistics". As a result of this change all the persons concerned became subject to the provision in the Penal Code which made the unauthorized disclosure of individual information collected for public administrative purposes punishable by imprisonment of up to one year or by a fine. In addition to staff of the Federal and Land statistical offices, these persons also include enumerators and interviewers as well as persons not employed by the offices mentioned who are engaged in processing Federal statistics or in similar tasks. Under the "Commitment" Act (Das Verpflichtungsgesetz), persons in this position who are not officials must formally undertake to fulfil their obligations in a conscientious manner, in particular the obligation relating to the confidentiality of statistical information.

The provisions waiving the duty to inform and notify tax offices was made more precise in that it now applies to persons and agencies engaged in the compilation of Federal and Land statistics.

In addition, it was expressly stated that information gained from statistical data on individuals may not be used for taking measures against the person concerned. This prohibition on using data in a way which is prejudicial to an individual ensures that personal data cannot be used by the public authorities when making decisions affecting individuals and, as such, it is likely to allay any irrational fears which the person questioned may sometimes experience.

The confidentiality of statistical information was initially extended in such a way that all persons not engaged in compiling Federal statistics but who received detailed statistical information had to observe confidentiality in the same way as persons and agencies involved in compiling the statistics mentioned. This closed a loophole in the FeSA.

Moreover, it was expressly stated that the transfer of personal data between persons and agencies engaged in compiling federal statistics is permissible if such transfers are required for the purpose of compiling the statistics. As a result, these persons and agencies also have a fundamental duty to observe the absolute confidentiality of personal data.

The provisions regarding the special treatment of identifying characteristics derived from the German data protection legislation. They stipulate the deletion of any data, in particular names and addresses, which can be used to identify persons obliged to give information and other persons concerned if such data are no longer required to carry out work connected with the compilation of federal statistics. The names and addresses of persons obliged to provide information must be separated from the other data and be kept in a special locked container. These provisions, incorporated into the Act at the instigation of the Federal Commissioner for Data Protection, are designed to take account of the special features associated with the processing of data for statistical purposes. Since the individual is of interest not as a person but only as a statistical parameter, data on him should be processed in as anonymous a form as possible, on condition that this can be justified on statistical grounds and does not involve an excessive amount of work. Thus, names and addresses must be physically separated from the data to be evaluated and kept in a safe place. If they are not required for statistical purposes, no identifiers (which may include not only names and addresses but also other prominent characteristics) can be stored in any shape or form.

On the other hand, when the FSA's provisions regarding the confidentiality of statistical information were drafted, it was necessary to incorporate additional cases where the duty to observe the confidentiality of statistical information did not apply.

On the basis of relevant provisions in the Federal Data Protection Act, the duty to observe confidentiality no longer applies if "in a specific case the person concerned expressly consents to the transmission or publication of the personal data provided by him". In this way, those who drew up the law indicated that they did not regard it as their duty to forbid the disclosure of personal data if the individual concerned did not appear to raise any objections to this.

In addition, in order to meet the greatly increased demand for personal data, especially for research purposes, a provision was included which makes it possible to transmit "depersonalized" personal data. This provision states that the data must be altered in such a way "that they can no longer be attributed to persons obliged to provide information or other individuals". If this condition is met,

the Federal Statistical Office or the Land statistical offices may pass on the data to anyone interested. Absolute anonymity of personal data is almost impossible to achieve and, therefore, when this provision is applied, it must be possible in each individual case to provide data which cannot be "re-personalized" and which, at the same time, provide the person receiving it with sufficient information.

The group of potential data recipients was also widened. It now also includes "other officials and persons assigned special tasks by the public authorities". The former comprise persons who carry out public duties in administrative bodies which were previously not regarded as being entitled to receive information, while the latter include the staff of scientific institutes carrying out work for public bodies. This group, too, is defined in such a way that criminal proceedings may be taken in the event of failure to observe confidentiality.

The persons added to the list of potential data recipients must fulfil additional conditions for the passing on of data.

Under these provisions, the transfer of personal data is permissible only "if such a transfer is authorized under the provisions governing the statistics subject to details of recipients and of the purpose of the transfer being given, and if the survey documents state that such a transfer will take place, giving the details mentioned". Those drawing up legislation relating to specific surveys must therefore decide for whom and for what practical purposes further use of the individual statistical information should be permitted. The duty to inform the persons questioned of this decision is designed to increase the transparency of possible data flows and hence facilitate the control of such flows.

Finally, in accordance with provisions in certain laws on specific surveys, every law of this kind which allows data to be passed on must state whether this may occur "with or without indication of names or of names and addresses". Questionnaires must also state whether or not this information will be given. This is intended in particular to take account of the principle of proportionality, which also covers the obligation to carry out public service tasks in a manner which is least detrimental to the persons concerned. The persons questioned must also be informed about the degree of anonymity of the data which might be passed on.

1.2 Supplementary provisions under data protection legislation

The Act on Protection against the Misuse of Personal Data during Data Processing (Federal Data Protection Act - FDPA) of 17 January 1977, already quoted several times above, contains general rules on the protection of personal data stored in data files. Unlike the FSA, the FDPA protects only data on natural persons. As it is a law of general application, more specific regulations, such as those relating to the confidentiality of statistical information, take precedence over it. However, its provisions give additional protection in those cases where facts are not, or not conclusively, governed by the duty to observe the confidentiality of statistical information.

This applies in particular to the confidentiality of data, the legal position of the persons concerned and the special measures required to protect data and ensure that data are processed in a proper manner.

While the duty to observe the confidentiality of statistical information is designed to prevent personal data being disclosed to third parties, i.e. persons not connected with the agency where the data are stored, the duty to observe the confidentiality of data goes one step further in that it also regulates the flow of data within the storing agency. It forbids persons engaged in data processing "to process, disclose, give access to or otherwise use protected personal data without authorization for a purpose other than that of carrying out their legitimate tasks". Even the unauthorized passing of personal information from one department to another within a single public body may constitute an infringement of this regulation. To ensure that the staff of such bodies are made aware of this provision, they must formally undertake not to pass on data in this way.

Under the FDPA, persons on whom data are processed have the following special rights: the right to be informed about the data on them which have been stored; the right to rectify these data if they are erroneous; the right to block the data if it is impossible to establish whether they are correct or erroneous and if the conditions governing their original storage are no longer met; and the right to delete the data if their storage is inadmissible and the conditions governing storage which were originally met no longer apply.

Since the majority of statistical data are rendered anonymous at an early stage, these rights are of practical relevance only in the case of data which are stored in data files so that they can be used in subsequent surveys.

The FDPA requires that persons processing personal data introduce extensive technical and organizational measures to meet the requirements of the Act. These are referred to as "data security measures". This is because they are closely related to the measures taken in automatic data processing known as "data security". These are steps taken to ensure that data processing occurs in the proper way and involve protecting hardware, software and data from loss, damage and misuse. While these measures relate only to automatic data processing and are taken to help the persons processing the data, the "data security measures" provided for in the FDPA apply also to manual data processing and are designed to protect the individuals concerned as well as act as an instrument to ensure that the practical provisions of the law are observed. In many instances a measure will prove to be of value both for "data security" and for the implementation of data protection provisions.

These measures include controls on physical access to data processing equipment, physical exit controls to prevent the removal of storage media, storage controls to prevent certain types of improper data processing, controls to prevent unauthorized persons from using data processing systems, access controls to ensure that users do not obtain unauthorized access, transmission and input controls, order controls to ensure that processing takes place in accordance with instructions, transport controls of the storage media and checks to ensure that the organizational structure within a public authority or agency is such as to meet the special requirements of data protection.

Since all these measures are costly, they are to be carried out only on condition that the cost involved is in reasonable proportion to the protection which it is hoped to achieve.

Where data are processed by agencies which do not form part of statistical offices but work on their behalf, there is a duty to carefully select the agency paying special regard to the suitability of the technical and organizational measures which it has taken.

In addition to the provision in the Penal Code which provides for sanctions against persons who fail to observe the confidentiality of statistical information, the FDPA also includes a penal provision which makes it a punishable offence to alter, recall or procure protected personal data from data files locked away in containers. Penalties range from a fine to a maximum of two years' imprisonment.

To enable the general public to monitor the data processing activities of public authorities and to improve the legal protection of the individual, details of stored data are published in the Federal Gazette. These details, to be published immediately after the data have first been stored, include the type of data involved, the tasks for which they need to be obtained, the group of persons concerned, the agencies to which the data will be regularly transmitted and the nature of the data which are to be transmitted. As regards statistical data, the FDPA requires details to be published only if the general public has not already received the information concerned through other channels, e.g. via the publication of the legal act governing the survey in question.

As publications of this kind are unable to produce any real improvement in the legal situation of the persons concerned owing to the huge volume of information to be communicated, they are in future to be replaced by extending the data-file register which must be kept by the Federal Commissioner for Data Protection.

Responsibility for ensuring that the Federal Statistical Office observes the data protection provisions is assigned to the Federal Minister of the Interior (FMI).

In particular, the latter must ensure that a record is kept of the type of personal data being stored, the tasks for which these data are required and the identity of those who regularly receive them.

The Federal Statistical Office therefore keeps records of stored data which contain all the requisite information, relating to all the physical features of a data file. Extracts from these records are forwarded to the FMI.

The FMI has also to set up a system of supervision to ensure that the data processing programs to be used for processing personal data are employed in a proper manner. The first task of this system is to ensure that the programs themselves meet the requirements of data protection legislation. The aim of this supervision is to safeguard the integrity of the program throughout the period when it is used.

Under an Order issued by the FMI, an internal "data protection commissioner" has been appointed at the Federal Statistical Office. His task is to give advice on data protection matters and to ensure that data protection provisions are observed.

The FMI has so far not issued any general administrative regulations regarding the implementation of the FDPA in his area of competence.

Under the FDPA, the office of Federal Commissioner for Data Protection was created as an overall external supervisory authority to ensure that data protection provisions are observed. The Commissioner must be provided with any information which he requests and be allowed to inspect any documents connected with the processing of personal data. In addition, he has unrestricted access to all official premises. The Commissioner must be notified of all data files which are operated automatically. Details are then entered in a register which anyone may inspect.

If the Commissioner ascertains that the federal administrative authorities have contravened data protection provisions, he may submit a complaint to the supreme federal authority with responsibility in that field. He may, at the same time, suggest ways of remedying the deficiencies and improving data protection. The Commissioner, to whom any citizen may submit a complaint, presents a report on his work once a year to the Bundestag.

The Länder have also appointed data protection commissioners, although their legal status differs considerably in some respects from that of the Federal Commissioner. They are responsible for supervising the administrative procedures used in cases where the Länder implement legislation relating to federal statistics.

1.3 New requirements arising from the right of the individual to control data relating to himself

With its decision on the 1983 Census of Population Act, the Federal Constitutional Court (FCC) expressly acknowledged the existence of the individual's right to control information relating to himself (henceforward referred to as the "information control right") as a concretization of the general right to privacy. This has far-reaching consequences for the collection and processing of personal data.

1.3.1 Definition of concept

The Court interpreted the information control right as meaning that the individual has the power, in principle, to control the disclosure and use of personal data relating to himself. A legal system "in which citizens are no longer able to know who knows what about them or when and under what circumstances such information is collected" cannot therefore be reconciled with the information control right.

This and other statements made by the Court clearly show that it does not regard the information control right purely as a right of individual liberty but as a precondition of a properly-functioning free democracy.

The information control right is not, however, an absolute right. In fact, it may be restricted if certain constitutional conditions are met. Thus, according to the FCC, a distinction must be made between personal data collected and processed in an individualized, non-anonymous form and personal data to be used for statistical purposes. The conditions governing statistical surveys are less stringent since these surveys serve many purposes. Thus, for example, when personal data are collected for statistical purposes, it is not necessary to specify a definite purpose. To compensate for this, however, the collection and processing of information within the information system must be subject to appropriate restrictions.

1.3.2 General requirements for the collection and processing of personal data for statistical purposes

Restrictions on the information control right which are in the overriding public interest are permissible by law.

While the constitutional proviso primarily applied hitherto to the collection of individual statistical information, it is quite clear that legislators must now draw up provisions covering events occurring at every stage of data collection and processing which significantly affect the fulfilment of the fundamental rights of the individuals concerned. For this purpose, they must issue regulations which conform with the principles of legislative clarity and or proportionality.

The principle of legislative clarity means that the law - where necessary, together with the background documents - must unequivocally indicate the nature and scale of the data collection and data processing concerned. The law must, therefore, at the very least, list the basic variables on which the citizens are to be questioned. It must also be made clear which data are to be used for statistical evaluation and which are needed to facilitate the actual carrying out of a survey. In this connection there are in principle no "insignificant" data, since the credibility of protection is determined not only by the sensitivity of a particular item of data but also, and in particular, by the scope for using it.

The proportionality criterion involves the principle of "prohibition of excess". Under this, the actual measures to collect information which are planned must always be those compatible with achieving the desired administrative purpose which represent the least inconvenience for the persons concerned. "Superfluous data", i.e. data which can be of no

assistance in performing public duties, and data on the individual's absolutely inaccessible private life ("data without social significance") may not be collected or processed.

Those responsible for drawing up legislation must also consider methodological questions. Thus, for example, they must ascertain whether a full survey could be replaced by a sample survey or whether a survey where there is an obligation to provide information could be replaced by one where information is given voluntarily. Similarly, they must give precedence to a "depersonalized" survey as opposed to one where personal data remain identifiable if the purpose laid down can be achieved equally well in this way.

The special treatment accorded to individual statistical information is based on the assumption that the latter will be used for statistical purposes only. Consequently, the FCC has laid down that it may be passed on only in anonymous form, although it is sufficient if this anonymity is of a *de facto* kind. The passing on of information in an individualized, non-anonymous form is possible only if the purpose for which it is to be used is laid down by law.

A factor of great practical significance and one which was ultimately responsible for the failure to carry out the population census was the great value attached by the Court in respect of the information control right to the principle that statistical activities should be separated from administrative enforcement procedures.

As regards the plans to update registers of residents using the statistical data obtained, the Court stated that a legal provision governing the compilation of statistics may not link together things which "tended to be incompatible". In particular, the Court considered it incompatible that, on the one hand, the Census of Population Act guaranteed the non-detrimental use of the data collected and that, on the other hand, one of the purposes of the register was in fact to enable measures to be taken against individuals.

The FCC's statements will have to be carefully examined to determine exactly what they mean. However, one point is already clear: the use of non-anonymous statistical data for administrative enforcement purposes, which would appear to be wholly reasonable since it avoids asking questions on two separate occasions and involves less administrative work, is permissible only if very restrictive conditions are met.

This conclusion is reinforced by the reservations which the Court itself expressed regarding a "combined survey", that is, a survey in which, in its own words, personal data were collected simultaneously on various questionnaires to be used for statistical purposes and the enforcement of administrative decisions. These reservations are based on the fact that "the linking together of two different purposes involving different requirements is highly

disconcerting for the individual citizen who cannot fully understand the scope of automatic data processing and, as such, can jeopardize the reliability of the information and its suitability for statistical purposes".

On the organizational side, the requirement that statistical activities should be effectively "sealed off" from the enforcement of administrative decisions complies with the basic principle of the separation of statistics and administrative enforcement. Where local authorities are involved in collecting data, as is the case, for example, with population censuses, it is they who have the greatest responsibility for measures which guarantee the separation of powers in respect of information. However, this responsibility is shared with all other agencies which are authorized to receive individual statistical information. For the Federal and Land Statistical Offices, the main means of ensuring that different areas are sealed off from one another is by protecting data from outside intervention. In this respect the Court shows its belief in the high degree of security existing in these offices.

As regards the use of data for internal statistical purposes, the FCC lays down certain rules aimed at safeguarding fundamental rights by means of procedures. Thus, in the first place, the Court limits the internal statistical use of data by declaring that it is inadmissible to link together personal data without authorization by law. In addition, the link between the data collected and the persons covered should be eliminated as soon as possible. The Court stipulates that statistical information on individuals should be made anonymous as soon as possible and that, at the same time, effective measures should be taken to prevent them being "re-personalized". The same purpose is served by the demand for greater precision in the separation and deletion provisions in the FSA.

Other organizational safeguards for the information control right are the requirements that surveys be conducted in such a way as to cause the least possible inconvenience to the individuals concerned, that enumerators and other survey officials should be carefully selected, and that data should be given effective protection during the collection process.

The information control right and the legal protection of the individual are also safeguarded by the duties imposed on public bodies engaged in statistical work to provide persons questioned with explanations and other information. In particular, these include the duty to explain a) the purposes of a survey, b) the meaning of any obligation to provide information and the various ways of fulfilling this obligation and c) how the data collected will be used. The attention of those questioned must also, where appropriate, be drawn to the voluntary nature of certain replies and to the separation and deletion provisions which apply.

The duty to keep records of all personal data of a statistical nature which are transmitted is also designed to reinforce the legal protection enjoyed by the persons concerned.

Finally, in view of the individual's inability to know exactly what is involved in automatic data processing and in order "to afford legal protection at an early stage by taking timely precautions", the FCC stresses the importance of the participation of independent data protection commissioners in providing effective protection.

2. CONSEQUENCES FOR FEDERAL STATISTICS

2.1 Consequences for legislators

The new requirements to be met by official statistics as a result of court decisions regarding the information control right can be fulfilled only with the assistance of those responsible for drawing up laws and regulations.

2.1.1 Consequences for those responsible for drawing up the FSA

The FCC has expressly called on legislators to draw up provisions covering a number of circumstances. Since the FSA reduces the need for legislation governing individual sets of statistics by itself laying down as many general rules as possible for the compilation of statistics, it is here that the legislators should concentrate their efforts to implement the information control right in specific areas.

In addition to provisions which make it unnecessary to have regulations for individual sets of statistics, it may also prove necessary to incorporate general rules in the FSA which would then have to be embodied in specific provisions in individual laws.

For the practical implementation of the information control right, a distinction must first be made between survey characteristics and auxiliary information.

Since the FCC considers that personal data collected for statistical purposes are subject to fewer restrictions when being processed solely because they immediately lose their identifying features, legislators must focus their attention on regulating what happens to the data during the period when it is possible, exceptionally, to identify the persons to whom the data relate using the auxiliary information which is still available.

In this connection, the first task will be to define the terms "survey characteristics" and "auxiliary information".

It is particularly important to make a clear-cut distinction here, since the data in question are to be processed in different ways.

Thus, legislators must precisely define at what point typical auxiliary information such as names and addresses must be separated from the survey characteristics. They must also specify the point at which the identifiers must be removed. If legislators feel unable to include generally binding provisions in the FSA, corresponding provisions must be incorporated into the laws governing individual sets of statistics. In this case, they will have to decide whether the FSA should not at least include a provision in the form of a blanket clause.

If legislators consider that the auxiliary information needs to be used for subsequent surveys, the FSA will also be the appropriate place to lay down provisions to this effect.

Other factors of general importance include provisions in survey procedures. Legislators must become much more active in this area than in the past, since the FCC attaches special significance to procedural provisions in safeguarding the information control right.

In the context of the FSA, this would primarily comprise provisions governing "typical" procedures such as the collection of data by enumerators, interviews, "voluntary" surveys and those where persons are obliged to provide information. Where necessary, legislators could draw up provisions on the selection of enumerators and interviewers which were aimed at preventing conflicts of interest arising for these persons because of their other activities, and other provisions on the legal situation of the persons questioned, particularly as regards meeting their obligation to provide information.

The FSA could also contain clauses to the effect that a statistical survey may be undertaken only by agencies which, in accordance with a legal provision, meet all the requirements for an effective separation of statistical work from other activities.

Generally binding provisions could also be included which specified under what conditions personal data may be regarded as anonymous or released for publication (e.g. in tabular form).

As an effective measure to prevent attempts to identify persons covered by data, legislators could incorporate a provision in the FSA making such attempts a punishable offence.

The FSA would also be a suitable place for laying down rules regarding the participation of the Federal Data Protection Commissioner in statistical surveys.

The requirement laid down by the FCC that legislators should ensure that questionnaires are compatible with the underlying legal provisions could also be covered by an additional clause in the FSA.

Other possible additions to the FSA would probably not be able to replace provisions relating to specific areas in the laws covering individual sets of statistics. For instance, a clause might be added which makes it permissible to pass on statistical data on individuals. However, assuming that the law continues to permit the transmission of statistical data on individuals, legislators will probably be able to do no more than lay down a general framework in the FSA within which specific laws could allow data to be passed on.

Similarly, owing to the circumstances peculiar to each survey, it is unlikely that a regulation covering the obligation of statistical authorities to provide persons questioned with relevant information could be adopted without specific provisions covering actual cases.

2.1.2 Consequences for legislators responsible for laws on specific sets of statistics

Legislators responsible for laws on specific sets of statistics are likely to be faced by a need to adopt a whole series of amendments to take account of the information control right. In future, when they draw up new laws governing statistics they must also consider methodological aspects.

There is virtually no legislation which distinguishes between survey characteristics and auxiliary information. If it was thought inadequate that, in the course of the administrative enforcement of laws relating to statistics, the persons questioned were told which questions were designed solely to facilitate the carrying out of the survey and which would be statistically analysed, this would be a sufficient reason to amend almost all the laws covering statistics.

In general, the survey procedure to be adopted should not be laid down in such a way that all the factors affecting fundamental rights are covered. Here, there is as yet no evidence as to what general provisions in the FSA can achieve. Amendments will probably be required at least in respect of surveys of an unusual type.

There are virtually no provisions regarding the facts which have to be explained to respondents when a survey takes

place. Here, too, it is doubtful whether a general provision in the FSA can fill this gap.

It also appears extremely unlikely that a separation and deletion provision in the FSA could meet the requirements laid down by the FCC for a sufficiently specific provision. All legislation governing individual sets of statistics which authorizes the collection of personal data might have to be supplemented by provisions of this kind.

However, there can be no doubt that the individual laws which provide for the passing on of statistical information relating to individuals need to be amended. Scarcely any of these laws take account of the limitations which in future will be imposed in information which is passed on.

Furthermore, in cases where it is planned to combine personal data from various statistical sources, such actions must be expressly covered in the individual laws governing the collection of the data which are to be combined.

If the legislators were to decide to enact individual laws on statistics relating to specific areas rather than draw up an "amending law", they would almost certainly have to examine methodological questions in the process. In the view of the FCC, the legislators must decide in every individual case whether the procedures used are in line with the latest scientific findings as regards the degree of inconvenience which they cause. If, for example, modern sampling techniques produce results which are just as good as those obtained by a full survey, instructions to hold a full survey would contravene the principle of proportionality and therefore be unconstitutional.

2.1.3 Consequences for persons drawing up laws and regulations

In some areas the FCC has not specified how the information control right is to be safeguarded. It has merely stated that the legislators are free to decide whether they will authorize those who draw up regulations to introduce relevant provisions or whether they will allow the administrative authorities scope to take their own decisions, with rules setting up a system of supervision to ensure that these decisions are carried out correctly. The main purpose here is to ensure proper administrative enforcement of the laws on statistics; for example, it is necessary to make sure that the questionnaires contain only questions which are covered by the underlying law.

The technical and organizational measures required for an effective segregation of statistical data should also be mentioned.

Measures to avoid conflicts of interest among enumerators and other survey officials need not be stipulated by law either.

2.2. Consequences for the conduct of statistical surveys

The judgement of the FCC on the 1983 Census of Population Act does not only have consequences for legislators. It also lays down guidelines which the authorities engaged in statistical work must observe when collecting, preparing, processing and presenting statistical data.

2.2.1 Collection of data

The obligation derived by the FCC from the principle of proportionality to the effect that the "least detrimental" means should be employed, i.e. the means which interfere least with fundamental rights, is of central importance especially during the data collection stage. It means, *inter alia*, that the statistical offices have a duty, when carrying out surveys, to consider whether the purposes for which the information is required could not be achieved with procedures which interfered less with fundamental rights. The degree of such interference depends to a large extent on how comprehensive the questions in a given survey are, the degree to which individual questions are concerned with the private life of the persons providing the information and whether those questioned in the survey are obliged to provide information.

Consequently, a survey involving voluntary participation is the "least detrimental" data collection procedure of all. However, even if participation in a statistical survey is voluntary, the information control right is guaranteed only if the person concerned can consciously decide whether or not to provide information and determine how data on him can be used. This presupposes that the person knows why the data are being collected. Thus, the statistical offices are also bound by their duty to provide explanations and other information in the case of surveys where information is given on a voluntary basis.

However, in cases where the provision of information must be compulsory, those responsible for the survey must determine whether the purpose for which the information is collected can be achieved by asking fewer questions and making those which are asked less closely related to the social status of the persons questioned. In this connection, special care must be taken to ensure that no data are collected which might lead to a social stigma being attached to the persons concerned.

The extent to which a statistical survey interferes with the rights of a person providing information is also determined by the identifiers such as names, addresses, personal distinguishing features, etc. which are used when the survey is carried out. If the purpose for which information is to be collected allows it, surveys based on anonymous data are to be given precedence over surveys which use identifiers during the collection stage. However, if identifiers are

included in the questionnaire they must be listed as auxiliary information, together with the survey characteristics, in the legislation covering the survey concerned. In accordance with the principle of proportionality, this auxiliary information must be kept to a minimum.

The degree of inconvenience caused by a survey is also affected by the number of persons who must provide information (the coverage of the survey) and their relationship to the total number of those who might be asked to provide information (sampling ratio). The greater the number of persons asked questions in a survey is and the higher the sampling ratio becomes, the greater the "detriment" caused by the survey is bound to be.

The degree of interference in the information control right is also affected by whether the survey is carried out using enumerators or interviewers or whether other agencies not connected with the statistical offices are called in. If it is not possible to incorporate guarantees of a legally proper involvement of these agencies in the legislation governing individual surveys, the services of such agencies must be dispensed with. If they are invited to take part, there must be no possibility of conflicts of interest and it must be possible to segregate these agencies from other administrative bodies.

To summarize what has been said about the consequences of the FCC's judgment on data collection activities, particularly with a view to future developments, it can be expected that:

- greater significance will in future be attached to "voluntary" surveys as far as federal statistics are concerned;
- fulfilling the obligation to provide explanations and other information will make survey work more costly and more time-consuming;
- discussion will in future focus more on the methodological aspects of data anonymity;
- the involvement of enumerators and interviewers and of data collection agencies not connected with the statistical offices will have to be properly justified.

2.2.2 Preparation and processing

The FCC's demand for specific conditions governing data processing as well as the requirements regarding the segregation of statistical sources and the removal of identifiers at an early stage are highly relevant to preparation and processing of work.

For the statistical offices, this means that more organizational records and data flow characteristics will have to be prepared so as to provide a clear description of the conditions under which data in individual surveys are

processed and the procedures involved in such processing, thus ensuring that the operations involved in the processing of personal data are made clear to anyone who is interested, including the data protection commissioners.

The requirement that data can be depersonalized at an early stage is met by separating and deleting, as soon as possible, auxiliary information which might help to identify the persons covered.

Effective segregation arrangements, normally based on a particular philosophy of data security requirements have already been drawn up in the course of the general implementation of Federal and Land data protection legislation and practical measures designed to meet them have been taken in the Federal and Land Statistical Offices. The latter's general approach has been described as exemplary by the FCC. Further legal provisions are not required in this area. However, particular attention will have to be paid in future to the security measures taken by third parties (e.g. service companies) which are occasionally asked to assist in carrying out surveys. Legal rules may, for example have to be drawn up to "segregate" such companies.

In the context of preparation and processing of work, a factor of particular significance for the information control right is the extent to which data from various statistical surveys are combined with one another. Where the data concerned are in the form of aggregates, this is of no consequence to the individual who has provided information for one or more of the sets of statistics concerned. The "statistical" combining of personal data records for various surveys with identifying features removed are unlikely to pose any problems either. "A genuine" combination of personal, "identifiable" data records from various surveys by means of common identifiers must have legal sanction. Without such a sanction, it would be at variance with the judgment of the FCC.

To summarize what has been said about the consequences of the judgment for preparation and processing work, it is clear that:

- the accelerated introduction of organizational records and data flow charts will lead to additional expenditure;
- the separation and deletion obligation connected with the requirements that data should be depersonalized at an early stage will lead to changes in organizational procedures and a substantial increase in expenditure;

- the number of outside agencies involved in preparation work will gradually go down;
- it will become more difficult to combine personal "identifiable" data records from various sets of statistics.

2.2.3 Presentation

The presentation of data by the statistical offices is not affected by the FCC's judgment except where the publication or passing on of personal data is concerned. The judgment does not affect the dissemination of aggregate data via publications, statistical information system, etc.

According to the FCC's judgment, personal data collected in a survey may, assuming it is expressly authorized by law, be passed on, on condition that this is for the purpose of purely statistical processing by other agencies and that the measures required to protect the individual's right to privacy can be reliably guaranteed. On the other hand, the arguments put forward by the FCC indicate that it might be inadmissible to pass on personal "identifiable" data to facilitate the enforcement of administrative decisions. As a direct result of the judgment, the Federal Statistical Office has for the time being suspended the transmission of any personal data which retain identifying features until a decision is made regarding amendments to the relevant legal provisions. Depersonalized personal data, i.e. data on individuals which cannot be traced back to the persons covered, may continue to be passed on and published by the statistical offices.

Since the scope for passing on "identifiable" personal data has been drastically reduced, it is probable that demands for depersonalized personal data will increase. In the view of the FCC, the demands for the anonymity of personal data, as explained above, should not be set too high. Opinion is tending to favour a *de facto* anonymity, a solution which has already been discussed, *inter alia*, by experts in the field.

To summarize the above, it would appear that:

- the judgment does not affect practical work concerned with presentation of statistical results except for those exceptional cases where the results consist of personal data;
- the passing on of "identifiable" personal data will continue to be subject to drastic restrictions, probably in the longer term as well;
- in contrast, the passing on of depersonalized personal data will be made easier.

3. THE WAY OUT OF THE DILEMMA

3.1 Short-term reactions of legislators, the federal departments concerned and the statistical offices to the judgment of the Federal Constitutional Court

The discussion of the effects of the judgment on the work of legislators and on the actual conduct of statistical surveys by the statistical offices has already shown that many steps have to be taken as a result of the judgment. Not all the work which is needed can be carried out at once. The question therefore arises as to the extent to which the legislators concerned and the administrative authorities have a right to defer the requisite implementation of the Court's judgment.

3.1.1 Deferment of implementation of Constitutional Court's guidelines.

When a statutory rule is incompatible with the Basic Law, the rule normally has to be declared null and void. However, if the present legal arrangements governing statistics can be implemented in a constitutional manner, bearing in mind the decision of the FCC, it is not absolutely imperative, at least in theory, to withdraw the legislation on which these arrangements are based. If such a view is taken, it would appear sufficient that the legislators rectify the state of affairs which is unconstitutional within a suitable period.

In practice, however, deferring implementation in this way would not be of much value because, although the statistical offices had stated it might be possible to carry out the 1984 Microcensus in a manner which did not conflict with the judgment of the FCC, the survey in question was postponed. It still appears doubtful whether a census on distributive trades and catering due to be held in 1985 can be carried out on the basis of a law introduced before the FCC delivered its judgment on the census of population, although the statistical offices are endeavouring to organize the survey in such a way that it conforms with the FCC's guidelines.

3.1.2 Bringing collection and processing procedures in the statistical offices into line with the guidelines

Most current sets of statistics are in fact compiled in accordance with the FCC's guidelines.

The measures taken in this regard include the statistical offices' efforts to make the contents of questionnaires used in individual surveys conform as closely as possible to the working of the law on which the survey in question is based

and, where doubts arise, to dispense with one or more questions in the questionnaire in accordance with the principle of legislative clarity.

The statistical offices are also attempting to comply with the obligation to provide explanations and to the information laid down by the Constitutional Court and are endeavouring to depersonalize data at an early stage by designing survey documents in an appropriate way (with identifiers on detachable sheets) and by taking organizational measures affecting data preparation.

Compared with the situation before the FCC's judgment, this involves additional work. The extra expenditure thereby incurred may be regarded as the price which has to be paid for improved data protection.

The steps taken to adjust to the new situation include the practical suspension by the statistical offices of the provisions regarding the passing on of personal data. Provisions of this kind are found in no fewer than 30 laws governing individual sets of statistics.

3.1.3 Requisite amendment of legislation governing individual sets of statistics

Although the statistical offices are endeavouring to organize data collection in a constitutional manner in accordance with the guidelines laid down by the FCC, the requisite amendments must be made to legislation on specific sets of statistics by the legislators and, during the preparatory stage, by the responsible federal departments, since such amendments cannot be postponed indefinitely.

The preparatory work on drawing up a new Census of Population Act began very soon after the FCC's judgment was handed down and will soon be completed. The draft of a Microcensus Act complying with the principles embodied in the Court's judgment has also been prepared, as have drafts for an amended law on higher education statistics.

Work is also under way on drawing up an omnibus law (amending law) to bring the legislation underlying the other current sets of statistics into line with the guidelines contained in the FCC's judgment. In particular, this affects the provisions governing the passing on of personal data contained in over 30 specific laws, the inclusion of more precise details in survey programmes and, finally, organizational and procedural rules.

3.1.4 Further development of the Federal Statistics Act

In addition to the adaptation of legislation on individual sets of statistics the FCC's judgment has also made it necessary to amend the basic Federal Statistics Act. This is now taking place, together with the revision of the individual laws on statistics.

The revision of the Federal Statistics Act also, as mentioned above, mainly affects the provisions regarding exceptions to the confidentiality of statistical information, the passing on of data and procedural rules. It remains to be seen whether the amendments which have to be made to the Federal Statistics Act can be used not only to bring the Act into line with the FCC's guidelines but also to introduce legal provisions for important innovations in the further development of federal statistics.

3.2 Longer-term arrangements to bring the plans, methods, procedures and organizational measures adopted for federal statistics into line with the judgment

The revision of the relevant legislation in accordance with the guidelines contained in the FCC's judgment means, first of all, that the compilation of federal statistics must be made to conform with standards imposed from outside by the highest court in the land. In the longer term, the most significant factor will be whether those responsible for federal statistics can take the principles embodied in the judgment and use them as a basis for actively shaping the plans, methods, procedures and organizational measures required for such statistics. For this purpose, the judgment offers a number of points of reference, which will be discussed in greater detail below.

3.2.1 The development of a modern demand-orientated programme

In its judgment the FCC stressed how important statistics were for carrying out vigorous policies in conformity with the Basic Law. However, if federal statistics are to provide comprehensive, continuous and up-to-date information on the economic, social and ecological situation, it is especially important that they be geared to respond to urgent needs. For this purpose, all the bodies engaged in drawing up plans for federal statistics should give early consideration to the question of which problems will occur in future and which will require an urgent solution.

This is the only way to ensure that the information required for decision-making can be made available in good time.

Therefore, over the next few years, the planning of the programme of statistics is to be based on present and future social, economic and technical developments and will cover all important official activities.

Under this plan, programme development work will be carried out in stages: long-term overall planning, long-term outline planning and short-term detailed planning. During the overall planning stage, the general needs for quantitative information both for the present and for the foreseeable future will be determined and described. A second stage (outline planning) will follow, based on this overall planning. The object of this stage is to establish the extent to which the general need for information previously determined can be met by an overall statistical system. By comparing the need for information thus determined (information useful for statistical purposes) with existing data, it is possible to identify shortfalls which must then be met rationally in a way which imposes the least possible burden on the persons questioned. This will be followed by a third stage (detailed planning), during which plans for individual sets of statistics will be drawn up, based on the outline planning.

3.2.2 The discussion of methods and procedures

The discussion of methods and procedures will become important during the outline planning stage, if not earlier, i.e. during the stage when plans have to be drawn up to remedy data shortfalls in a rational manner which imposes the least possible burden on those providing the information.

As mentioned above, the FCC has imposed a duty both on legislators and on the authorities who plan the surveys to select, from among the data-collection methods and procedures suitable for the particular purpose in mind, those which impose the least burden on the persons questioned, i.e. those which least interfere with the individual's right to privacy.

Little progress is likely to be made with surveys of a traditional type, particularly in the case of new fields of investigation now being covered by statistics (e.g. the black economy, the use of new technologies, the social situation of certain groups of the population). The many different demands which have to be met when examining these areas require that those responsible for compiling federal statistics be much more flexible than in the past when designing and carrying out surveys.

Making greater flexibility a reality poses particularly difficult problems for legislators. However, only legislation can prepare the ground to enable statistics in

the Federal Republic of Germany to gain the necessary flexibility, of the kind taken for granted in other western countries when official statistics are compiled.

In this connection, an important contribution could be made by making legal provision for small-scale surveys with voluntary participation, which could be carried out with simpler procedures than those used for normal surveys where the giving of information is compulsory and for the large-scale surveys based on voluntary cooperation which have been carried out so far. Particularly in view of the new fields of investigation now being introduced, legislators would be taking on an impossible task if they attempted to specify every characteristic which was to be covered by the survey programme.

The inclusion in the Federal Statistics Act of appropriate legal provisions governing the organization and the conduct of small-scale surveys based on voluntary participation is an essential prerequisite if those responsible for compiling federal statistics are in future to be able to meet the demands made upon them.

3.2.3 Further developments in the organization of federal statistics

In explaining its judgment, the FCC laid stress on the principle of separating administrative enforcement and statistical work. As already mentioned, this reflects in particular the view that data collected for a specific administrative purpose may in principle be used only for this purpose and, where appropriate, for statistical purposes as well. On the other hand, "identifiable" personal data collected for statistical purposes may not be used for administrative purposes.

The FCC's view underlines the relevance of the so-called "one-way-street system" which has already been frequently discussed by statisticians. The basic idea is that the system of federal statistics, as in certain other western countries, should be developed so that it becomes a sort of one-way-street for personal data. Under this system, the statistical offices will obtain their data by means of primary surveys which they carry out themselves and by using information from public administrative authorities for purely statistical purposes ("secondary statistics"). Combined surveys designed to serve the purposes of both administrative enforcement and statistical compilation will not be allowed. In these "one-way street" models steps must also be taken to ensure that statistical offices have access to all administrative records and that, when the latter contain information of relevance to the statistics as a whole, the statistical offices should be able to use this information, at least in anonymous form. This would ensure that the largest possible volume of statistically analysable data flowed into the system of federal statistics.

On the output side, the one-way-street principle would make it impossible to pass on identifiable personal data. Such data would be made available to users only in a depersonalized form.

A broad flow of statistically analysable data into the system of federal statistics would, however, also require changes in the rules for processing the material. Most standard tabulation programs would have to be shortened, while special and additional processing operations would become more important. To meet these requirements, greater attention would have to be focused on individual problems and projects, which would also affect the internal organization of the statistical offices. Given the situation described, these steps would automatically produce a greater flexibility in planning and in work procedures.

, *

* *

DISCUSSION

M. DE VRIES (*Centraal Bureau voor de Statistiek*) opened the discussion:

The judgment of the German Federal Constitutional Court had introduced a state of real uncertainty into the work of compiling statistics and was of considerable interest and importance to other countries also, and particularly to the Netherlands. Mr. De Vries asked the following questions arising from the paper:

- (1) To what extent were personalized data arising from surveys available to the staff of the Statistisches Bundesamt?
- (2) It was mentioned that the costs involved in data protection must be in reasonable proportion to the protection achieved: how were these norms determined?
- (3) Was there a contradiction between the statement in Section 1.3.1 that, when personal data were collected for statistical use, it was not necessary to specify a definite purpose and the statement in section 1.3.2 that there was a duty to explain to persons questioned the purposes of a survey?
- (4) What was now the position as regards the passing on to third parties of statistical data on individuals?

- (5) What was the position as regards the passing on, for administrative purposes, of information collected for statistical purposes?

Before proceeding with the discussion, Mr. WERNER and Mr. SUDFELD (Statistisches Bundesamt) replied to these questions of Mr. de Vries:

- (1) Personalized data, with identification, from surveys were accessible only to those officials directly concerned with the organization of the surveys. The material necessary for the identification of individuals was removed from the records at as early a stage as possible in the processing and thereafter the magnetic tapes were available to members of staff of the Statistisches Bundesamt needing to work on them. Insofar as the combining of individualized data from different sources was concerned, this would require a specific legal authorization.
- (2) On the relationship between data protection measures and their costs, the Federal Protection Law gave some guidance. The relevance of the measures and their cost must be in reasonable relationship to the protection afforded and the risks to which the data were exposed. In deciding such questions the Data Protection officials as well as of the official working with the data were consulted.
- (3) The Constitutional Court recognized that a statistical inquiry was in a somewhat different situation to that of an inquiry by a police or registration official. Whereas in the latter cases, the precise purpose of the inquiry must be stated, a certain flexibility was acceptable in the case of a statistical inquiry.
- (4) It was mentioned in the paper that the passing on to third parties of information on identifiable individuals had been stopped for the time being and there was no doubt that such transfers would be very limited in the future. On the other hand the Court allowed the passing on of anonymized data and had made the conditions required for doing so more liberal. In particular no great obstacles should be put in the way of supplying data for scientific research.
- (5) With regard to the passing on of data for administrative purposes, insofar as aggregate data were concerned there was no problem. The passing on of individualized data was, however, completely forbidden. An inquiry having

both administrative and statistical purposes was no longer permitted. Statistical data might be extracted from administrative records although the position was still not clear where the administrative records related to identifiable individuals.

Mr. BEGUE (*Institut National de la Statistique et des Etudes Economiques*) remarked:

Referring to the use of administrative records for statistical purposes in the Federal Republic he asked whether, when a new system of administrative records was being introduced, there existed arrangements for consultation with the statisticians to ensure that their needs were taken into account. Such arrangements existed in France, organized by a National Council for Statistical Information.

Mr. FLAHERTY (*University of Western Ontario*) intervened:

He asked the authors to comment on how, since there were good Statistics Acts and good Federal Data Protection Acts in existence in April 1983, the census of population had nevertheless run into difficulties. He noted that the Federal Minister for the Interior had appointed a data protection officer to work in the Statistisches Bundesamt and asked for some information on the status and method of work on this official. He expressed surprise at the remarks (Section 1.3.2) that the use of non-anonymous statistical data for administrative purposes appeared wholly reasonable, in view of the considerable controversy on this point.

Mr. RAPAPORT (*Statistics Sweden*) asked:

Whether it was now possible to carry out longitudinal surveys in the Federal Republic following the judgment of the Constitutional Court.

Mr. BARNES (*Office of Population, Censuses and Surveys*) enquired:

Whether the Constitutional Court defined what was meant by "depersonalization" of records or gave any advice as to where lines should be drawn. Clearly name and address identified individuals but sometimes, even without name and address, unique combinations of characteristics, especially in small geographic areas, might be sufficient to identify people.

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

A remark of Mr. Rapaport said that future strategy should be to stress to the public the value and importance of

statistics to the community and at the same time compare the problems associated with the requirements of data protection and hence eliminate them in this way. One could conclude from this remark that the extent to which data were to be protected should be determined by the persons involved. In the Federal Republic of Germany at least, this would not work. According to the judgment of the Federal Constitutional Court on the Population Census Act, data protection had constitutional status. The guidelines and administrators were not therefore normally in a position to seek the most practical solution but must base their decisions on the basic principles formulated by the Federal Constitutional Court. Dr. Baumann remarked that he was rather irritated by the reaction of some European statisticians to this judgment, because the decision was clearly in support of statistics and demonstrated how important and indispensable they were for any policy decision. The judgment freed statistics from the dual function they had to fulfil if data were also to serve an administrative purpose and at the same time gave the citizen the assurance that his data would be used exclusively for statistical purposes. While a number of points in the judgment required explanation, the statements on statistics were unambiguous and the present bills on the population and microcensus made this clear. Both bills maintained the principles formulated by the Federal Constitutional Court. In practical terms their implementation would certainly be more complicated for the administration than the previous procedures but they did ensure a high degree of openness in statistical operations.

It had not been expressed sufficiently clearly that the topic of the seminar, "Protection of privacy" had been considered primarily from the point of view of the processing of data recorded for statistical purposes in the discussion. The judgment of the Federal Constitutional Court, however, quite clearly stated that this kind of protection should be introduced much earlier: the legislator had to take action to ensure the protection of privacy during the actual recording of data. For this reason according to the judgment, the following measures were necessary:

- the person who was obliged to supply information had a right to explanations and information, about which he should be notified in writing;
- the person obliged to supply information should be allowed to hand the questionnaire to the enumerator - who would not have inspected it - in a sealed envelope or to send it by post;

- it should be possible for the person obliged to supply information to give this only to enumerators who had absolutely no conflict of interests;
- personnel and organizational measures should be adopted to ensure that during the survey the offices involved were hermetically sealed from the other administrative departments.

Mr. JENSEN (*Danmarks Statistik*) remarked as follows:

The principle of "information control right" could be applied only in a very general way, since no one could control his data through all the stages of processing in a statistical office. The creation of ways and means to guarantee information control in the general sense was a matter which should be dealt with in the legislation on data protection. This was the case in Denmark. A specific set of rules was formulated for every single register and these detailed rules were made public. In this way the Danish population had every opportunity to follow and, in a general sense, control the use and the circulation of individual information.

Mr. HARRIS (*EUROSTAT*) pointed out:

The present position as regards German statistics was a matter of great concern to the Commission. The Commission was being obliged to tailor Community legislation to meet German requirements and this situation gave rise to problems with other countries. It also gave rise to problems with Community legislation since the Commission was, under the Treaties, entitled to ask Member Countries for the information it needed to do its work. Noting the recommendations in favour of voluntary surveys and public relations he felt that there was need for a profession of "official statistician". Such professionals would be not only numerate but also highly literate with a feel for politics and publicity. He welcomed Mr. Werner's thesis that a fairly general statement of use in regard to statistical surveys should be acceptable. If one had conducted a household survey for use in employment policy, it would be far too rigid an interpretation of the Court's judgment to forbid certain cross-classifications simply because they had not been and could not have been foreseen at the outset.

Mr. SCHOETTL referred to the principle of proportionality, well known in France, whereby the limitations imposed by the police on liberty of individuals should be strictly confined to the needs of the general policy aim involved. He asked

whether the application of this principle in the collection of German census statistics was accepted unanimously or whether it gave rise to criticism. He asked whether there was not at least some debate in Germany on the principle of "information control right". The logic of this principle was that it conferred on individuals a veritable copyright on all the information relating to them. This could affect not only the compilation of statistics, but also the very functioning of the State itself.

Mr. BIEVER (*Commission consultative pour la protection des données nominatives*) made the following points:

- (i) Article 8 of the European Convention on Human Rights guaranteed the right to privacy. It was not evident that having regard to its jurisprudence the Court of Human Rights in a case similar to that before the Federal Constitutional Court, would decide exactly the same thing (judgment Malone).
- (ii) The distinction between personal and anonymous data was sometimes unclear, particularly in the case of identification as belonging to a group.
- (iii) The problem of protection of personal data and meeting the needs of statisticians must be reconciled. Otherwise there was the risk that the experience of the Federal Republic would be repeated in other countries.
- (iv) If one wished to give the public more complex information about the value of statistics, one should not, at the same time, ask for their interconnection with other files held by the administration. These two procedures risked being profoundly contradictory. The principle of finality which should apply to each data base was thus violated.

The author (Mr. WERNER) replied to the main points raised:

With regard to the question of Mr. Begué, in order for administrative data relating to individuals to be used for statistical purposes in the future, the uses would have to be clearly specified in advance in the enabling legislation. The prospects for such an arrangement in the present climate of opinion were poor, as the necessary political will was lacking.

He agreed with Mr. Flaherty that the paper left a gap in that it did not discuss the failure of the 1983 census despite the apparent adequacy of the data protection

legislation in force at that time. However the authors had confined themselves to the legal situation and did not attempt a description of the psychological situation prevailing at the time of the census.

The data protection officer in the Statistisches Bundesamt had a double function: he was leader of the legal and data protection groups and had a grade immediately under that of head of department. In the past, contact with the data protection office had been mainly in regard to specific problems. Since the Court judgment, a committee had been set up to ensure regular contact between the Statistisches Bundesamt, the Federal data protection office and the Länder.

On Mr. Flaherty's third question, it was a simple statement of fact that a joint statistical/administrative survey could, in certain cases be the most efficient arrangement. The paper pointed out that the conditions under which such a survey could be carried out were very restrictive and needed to be examined.

In reply to Mr. Rapaport's question, longitudinal studies were still possible provided that this use of the data was previewed in the legal basis for the study. It must have been stated that the data might be linked over a period of time, so that an immediate anonymization of the data was not possible.

On the question of "depersonalization" posed by Mr. Barnes, the Karlsruhe judgment simply spoke of the practical anonymization of records, which was rather less precise than the principles contained in the existing Statistics Acts. This left rather greater scope for the Statistisches Bundesamt which must examine the practical situation with regard to each inquiry and to assess the risk that data might be identified from it.

He said that Mr. Schoettl was incorrect in believing that the Karlsruhe judgment conferred on the individual an absolute right of control over his own data. This right could be limited by the appropriate legislation. This, he thought, was consistent with the view of the European Court of Human Rights mentioned by Mr. Biwer, that there were no absolute rights in the protection of information, at least insofar as information of concern to the State was concerned. The Court recognized that there was such an absolute right in regard to information of an intimate nature.

The author (Mr. SUDFELD) continued:

He referred to the situation prevailing in the Federal Republic in the spring of 1983 which led to the collapse of the proposed census and which was due to a chain of unfortunate factors going far beyond questions of statistics. The 1983 Census Act had been passed unanimously in 1982 and there had been nothing to suggest that a massive protest movement against the census would build up in a period of a few months. Because the elections for the Bundestag coincided with the period of preparation for the census, the politicians who had voted for the census either gave no support or very half-hearted support in the face of the developing criticism. He doubted whether the implications of the judgment of the Constitutional Court, pronouncing in favour of the population census, had penetrated to the general public. Consequently he feared that there still existed considerable ill-will on the part of the public towards the census. He did not think this was an exclusively German problem: given equally unfavourable circumstances, the same situation could arise in other countries.

STATISTICS AND ADMINISTRATION

Prof. Dr. W. BEGEER, W.F.M. DE VRIES, H.D. DUKKER
Centraal Bureau voor de Statistiek

SUMMARY

This paper by the Dutch Central Bureau of Statistics provides an outline of the particular status that a bureau of official statistics should possess within society. This is illustrated by noting the difference between administrative and statistical information: in the process, the terms "administrative immunity" and "potentially censurable or vulnerable entity" are introduced and explained.

0. INTRODUCTION

Although the title of the seminar is "protection of privacy", and these words are generally taken to mean maintaining the secrecy of individual data, the authors intend to give a wider meaning to the subject.

Public servants are generally forbidden to discuss individual data with unauthorized persons or to release such data to them. Their work will, however, make it necessary to use individual data and this will often result in the exchange of data with colleagues and the transmission of data to other - possible even prosecuting - authorities. In such cases, secrets are still being kept from unauthorized persons, but the group of authorized persons can be very large and, to some extent, defined by the people involved themselves.

The authors feel that in the field of statistical information, the protection of privacy should be more widely interpreted than in the cases described above. In order to develop this idea, they make a distinction between information needed for administrative purposes and statistical information, which, in turn, involves a discussion of the specific aim of statistical information. With an eye to the importance of good statistical information, they describe the status which should, in their view, be accorded to "official" statistical offices and also extend the concept of privacy by using terms such as "administrative immunity" and "potentially censurable or vulnerable entity".

1. ADMINISTRATIVE INFORMATION AND STATISTICAL INFORMATION

Administrative information is taken to mean information at the level of the individual, which is required for administrative or governmental purposes. In principle, this description involves a process of interaction between the administrative or governmental body and the individual entities. Information about individuals is used for administrative or government action that concerns the individual entities. In some cases, the information is already available and is used as needed.

In other cases, (additional) information is requested before action is taken. One well known example is the way that income tax is organized. A certain level of basic information is permanently available, comprising at least the name and address of taxable persons. In most cases, their old files are also available. In order to calculate the taxes payable, additional information (tax return) is requested, whereupon, perhaps after checking the requested information, the tax payable is determined.

Statistical information is taken to mean information about groups of individual elementary units, compiled from information about individual entities, that has been transformed in such a way that it is not possible to identify information from specific entities.

This description contains a number of concepts that require further explanation. Firstly, it always concerns information about groups or information about "phenomena". The typically American expression "her statistics are" concerns another meaning of statistics than the one we have in mind. Information about a group can be a total, but statistics are more often expressed as (frequency) distributions across categories. Additionally, group information is based on information about individual entities. It may comprise information from all the individual entities belonging to the group (comprehensive survey) or from only one part of it (sample). It is even possible for information obtained at the level of individuals to concern entities that do not, themselves, belong to the group involved. At this point one is confronted with the well-known difference between "sample population" and "target population". Where certain phenomena are the subject of time series, the statistical information obtained is often based on (parts of) groups whose composition is subject to variation.

In all these cases, however, information about individuals is transformed into information about groups in such a way that a reverse transformation is not possible. This last requirement can have far-reaching consequences for the statistical information to be obtained. After all, when there are few entities in a given category and the category is described in terms of a large number of attributes, it

may become possible to identify the individual entities concerned. It is also possible that different "statistics" about a single group may be combined and thus lead to identification of individual entities. In such cases, one refers to "disclosure" of individual data. This is not a field in which a statistician may adopt a purely deterministic approach, since the vital question is not whether it is possible to identify individual entities with complete certainty - even the possibility of making statements, with a high degree of probability, about individual entities is, in fact, unacceptable. Conversely, it is never possible to state with absolute certainty that, in a given situation, disclosure of individual data is impossible. The rather normative pronouncements made above are to be regarded as setting standards, applying to the concept of "statistical information" within the framework of "official" statistics. This rider contains two constraints. Firstly, we would not apply these standards to statistical information in general and, secondly, we do not state that statistical offices should always aim, regardless of the circumstances, at non-identifiability or the avoidance of any possibility of disclosure. The question of standards applying to offices for "official" statistics will be discussed later.

By differentiating between administrative information and statistical information we did not mean that the latter cannot be obtained from the former. The individual entities supplying the information required for administrative purposes will, of course, bear in mind the possible consequences of the information supplied from the administrative or governmental viewpoint. As a result, the information supplied may to some extent be distorted (or touched up?). From another angle, it is quite conceivable that certain forms of administrative information will be of such a high quality that for statistical purposes these are preferred to possible alternatives.

The aim of statistical information (in the context of "official" statistics) is quite different, and on another level, from that of administrative information. It may be the collection of information for the preparation, monitoring and evaluation of policies. The word "policy" is taken here to mean the adoption of measures or the taking of action in relation to phenomena or groups, not to their individual entities. Both purposes require good statistical information (although, of course, good information is also needed for administrative purposes).

2. THE STATUS WHICH SHOULD BE ACCORDED TO "OFFICIAL" STATISTICAL OFFICES

Bearing in mind the importance of good statistical information, it is desirable for an office of official statistics to be "independent". By "independent" we mean that decisions concerning the statistics programme are not

influenced by (party) political preferences. Among other things, that means that there must be no possibility of suppressing or delaying the availability of information that is politically unwelcome. (Conversely, there must be safeguards to prevent the office concerned taking a completely unilateral decision not to compile certain information that may be wanted by others.) Where a certain programme is being carried out by the office, it must be impossible for any outside organization to influence the statistical results. In other words, "adjustment" of the results, by organizations outside the office must be prevented. In this context, a description of the status of the Nederlands Centraal Bureau voor de Statistiek (CBS)(Netherlands Central Bureau of Statistics) can serve as an example.

The cornerstone of administrative statistical policy in the Netherlands is the Royal Decree of 9 January 1899 (Official Journal 43), also known as the Decree establishing the Central Bureau, and the Central Commission for Statistics (CCS).

The first paragraph of article 2 of the aforesaid Royal Decree reads: "The Central Bureau of Statistics collects, processes and publishes, insofar as available resources allow, those statistical assignments considered by the Director General to be of practical or scientific value".

The wording makes it clear that statistical policy is, in the first place, in the hands of the Director General of the CBS. His decisions are, however, subject to constraints and he is accountable to higher authority.

The first constraint is clear: the available resources. The Minister of Economic Affairs is politically responsible for the CBS and the CCS. He influences statistical policy in the first place, - subject to approval by Parliament - by making resources available.

The second constraint on the decisions of the Director General is to be found in the second paragraph of article 2. This runs: "The Bureau will undertake new statistical research or assignments, and bring research or assignments to an end, only with the approval of the Central Commission for Statistics".

This authorization system does, of course, give the Central Commission for Statistics very considerable influence on statistical policy.

Like every public servant, the Director General is accountable to his Minister. As part of such accountability, nevertheless, he can invoke his dependence on the authorization policy of the CCS. Similarly, the accountability of the Minister to Parliament is tempered by the authorization policy that the CCS is empowered to define.

The third constraint on the decision of the Director General follows from the third paragraph of the aforesaid article 2. This reads: "The Central Commission, acting either on its own initiative or on behalf of our Minister of Economic Affairs, instructs the Director General of the Bureau to collect, process and publish statistical assignments. He is bound to accede to this instruction unless, as a result of an objection to an assignment decided on by the Commission itself, he appeals to our Minister."

The aforesaid constraint is a policy tool for the Minister and a right of the CCS - though subject to the final responsibility of the Minister of Economic Affairs. Article 2, in its entirety, is a balance structure in which the government and parliament always form the keystone, notwithstanding the responsibilities assigned to the CCS and to the Director General of the CBS.

The independent position described above is one prerequisite for obtaining good statistical information. Such a position can guarantee two desirable effects: everyone in the public service is entitled to assume that the statistical information is not manipulated in any way and that all work is "for the general good". One consequence of this may be that a greater willingness on the part of the individual entities to provide information exists or can be promoted than would be the case in the absence of independence. In addition, independence is a virtually essential prerequisite if the administrative immunity discussed later on is to be credible.

3. ADMINISTRATIVE IMMUNITY

Administrative immunity is taken to mean that information supplied by an individual (for statistical purposes) will not lead to any administrative consequences. The aim here is to ensure that the respondent can be "relaxed" in supplying the information. Particularly where the information concerned involves (planned) action, attitudes or relationships which could be linked to administrative consequences, the guarantee of administrative immunity can result in more accurate information at the level of the individual than would be the case were it to be collected as administrative information. We approach this subject with some caution. In the case of statistics about personal or household incomes, it is often better to use the appropriate administrative information. In the Netherlands at least, the information concerned is so complex that the respondent often has to take a great deal of trouble over it and to consult his or her own bookkeeping in order to provide correct data. This trouble is taken for a tax

return - because the persons' own interests are involved and, in principle, checks by tax authorities can be expected - but probably not for statistical research on a voluntary basis. On the other hand, income (or parts of same) that are not revealed to, or discovered by, tax authorities may well emerge through a specific statistical survey where administrative immunity is granted. It will be clear that each case must be studied on its merits to see whether the procurement of statistical information requires direct surveys or the use of administrative information.

In this respect, it should be noted that where administrative information is used by the "official office of statistics", administrative immunity cannot and will not be granted by the body first receiving individual information. The office of statistics must, however, use the data so obtained exclusively for statistical purposes. This means, for example, that if the statistical office finds inconsistencies in the original data provided by the individual, or discovers on the basis of information from another source that in some cases the individual information is incorrect, no statement to this effect will be made to the body originally receiving the information. Should the statistical office receive information at individual level from two or more administrative organizations and be able to link together information about the same individual entity, this linked information may not be "fed back" to the administrative organizations involved. It is clear that this - explicit - methodology can give rise to, or strengthen, a feeling of trust in the statistical office. With reference to our earlier comments about the desirable "independent" status of an "official office of statistics", it is now more evident that this status can help to promote public confidence in the administrative immunity granted.

We consider it very important that this "independent" position be stressed and a complete detachment from administrative processes be maintained. The present development of society makes it increasingly possible for data bases to be automated and thus mutually connected. Partly as a result of this, individual entities adopt a reserved attitude towards the provision of information - a reservation that expresses itself as resistance. "Official offices of statistics" should, in our opinion, project an image and adopt methodologies of such a kind as to make any reservation or resistance with respect to these particular offices quite unnecessary.

Within this context, census experience is very important. Either legislatively or in practice, there is a link between pure statistical work and assistance to administrative organizations. After all, a census is often used to check the population register(s). If this is based on appropriate legislation, it is clear to everyone that administrative immunity is non-existent. Provided that this does not

endanger the quality of the statistical information (in our opinion the main aim of a census), no objection can, in principle, be made. It has been found, however, that this intermingling of functions does indeed endanger statistical information. Although it is not possible to guarantee that personal reservations or resistance with respect to a census would decrease if this intermingling of the two functions were abolished, we believe that in the present circumstances it is important to move in that direction wherever this has not yet occurred.

4. THE CONCEPT OF "POTENTIALLY CENSURABLE OR VULNERABLE ENTITY"

Earlier on, a distinction was drawn between administrative and statistical information - both terms then being explained. We also hope that the concept of "administrative immunity" is now clearly understood. We now come to a concept that is much more difficult to describe, and probably even more difficult to apply - that of "potentially censurable (or vulnerable) entity". To some extent, this concept is an extension of "administrative immunity". In the latter case, the important thing was that the individual entities supplying information for statistical purposes should suffer no direct consequences as a result.

Sooner or later, statistical information can have consequences either for all or a subgroup of the respondents or for a "third party". Since statistical information generally plays a part in governmental policy, many examples can be given of indirect consequences which can be linked to information supplied by respondents. In itself, we do not see that this gives rise to any great problems. In a democratic system, statistical information that has been obtained and supplied in an independent way is extremely valuable. Those individual citizens and directors of companies and organizations who recognize this value will - at least that is our hope - be prepared to supply individual information which, once transformed into statistical information, can result in policy measures being taken by the government.

It is, however, not impossible that certain groups will cultivate a mood of opposition to particular statistical research because of their general desire to deprive the government of the instruments it needs to carry out its policies. The well-known resistance to censuses, even where these have a purely statistical purpose, can to some extent be seen as resistance to a government whose activities permeate many aspects of the daily life of citizens. This

phenomenon is part of a problem which concerns the entire structure of the State. Official offices of statistics can do very little to alter this situation, but in making that comment we have assumed that the census has a purely statistical purpose. If that is not the case, it is advisable to take action which will sever any link between statistical and administrative information.

The situation is rather different where certain groups oppose statistical surveys, or attempt in other ways to cause them to fail, because they feel that the statistical information can give rise to measures aimed particularly against them. For the sake of example, let us just assume that statistical research - where administrative immunity has been granted - makes it possible to identify a group particularly active in tax evasion. Such information can give rise to two, quite different, types of measures. One of these is a change in policy to make it more difficult or impossible for the group concerned to evade taxation. The other type is a stricter policy of detection and prosecution which focuses to a greater extent on the group concerned.

The possibility of this latter type of measure is one of the reasons for our comment that the concept of "potentially censurable entity" can be an extension of the concept of "administrative immunity". After all, although no direct consequences result from the statistical information supplied, a certain group nevertheless becomes the target of administrators government officials. In such cases, we feel that the expression "potentially censurable entity" is appropriate.

To a certain extent, a comparable situation would exist if statistical research were to reveal that certain (small) subgroups were especially liable to be involved in particular - for example criminal - actions. The subgroup can be an ethnic minority or a group restricted to a particular area or suburb. A very real danger then exists that the statistical information concerned may result in discriminatory treatment based on the mere fact that someone belongs to that subgroup. In this case, too, statistical research gives rise to a potentially censurable entity.

At the beginning of this paragraph, we also noted the possibility of consequences for or measures directed against a "third party". To illustrate this, we would like to present three examples.

The first example involves the possibility of statistical research results proving that the policy adopted by a democratically-controlled organization (government, local council) has undesirable consequences. This organization, then, is our "third party". In this case, we should not

speak of a potentially censurable entity. After all, we believe that statistical information must be usable for determining and monitoring policy.

As a second example we can take a statistical survey of examination results in science subjects. If the statistics are split up to show examination results achieved by the students of different establishments, a "third party" is created since, although the examination results relate to individual students, the results achieved by economics students at the University of "X" constitute data about the Economic Sciences faculty at that university. The statistics could make it possible to compare data concerning the Economics faculties at different universities. In principle, this then creates a "potentially censurable entity". After all, inferior results at one faculty compared to another can be taken to suggest that teaching standards are lower there, or something similar.

A third example can be drawn from statistics on causes of death. If such statistics are split up into small regions, differences between these regions can become apparent. This can give rise to the formulation of (scientific) hypotheses or the testing of previously formulated ones. Let us now assume that in one particular respect a certain region compares very badly and that only one doctor is working in that region. In this case, this doctor is our "third party" and a potentially censurable entity.

In general, we believe it is necessary to make a distinction between a situation in which all - or at least certain - of the respondents can suffer adverse consequences from statistical information and a situation in which a "third party" exists. In the first case, the concept of "potentially censurable entity" can be seen as an extension of administrative immunity. In the latter case, the concept is more individual.

In addition, we have also shown by examples that some cases do not necessarily involve a potentially censurable entity, while others do. We hope that our treatment of the examples has made it clear that there are no black and white situations. In fact, it is a case of drawing a dividing line somewhere on a continuum.

The question now arises as to what operational value should be allocated to the concept of a potentially censurable

entity. We believe that whenever the question of a potentially censurable entity can arise, it is necessary to be very reserved in supplying statistical information.

In contrast to an individual scientific researcher, or a purely scientific research team, an official office of statistics has to exercise restraint in the publication of research results. Such an office is usually part of central government, but should have a separate status within the government structure and take great care to maintain that status. Other government organs exist for the preparation, execution and supervision of policy, and others still for carrying out administrative, governmental, investigative and prosecutory tasks. Above all, an official office of statistics must ensure that confidence in its independence and integrity is maintained. It provides statistical information on behalf of all echelons of society, and can do so only if it avoids any appearance whatever of participation in the administrative and governmental system. This may mean that certain detailed statistical information cannot be made available. Feelings of frustration may result because from a coldly scientific viewpoint this information may be considered valuable or useful in terms of measures to be taken. This code of conduct must not be based on more or less arbitrary principles, and in order to avoid this danger it must be possible to discuss the policy adopted by the official office of statistics by means of concrete situations, for example within the framework of a statistical commission.

5. CONCLUSION

This paper presents and explains a number of concepts. Statements are also made about the desirable status of an official office of statistics and about the policy that such an office should pursue. To some extent, this status and policy may be motivated by pragmatic considerations. By this we mean that a particular status and policy can help to improve the statistical information obtained, by promoting confidence in the office. The very choice of a pragmatic argument is an ethical process, and we make no secret of the fact that our statements are also motivated by outright ethical considerations. The hypothetical example drawn from statistics on causes of death, whereby a certain doctor could be named in relation with unfavourable "results", is relevant in this context. Let us assume for a moment that the doctor concerned did, in fact, deserve to be denounced, as his poor professional standards could result in the death of a number of people. Is concealment of such statistical information justified? Although we would find the decision very difficult if the example did occur in reality, we would

still be inclined to answer "yes". If we adopted any other attitude, there would be nothing whatever to stop the existence of other potentially censurable entities being revealed, or being notified through professional channels to detection systems.

We hope that the concepts we have described, and the views expressed, will provide food for discussion.

*

*

*

DISCUSSION

Mr. REDFERN *opened the discussion:*

He addressed himself to three main points out of the many interesting and provocative points in the Dutch paper. He agreed that the independence of the statistical organization from the administrative departments and from central government in general, as was the situation in the Netherlands, Denmark, and several other countries, was highly desirable in order (1) to maintain the confidentiality of personal data and (2) to ensure a fair presentation of the statistical results. In the United Kingdom the statistical service had always been firmly a part of central government but the two objectives mentioned had been assured to a large extent by the statisticians' adherence to their professional ethics and traditions.

He believed that the method of record linkage would be increasingly used to provide statistics in the future and this method required that the organization carrying out the linkage, was and was seen to be, independent to administrative departments. He supported the views in the paper in being firmly against censuses and surveys carried out for statistical purposes being used for the up-dating of population registers. However, he disagreed with the final theme of the paper, that the publication of information about any "potentially censurable entity" should be restricted. He believed that the principles of confidentiality should be applied to individuals but not to groups. The main aim of the census of population was to distinguish the characteristics of one group from those of another and this almost inevitably resulted in the possibility of administrative action which would favour one group to the detriment of another.

Mr. EMBLETON (*Central Statistics Office, Dublin*) stated:

References to general opposition to the census of population should not be interpreted as implying total opposition. There were variations between countries but in general the opposition came from a well organized and articulate minority who managed to receive extensive media coverage. There was a danger that strict adherence to the concept of "potentially censurable or vulnerable entity" would result in a position where a decreasing proportion of information collected at official inquiries was published.

Mr. TØNDER (*Central Bureau of Statistics, Oslo*) intervened:

He expressed surprise at the idea of "potentially censurable or vulnerable entity". He did not think that if a group of persons felt that statistical information could give rise to measures aimed particularly against them, this constituted a problem of privacy in the sense of the discussion of the seminar. He thought that if the statistics showed that a policy adopted by a democratically controlled organization was bad, these statistics should be published. The only case in which statistics might be withheld was where there were doubts about their quality of accuracy.

Mr. BEGUE (*Institut de la Statistique et des Etudes Economiques*) remarked:

Statistics almost always showed up differences between groups. For example they showed the differences in wages between men and women, leading to policies reducing the advantages enjoyed by men. Other examples could be given, dealing with grouping by nationality, by occupation, by age. He thought it would not be possible to set objective limits for the application of the principle of "potentially censurable or vulnerable entity" and in the extreme case it would not be possible to publish any statistics at all.

Mr. BUERGIN (*Statistisches Bundesamt*) added:

He agreed with Mr. Tønder that statistical data should not be withheld provided the conditions of accuracy and anonymity were met and data-protection rules complied with. He quoted the example of "cause of death" statistics which were important to identify regions with particular health problems. The statistics were grouped so that individual cases did not appear, before being passed on for scientific study.

Mr. HARRIS (*EUROSTAT*) noted:

The comments on the paper had been mainly directed to the subject of "potentially censurable or vulnerable entity" and considered that Professor Begeer had done a great service in drawing attention to this very real problem. In the United Kingdom it had particular reference to the laws on racism. There might be strong reasons for not publishing figures which showed a certain racial group in a particularly unfavourable light.

Mr. BRUENGGER (*Bundesamt für Statistik, Bern*) stated:

Switzerland had viewed the developments on the German 1983 census with great concern, as there had been risk of a protest movement also building up in Switzerland. Referring to Section 3 of the paper he agreed with the German position that statisticians should not be given a "blank cheque" as regards data linkage; each use of a data register for statistical purposes or record linkage should have specific legal authorization.

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

In reference to the notion of "potentially censurable or vulnerable entity" which brought into question the whole subject of statistics as a science, all statistics could be interpreted incorrectly. Statisticians had a responsibility to present the facts fairly and accurately but it was not their responsibility if the figures were subsequently manipulated for some political purpose.

The Chairman intervened:

He asked what would be the position if a government minister asked for information concerning individuals.

The author replied to the discussion:

He firstly answered the Chairman's question. The minister would not get information on individuals; no one would get such information. The CBS had complete autonomy in ensuring that data on individuals remained confidential.

He was pleased that the paper had provoked such a lively discussion and remarked that he was not inventing problems but merely posing them. His experience as Director General of the CBS had made him very familiar with the practical

problems arising and it was simply not true that one could publish everything, provided only that it was accurate and that nothing confidential was divulged. The CBS was required to publish useful information and his task as Director General was to judge what was useful. He would be very pleased if a conference could define what information should be published. In practice, however, he had to make rules and set limits and try to observe them in a consistent fashion.

With regard to Mr. Embleton's remark that only minority groups were opposed to the census, he thought this might be so initially but these groups were effective in creating a climate of opinion which turned politicians and the silent majority against the census. At all events a pilot survey revealed that a non-response of 40 percent was to be expected at a census in 1985 and the Dutch Government decided in the circumstances not to proceed with the inquiry.

THE N.I.S. AND PROTECTION OF PERSONAL PRIVACY

E. VAN LANGENDONCK
Institut National de Statistique

1. INTRODUCTION

Nature protection, protection of the environment, protection of privacy - phrases which are very much in vogue these days and are inextricably linked to Man's identity and quality of life.

Science and technology have advanced to a level where personal privacy is threatened and there are demands for protective measures. The more discussion, justified or not, there is about protection, the more we see, or think we see, our privacy eroded or threatened.

But there is a very real danger of taking things too far. Privacy might very well find itself ultimately the victim of excessive "protection measures", measures which impinge heavily on the very freedom we so much value.

The Nationaal Instituut voor de Statistiek (National Institute for Statistics - NIS) in Belgium, whose work involves the collection and recording of personal data, is certainly no idle bystander in this field: it is closely involved in all these problems and we therefore propose to examine the present situation in Belgium.

2. THE BELGIAN NATIONAL INSTITUTE FOR STATISTICS

The Royal Decree of 7 August 1939 provided for the centralization of all statistical services in Belgium, and part of the work assigned to the National Institute for Statistics was to carry out the major periodical censuses and surveys and compile all national statistics. Surveys and information recording activities which form part of the everyday work of other administrative departments and cannot, practically, be separated from that work would continue to be carried out by those departments.

The responsibilities of the NIS were consolidated, so to speak, by the Law of 4 July 1962, which empowered the government to carry out statistical and other surveys on the country's demographic, economic and social situation and constitutes the special legal basis for the surveys carried out by the NIS. This Law provides that the surveys required under its provisions must be carried out by the NIS.

However, as we will see below, the Law also makes an exception - as did the earlier Royal Decree of 7 August 1939, whereby surveys can be delegated to other government departments.

3. STATISTICS ACT OF 4 JULY 1962

The Statistics Act of 4 July 1962 distinguishes two principal types of survey: statistical surveys carried out for purely documentary purposes (Chapter I, Articles 1 to 4) and surveys carried out for administrative purposes (Chapter II, Articles 5 to 8).

The purpose of the documentary statistical surveys is to collect personal data which may be used only by the National Institute for Statistics and only for drawing up anonymous general statistics. Personal data are thus collected on a strictly confidential basis and can be used only for this purpose.

All surveys carried out by the NIS should theoretically fall into this category. In fact, however, the Institute carried out certain surveys, in particular to monitor economic activity, while other departments dealing with economic questions, such as the Ministries of Economic Affairs, Agriculture and Transport, also need personal data of the kind collected by the NIS for statistical purposes, either in order to carry out their work with a view to possible future administrative regulations or for drawing up or implementing regulations.

To avoid duplication, therefore, the Statistics Act of 4 July 1962 enables the NIS to carry out administrative surveys as well: when personal data are indispensable for preparing, drawing up or implementing an administrative regulation, the Crown can require special surveys to be carried out so that the data can be made available to designated government departments, with the exception of the tax offices.

The NIS draws up statistics using these data, and the personal data can be transmitted to the designated government departments.

In theory, these special surveys carried out for administrative purposes should also be carried out by the NIS, but the Law, as we saw above, makes an exception if the

survey is part of the department's everyday administrative work and is inextricably linked to it because of its special purpose.

The majority of the special surveys which come under Article 5 - or rather, all with occasional exceptions - are carried out by the NIS, which makes the relevant personal data available to the designated department as well as using them to draw up its own statistics.

4. DIFFICULTIES OF OBTAINING INFORMATION

The recession and budgetary restrictions have also left their mark on the NIS's activities. There is a growing feeling, even among the administrators, that some of the statistics being compiled no longer serve any useful purpose and are of no value, that too many details are being asked for, that some of the information being collected is already available elsewhere, that more and better use should be made of the existing administrative files. The national register of natural persons would solve a number of problems and simplify the work of the government departments, by acting as a focal point for information.

The Institute has on several occasions tried to cut down the amount of information it requests and to simplify the questionnaires. However, nothing has come of any of the discussions. In fact, the Institute has had its work cut out to avoid enlarging the questionnaires. Proposals to abolish this or that set of statistics have always been met by opposition from interested parties who wanted it retained.

Businesses, the first victims of the recession, are trying to cut back wherever they can. Often, in their view, statistics are of no value and certainly of no direct benefit to industry, with the result that they pay increasingly less attention to statistical returns and answer questionnaires less and less carefully and less and less promptly.

Many ordinary people do not know the first thing about statistics, let alone believe that they serve any useful purpose, coming to regard the requirement to give personal information increasingly as an infringement of their privacy and freedom. Opposition to giving certain information is growing, whether in the form of avoidance or of plain refusal.

Automation and the existence of data banks have only aggravated concern about the erosion of personal privacy, an accusing finger is pointed at the computer, which can consolidate data from different data files and in no time at all build a complete picture of a person or business.

The cries of alarm are getting louder in Belgium too and they have not been without result.

Apart from the Statistics Act of 4 July 1962, which, as we have seen, provides an adequate guarantee for the protection of the individual, there are two other documents worth noting in the field of statistics: the Law of 8 August 1983 concerning a national register of natural persons, which, as we will see later, includes a number of measures governing the use of stored data, and the Bill for the protection of certain aspects of personal privacy, submitted to Parliament in November 1983.

We will try below to show how these two texts will or could affect the work of the NIS and what the Institute's position is regarding these questions.

5. LAW OF 8 AUGUST 1983 CONCERNING A NATIONAL REGISTER OF NATURAL PERSONS

The Law of 8 August 1983 lays down the conditions for a national register of natural persons.

The national register is a computerized system for the input, storage and transmission of information concerning the identity of natural persons.

The purpose of the national register is not to centralize information about the population but to safeguard the transmission or retrieval of certain basic data specified in the Law. The following data are contained in the register: name and forenames, date and place of birth, sex, nationality, main place of residence, date and place of death, profession, civil status, composition of family.

It includes persons registered in the municipal residents' or aliens' registers and in the registers held at diplomatic missions or consulates abroad.

When a person is first entered in the national register, he is given an eleven-figure identification number. The identification number, which is also entered in the municipal residents' register, simplifies data therefore between the various government departments and organizations of public benefit authorized to use this identification number and exchange data of a personal nature.

Access to the national register can be extended by Royal Decree to the public authorities (national, district, municipal-provincial and local), the institutions of public benefit referred to in the Law of 16 March 1954 (e.g. Social Security Service, National Employment Office), notaries and sheriff's officers, in respect of information which they are authorized to have by virtue of a Law or Decree.

Access to the register can be extended to certain institutions under Belgian law which perform functions in the public interest (e.g. mutual insurance societies) after a decision by the Council of Ministers following consultation with the Advisory Committee for the Protection of Privacy set up by the Law of 8 August 1983. The composition and rules of procedure of the Committee are governed by Royal Decree of 20 April 1984.

The identification number may not be used without authorization or for any purpose other than that authorized. This is to prevent private institutions requesting individuals to tell them their identification number.

The identification number may only be used after consultation with the Committee and authorization by a Royal Decree discussed by the Council of Ministers. The Committee's opinion is published together with the Royal Decree in the Belgian Official Gazette. The Royal Decree lays down the purposes and limits of its use.

It has not yet been decided precisely what it will be possible to use the identification numbers for, but they should at least be able to be used in dealings with the national register and with the person concerned or with persons legally obliged to supply information about that person, such as an employer in his dealing with the National Social Security Service. Measures must also be taken to ensure that the number cannot be used to collate data illegally.

The job of the Advisory Committee for the Protection of Privacy is, either on its own initiative or at the request of the Minister of Justice, to give its opinion on all questions involving the protection of privacy in connection with the Law, with due regard to developments in the use of automated information management technology. The Committee is authorized to carry out *in situ* checks and also looks into complaints made to it about the implementation of this Law.

What does this mean in practical terms for the NIS?

As we have seen, the King can grant access to the national register to the public authorities for information they are authorized to have by virtue of a Decree or Law.

The first stage in the implementation of this provision, after obtaining the favourable opinion of the Council of State, was to submit for the King's approval a draft Decree giving NIS officials to be designated by the Minister for Economic Affairs access to information in the national register. However, this does not include the identification number, without which the NIS cannot draw up certain anonymous general statistics.

As we saw above, the use of identification numbers cannot be allowed without the approval of the Advisory Committee for the Protection of Privacy. To obtain this, the Institute will have to explain why it needs the numbers, what it intends to do with them and what links will be established with existing administrative data files for the purpose of drawing up anonymous general statistics.

The collation of personal data, even for statistical purposes, is one of the most sensitive areas in the use of personal information, especially in the eyes of the general public. However, the collation of data is in fact in the general interest.

The NIS will therefore have to produce a lengthy document carefully setting out its programme and its procedures.

The draft Royal Decree authorizing the use of identification numbers and determining the purposes and limits of this use, must then be submitted for consideration by the Council of Ministers.

However, it is not enough for the NIS simply to have access to the identification numbers of the national register: other departments and services which can use these numbers must be authorized to transmit them together with other data to the NIS to enable the Institute to make certain combinations for the sole purpose of drawing up anonymous general statistics.

The NIS is confident that its aims will be realized.

6. BILL FOR THE PROTECTION OF CERTAIN ASPECTS OF PERSONAL PRIVACY

On 10 November 1983, a Bill for the protection of certain aspects of personal privacy was submitted to Parliament by the Minister of Justice.

One of the aspects dealt with in this Bill (Chapter II) is protection against the erosion of the personal privacy of natural persons caused by the use of automatic processing of personal data by the government or private sectors.

The automatic processing of personal data means any set of operations carried out in whole or in part by automated means which involves the inputting, storage, alteration, erasure, retrieval or dissemination of personal data, and the carrying out of logical or arithmetical operations on those data.

Personal data are considered to be data relating to an identified or identifiable natural person.

Persons asked to provide personal data must be informed of the compulsory or non-compulsory nature of the answers, the possible consequences of a refusal to supply all or part of the information requested, the purpose for which the data will be used, and the persons or categories of persons authorized to receive the data.

The person concerned has the right of access to the data and can demand amendments. This right does not exist, however, in the case of automated data files of which the sole purpose is the compilation and dissemination of anonymous general statistics, because these data do not constitute a threat to personal privacy.

Certain data can have such an effect on a person's life that processing is either forbidden or subject to strict control.

The computerized processing of personal data relating directly or indirectly to race or ethnic origin, sexual behaviour, political, ideological or religious beliefs or activities and membership of a trade union or friendly society is in principle forbidden.

Persons in charge of computerized files of personal data must keep records showing for each processing operation the nature of the data, the purpose, the interconnections involved and the persons to whom the personal data will be given, and they must supervise the updating, alteration or erasure of data.

Before computerized files of personal data are opened, a report must be submitted to the Ministry of Justice giving all the information necessary for the identification of all the computerized files and the subsequent monitoring of their operation.

The Ministry of Justice keeps a register of computerized files of personal data: it is accessible to the public so that anyone can find out about the existence of a file, its purpose, the data processed, the conditions of use and conditions of access.

Interconnections, links or any other method of collating data or computerized files of personal data, to safeguard personal privacy, can be forbidden or regulated by means of a Royal Decree discussed by the Council of Ministers after the opinion of the Advisory Committee for the Protection of Privacy has been heard.

The transmission in any form between Belgian territory and another country of personal data which have been processed automatically or compiled for such processing can for the same reasons also be forbidden, made subject to prior authorization or otherwise regulated, in accordance with a Royal Decree discussed by the Council of Ministers after the opinion of the Advisory Committee has been heard.

The King can thus make use of the authority vested in him, within the limits of any international treaty which takes precedence over national law.

The job of the Advisory Committee for the Protection of Privacy - the Committee set up by the Law of 8 August 1983 on the national register - is, either on its own initiative or at the request of the government, parliament, the attorneys general or the courts, to give opinions on any matter relating to the application of this Law, with due regard *inter alia* to technological developments.

However, none of the provisions listed above - apart from the one relating to the Advisory Committee - are applicable to computerized files of personal data kept by international public-law institutions of which Belgium is a member, or to the computerized files of the National Institute for Statistics, except in the case of the special surveys referred to in Article 5 of the Statistics Act of 4 July 1962.

To carry out the work assigned to it by the Royal Decree of 7 August 1939, the Institute must have access to the greatest possible amount of personal data. It may use these data only for drawing up anonymous general statistics and the publication, dissemination or communication of personal data is forbidden. This restriction is laid down specifically in Chapter 1 of the Act of 4 July 1962 on surveys carried out for purely documentary purposes.

Although the NIS is a Directorate General of the Ministry of Economic Affairs, it regards any other Directorate General of this department and any other department or service as a third party. The fiscal legislation also provides that the NIS is beyond the tax authorities' powers of investigation.

The NIS thus provides the necessary guarantees for the use and confidentiality of personal data supplied by natural and legal persons, so that there is no danger of the erosion of personal privacy even when the data are processed by computer. The Bill therefore rightly provides that the Law should not be applicable to this.

The Law will, however - at least unless it is amended - apply to the special surveys referred to in Article 5 of the Statistics Act of 4 July 1962.

The only possible explanation as to why these should still be subject to the Law is that the personal data collected can be communicated to another ministerial department.

This means that persons supplying information for any survey based on Article 5 (mainly agricultural statistics and monthly statistics on industrial activity) must be made aware of the obligatory or non-obligatory nature of the answers, the consequences of a refusal to give information, the purposes for which the data are required and the categories of persons authorized to receive the information.

This is not such a problem for the NIS: what is more serious is the right of inspection of the person concerned, the fact that for each processing operation a document has to be drawn up stating the nature of the data processed, the purpose of the processing, the interconnections involved etc., that a report has to be submitted to the Ministry of Justice before computerized files of personal data can be opened, giving all the information necessary for the identification of all the computerized files and the subsequent monitoring of their operation.

As has been said earlier, in accordance with Article 5 of the Statistics Act, the NIS can carry out surveys to provide government departments expressly designated by the King in an implementing Decree with information about individuals which is indispensable for an administrative regulation. The Act also provides that the NIS may use this personal information to compile anonymous general statistics, as it can data collected under Article 1 (surveys intended solely for documentary purposes).

The reasons why the Law of 4 July 1962 made specific provision for the transmission of personal data to other government departments was to avoid duplication: otherwise, the departments concerned would be forced to carry out the same surveys as the NIS, which would mean unnecessary extra work for both the respondents and the administration.

There is no difference between the computerized processing by the NIS of data obtained under Article 5 and the processing of other data collected by the Institute. In the Institute's view the confidentiality of the former is adequately safeguarded and publication is forbidden, so that there is no danger of erosion of personal privacy through computerized processing.

Indeed, the government department which received personal data from the NIS in accordance with the Royal Decree providing for the survey will in any case be subject to the new Law if it processes the data automatically.

We might add that the data obtained under Article 5 of the Statistics Act are very often not data concerning individuals since they relate to legal and not natural persons.

It must be stressed that if the Bill becomes law in its present form, it will be a very serious obstacle to the Institute's work in view of the complex procedures proposed for authorization, transmission of data, right of inspection etc. Furthermore, it is not always possible to determine beforehand how the data collected will be processed or what statistics will be drawn up.

The existing laws and regulations, in particular the Act of 4 July 1962, provide the necessary safeguards for the use and confidentiality of the personal data collected - including those collected under Article 5 of the Act - so that the NIS feels that the fears of a threat to personal privacy are unfounded.

In view of all these factors, it is felt that the restrictions provided for in the Bill should be deleted, so that the NIS's computerized files of personal data will be totally excluded from the provisions of Chapter II.

However, this is not all.

Administrative, and above all fiscal, data coming from other authorities which collected them, are frequently used by the NIS, in accordance with the Royal Decree of 7 August 1939 on the centralization of statistics, frequently used by the NIS to compile very useful anonymous general statistics. This method avoids duplicating the collection of data and therefore saves money for the administration and for the individuals concerned.

The communication of data from other computerized files to the NIS for the purpose of drawing up anonymous general statistics, should also be excluded for the provisions of Chapter II.

Under the Bill as it stands, before transferring data to the NIS, the department which collected them would have to go through a series of formalities which will inevitably cost time and money. Furthermore, data can only be transferred if the persons from whom the data were collected have first been informed. Statistical processing of previously collected data which the NIS subsequently finds out about, or which it becomes necessary or desirable at some point to process to obtain general statistics, would therefore be prevented.

Excluding the communication to the NIS of data collected by other departments from the provisions of the law would therefore help the smooth operation of the statistical services.

This would mean that the persons concerned would not be informed that they are also included in the computerized file of the NIS, but this would be in no way damaging in their interests. It is only useful for a person to know that he is included in a computerized file in that this gives him the opportunity to be informed of the data held on him and correct any inaccuracies. Since the person involved is already entered in a computerized file of the department which originally collected the data, he had the opportunity of correcting any inaccuracies through that department, which would in turn have informed the NIS.

The Bill also provides, on the other hand, that the right of inspection cannot be invoked in the case of computerized files of personal data, the sole purpose of which is to compile and disseminate anonymous general statistics because, according to the Explanatory memorandum, these data do not constitute a threat to personal privacy.

We can therefore conclude that the formalities laid down in the Bill for the communication of data recorded in a computerized file are unnecessary for the protection of the interests of the persons as far as communication of data to the NIS is concerned, and unnecessary formalities are best avoided.

The NIS will also endeavour to keep the special surveys carried out under Article 5 of the Statistics Act from coming within the scope of the law and to ensure that the communication of data from other computerized files of personal data to the NIS for the purpose of drawing up anonymous general statistics is also excluded from the provisions.

At its meeting on 3 September 1984, the Hoge Raad voor de Statistiek, the government's advisory body on statistics, expressed the same opinion and urged the Minister of Economic Affairs to make every effort to get the proposed amendments accepted.

Finally, we would also like to point out that the NIS's approach to confidentiality is even stricter than that provided for in the Bill. As has been said, in the Bill personal data, i.e. data which require special protection, are considered to be data relating to an identified or identifiable natural person.

The NIS, on the other hand, considers individual data, i.e. data which must be confidential, to be any information relating to a natural or legal person, whether identifiable or not.

It has been suggested, in particular by the university research centres, that the NIS is too strict and its attitude a serious hindrance to scientific research. In their view, there can be no objection to the communication of anonymous data which cannot be linked to a physical or legal person and their confidentiality should only be safeguarded if identification is possible.

They also refer to the Strasbourg Convention on the protection of individuals with regard to automatic processing of personal data, signed by Belgium on 7 May 1982.

Researchers also take every opportunity of drawing attention to the principles adopted at the Bellagio Conference in 1978 on privacy, confidentiality and the use of official micro-data for research and statistical purposes.

The availability of aggregated statistics or special tables does not satisfy the needs of the researcher, and according to one of the principles adopted by this Conference national statistical offices must ensure that researchers working in the public interest have ample and practical access to data, within the bounds of the accepted concepts of personal privacy and the legal requirements for the protection of confidentiality.

If there is any change in the Statistics Act - and in view of the constitutional reform this has become necessary - these questions will undoubtedly arise and reference will be made to the law on the protection of privacy, to the Strasbourg Convention on the protection of individuals with regard to automatic processing of personal data and to the principles adopted by the Bellagio Conference in order to obtain, above all for the sake of the researchers, a more flexible attitude on the part of the NIS towards anonymous data with no means of identifying individuals.

7. CONCLUSION

As the official statistical organization, the NIS is a scientific body and must therefore take extreme care of the documentation it compiles and ensure the confidentiality of personal data.

To ensure the accuracy and honesty of replies, the Institute has always made every effort to allay the fears of the respondents that their data might be published or used for anything other than the authorized purposes.

The strongest card of the NIS, the official body for compiling general statistics, is the guarantee of the confidentiality of personal data, a guarantee which is logically, and quite rightly, the counterpart of the right to request the necessary information. This guarantee should make it possible for the NIS to be exempted from certain measures for the protection of the individual - measures which might be desirable, useful or necessary elsewhere but which, for the official statistical service, are unnecessary, indeed superfluous, obstacles which threaten the normal, smooth and efficient running of the Institute.

*

* *

DISCUSSION

Professor WHELAN (*Economic and Social Research Institute, Dublin*) opened the discussion:

In his own country, Ireland, legislation regarding privacy and data protection had not yet been enacted or even widely discussed. However, he expected that it would become an important issue in the years ahead. Mr. Van Langendonck's paper described recent initiatives regarding data protection in Belgium, another relatively small country. The paper was therefore of particular interest. This paper raised a number of important issues and prompted several questions. It seemed that the recent legislation in Belgium might seriously inhibit the work of the INS and that the Institute was pressing to have various amendments implemented. The major problems appeared to arise not in relation to documentary or "pure" statistical surveys but in relation to studies which fulfilled a dual role, i.e., administrative surveys or the processing for statistical purposes of administrative data. It would be interesting to know in what ways the data from administrative surveys which the INS conducted were used by other Departments to whom they were transmitted. Were they used by these Departments to produce

aggregate (though possibly detailed) statistical results, or were they used to administer various schemes (such as grants etc.) on a person-by-person basis?

In Ireland there was no compulsory population register - the nearest equivalent was the Electoral Register which listed all those eligible to vote. It had, on occasion, been suggested that there should be a National Population Register and so the author's comment on the Belgian Register and the regulations governing its use were of considerable interest. He asked (a) what statistics the INS hoped to derive from the Register and (b) what "existing administrative data files" it hoped to link with the Register.

Professor Whelan's next comment related to the proposed legislation for the protection of certain aspects of personal privacy. The "computerized processing of personal data relating directly or indirectly to race or ethnic origin, sexual behaviour, political, ideological or religious beliefs or activities and membership of a trade union or friendly society" was in principle forbidden. Did this mean that survey research on all these topics was now prohibited in Belgium? If not, how were the exceptions for *bona fide* survey research on these topics embodied in the Bill?

The author replied:

The census operation involved using two forms. One was destined for the commune and was used to update the population register. The other was sent to the INS and was used for compiling statistics. The agricultural census provided an example of a combined administrative and statistical inquiry. The INS used this to compile global and anonymous statistics. The Ministry of Agriculture used individual data from this inquiry for paying compensation for frost-damage, obligatory slaughter of cattle etc. The population register might be consulted by interested persons. It was not linked with any other register.

He then suggested that it might be useful to send a questionnaire to the national authorities asking questions on the following lines:

- i) How is the census of population carried out?
- ii) Are the individual data used for administrative as well as for statistical purposes?

- iii) May the statistical office pass on data for identifiable individuals?
- iv) What impact do laws designed to protect the confidentiality of individual data have on statistical work?
- v) How independent is the national statistical service from other branches of government?
- vi) What is the position as regards the keeping of census material as national archives?

There was some discussion about how such an approach should be managed. In any case, some of the information would come from a study commissioned by EUROSTAT and already in progress of census practices in member countries.

DATA PROTECTION ISSUES AS THEY AFFECT POPULATION CENSUSES
AND SOCIAL SURVEYS IN THE UNITED KINGDOM

R. BARNES

Office of Population Censuses & Surveys, London

SUMMARY

The collection of information in population censuses and social surveys frequently presents the dilemma how to preserve individuals' rights with regard to confidentiality of the information they have given, on the one hand but manage to exploit data for *bona fide* statistical and research purposes as fully as possible, on the other. To a large extent, these are conflicting aims and quite where the line will be drawn will vary from country to country. This paper describes the way the problem has been dealt with by those who carry out population censuses and government social surveys in the UK.

0. INTRODUCTION

Before dealing with the specific position of the United Kingdom, this paper addresses some general issues of data protection.

First, it seems useful to draw a distinction between privacy and confidentiality - terms which are sometimes used interchangeably and concepts which are often confused. Privacy is concerned with the intrusiveness of the information that is demanded or requested of people in the first place - and their rights in this regard. Confidentiality is about what is done with the information after it has been obtained. The two are closely linked, but the issues are separate; if people are uncertain about confidentiality, they may be more likely to resent and to resist what they see as an invasion of their privacy. This paper is concerned with confidentiality as that is what data protection is principally about.

The second general point is perhaps self-evident. In most countries, population censuses and social surveys can be carried out only if the population is broadly supportive of them, or at least not antagonistic towards them. One general pre-requisite of public acceptance is belief in the confidentiality of the information and the assurances and undertakings that are given. The postponement of the 1983 Census in West Germany and the abandonment of the 1981 Census in the Netherlands are ample demonstrations of what can happen when public support and confidence are lacking.

A third general point is the underlying question of what is the basis of the case for data protection - whether in law or not.

- i. The case may rest primarily on some ideological or ethical principle: that is, is there a need to preserve confidentiality because of some felt moral obligation to do so?
- ii. Or it may stem from a realization that there is genuine public concern with regard to the use or misuse of personal data - especially where data collected for statistics are used for administrative purposes and especially with the greatly extended use of computer systems?
- iii. Or it may arise, at least in part, primarily to defend the census or survey against those who seek, for whatever reason, to discredit it?

The answer to these kind of questions will vary from country to country and from census to census or from survey to survey; but it is important to make some assessment of the answer in order to determine what shape data protection should take, if any. For example, if the primary case for data protection is along the lines of i. above, then simply to safeguard confidentiality is probably sufficient. On the other hand, if the case is mainly because of real public anxieties (ii. above) then it is important not just that confidentiality should be preserved but also that steps should be taken to ensure that the public is persuaded that it is so preserved.

This leads to a fourth general point. If a major element in the case for data protection is genuine public anxiety, what is the reason for that anxiety? This may be real, rather than imagined: for example, that under some regimes data may be used for purposes not made clear at the time the data were given but which are to the detriment of the individual; for example, the US Government proposal in wartime to use the 1940 U.S. Census to trace persons of Japanese extraction for internment purposes. Or, the fears and concerns may be largely imaginary - although no less potent for that; for example, in the UK at the present time there is a debate about whether the next population census should contain a question on ethnic origin; the antagonists fear that it is conceivable that one day a new and malevolent regime may use that data for malign purposes, such as compulsory repatriation.

Whether their fears are real or imagined, people are likely to be more worried about data protection when they are uncertain about the uses to which data are put (and this may explain why information given willingly for one purpose,

e.g. income details given to tax authorities, is often withheld for another, i.e. income is generally considered unacceptable as a census topic in some countries - including the UK).

1. THE UK POSITION

In many ways, recent experience in the UK in moving towards data protection legislation provides a good case study of the issues involved and serves as an illustration of many of them.

Until 1984 there was very little general legislative protection of statistical data. What there was, for official data, comprised specific Acts of Parliament relating to particular statistical collections - e.g. the Agricultural Statistics Act of 1979 and the Statistics of Trade Act of 1947. In the field of population censuses and social surveys there has been no legislative provision governing the conduct of surveys although the census has been, and still is, subject to the provisions of the 1920 Census Act for Great-Britain and the 1969 Census Act for Northern Ireland. These Acts provide certain safeguards of the confidentiality of data and prescribe penalties for breaches of these safeguards by the census takers. Section 8(2) of the 1920 Census Act for Great-Britain is as follows.

"If any person -

- (a) being employed in taking a census, without lawful authority publishes or communicates to any person otherwise than in the ordinary course of such employment and information acquired by him in the course of his employment; or
- (b) having possession of any information which to his knowledge has been disclosed in contravention of this Act publishes or communicates that information to any other person;

he shall be guilty of a misdemeanour, and shall on conviction be liable to imprisonment with or without hard labour for a term not exceeding two years or to a fine, or to both such imprisonment and fine."

The Census Acts are closely and diligently observed in practice but even so their interpretation is sometimes in doubt - even today. For example, one clause of the 1920 Act says that the census authorities "shall be subject to the control of, and comply with any directions given by, the Minister of Health". The interpretation of this was challenged last year, in evidence given to a Parliamentary Committee on Race Relations and Immigration, on the ground that this clause could enable any Minister of Health to

demand to see individual census records and so undermine the confidentiality undertakings given in the census. It so happened that legal opinion found that, in fact, the Registrar General "would not be bound, nor even entitled, to comply with such a direction and if he did so he would be committing [an] offence." The point of this account in the present context is that it shows the limited extent to which data are clearly protected by law.

Until now the UK has relied for data protection, not so much on the law but much more on good practice and a strong instinct for professional self-preservation. That is, a sense that if we do not offer reasonable undertakings of confidentiality and, equally important, if we do not observe these (and be seen to observe them), the time will not be far off when public cooperation is simply withdrawn. In the UK we are heavily dependent upon public goodwill and cooperation, not only for success in government social surveys (which are all voluntary) but also for success in the population census (which is a statutory requirement).

Until the advent of general data protection legislation, a number of measures have been taken in the UK to protect data and to maintain "good practice". First, the census: the following is a list of the main measures taken to ensure the confidentiality and security of the data collected during the 1981 Population Census.

- i. All enumerators were required to sign an undertaking of confidentiality and were made aware of penalties for not observing this. They were referred to the Census Acts which prescribe penalties for unlawful disclosure, and the 1911 Official Secrets Act. These Acts govern all people involved in taking censuses.
- ii. People who wished to enter their own details were allowed to do so on a separate form and seal it in an envelope to be handed to the enumerator so that other household members could not see the details.
- iii. As far as practical, enumerators were not used in areas where they were likely to be widely known. It was not possible to guarantee that the enumerator would not know anybody in the area (except by

- incurring huge travel costs by employing enumerators in areas far away from their homes). In these cases a sealed return system was also used so that completed forms were not seen by enumerators.
- iv. A security firm was used to transport boxes of completed forms to processing offices. This was under the protection of a special contract to prevent the unauthorized use of information by the security firm or any of its employees.
 - v. In England and Wales two processing offices were used so that the forms from the area around one office were processed by the other office, and vice versa - so even the coders and data processors would not see the forms relating to their neighbours. All Scottish census forms were processed in Edinburgh but the work was organized so that processing staff did not see the forms for the area in which they lived.
 - vi. Staff at the processing centres were required to wear passes whenever on the premises. Visitors were admitted only under escort and were never allowed to see census forms.
 - vii. After use the forms are kept under special secure storage conditions for 100 years - with protection against unauthorized access and against damage (fire, water/damp, mites, etc.).
 - viii. Names and addresses were not fed into the computer. Computer arrangements and general confidentiality provisions were inspected by independent assessors, the British Computer Society, who reported on the arrangements and stated that

"The Society is pleased to report that the procedures for the conversion, processing and storage of census information, and the procedures for making output available to users, have been planned with care, and that particular emphasis has been placed on designing the system to make breaches of confidentiality extremely unlikely. Similarly, the standards of physical security at all locations visited have been set at a high level. Most importantly, all the staff have shown an understanding of the problems and an attitude towards the issues of public confidence which promises that a high standard of vigilance will be maintained at the actual operation of the 1981 census."

- ix. Statistical tables for each small area were blurred slightly by deliberate injection of small random errors so that information about identifiable individuals of households could not be accidentally identified from the tables: in small area statistics each cell was modified by +1, 0 or -1. Also, at small area level, if the area had less than 25 persons or 8 households present then most of the cells were suppressed. The smallest area for which census statistics were compiled and published was the enumeration district (ed) which comprised, on average, 180 households.
- x. Finally, all unpublished tables were scrutinized by senior staff before release to make sure that individual identities were not revealed accidentally.

With government surveys carried out by the Office of Population Censuses and Surveys (OPCS), reliance is also firmly placed upon the diligence and integrity of those who undertake the enquiries. In fact, so far, the record is very good indeed and in over 40 years of survey research it is hard to remember a single blemish. Moreover, in very few surveys are names and addresses recorded and, where they are, they are usually not transferred to computer file.

In 1984, the measures which have been described have been reinforced by two important new developments: the passing of the Data Protection Act (to enable the UK to ratify the European Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data) and the publication of a Code of Practice for the Government Statistical Service and the Handling of Data Obtained from Statistical Inquiries. Details of the Data Protection Act are given in another paper from the UK although it should be said here that there has not yet been sufficient time, since its enactment in July 1984, for the provisions of the Act to have had a major impact on data protection of censuses and surveys. The Code of Practice was published in June 1984 and the eight points it contains are reproduced as Appendix A to this paper.

Naturally a price has to be paid for the protection of data, and this is in terms of restrictions placed on the dissemination and use of data.

Census data for example are only released in the form of statistical aggregates for areas not smaller than an enumeration district (ed) and even then, as has been described, additional safeguards are imposed by the introduction of random error. We do not, in the UK, make available any census records of identified individuals to anybody outside the census offices - even to other central

government departments - and neither, so far, have we made available public use sample tapes, with anonymous individual data, as is done in a number of other countries. This means that the many users of census data have to accept the ed's, or some geocoded reference such as a grid square or a post code, as a minimum building block for a real aggregation. And, perhaps even more importantly, they have to accept the variables which are explicit or derived as the census offices determine them. It is of course possible to produce new derivations or, in theory, to aggregate from small building blocks but the scope for this kind of tailor made approach is severely limited by the resources of the census offices.

The question of whether or not the census authorities are able, under existing legislation, to make available public use sample tapes is not altogether clear. The 1920 Census Act is open to different interpretations on this but the point was covered by a government statement in a White Paper on the 1981 Census of Population: Confidentiality and Computing (Cmd 8201), published in 1981 which stated that

"Such a tape should have to conform to three criteria. First, the item of data to be recorded on the tape would have to be chosen and coded so as to protect confidentiality. Second, the uses of the tape would have to be sufficient to justify its costs. Third, statutory authority for the release of the tape would be required. ...Moreover new legislation might be needed because there is doubt whether the Census Act of 1920 provides statutory authority.

The Registrars General invite the British Computer Society or any other interested body or person to make proposals on the form of a public use tape that would serve a wide range of users whilst effectively protecting confidentiality."

So far no proposals have been put forward.

The price paid with regard to sample survey data is in many ways similar, although in this case, because the data are not collected under statute, we do provide computer tapes (by definition sample tapes) containing individual records - but not names or addresses. This we do either direct or

through a Data Archive which is attached to one of the universities. But one of the conditions of release of these tapes is that, before they leave OPCS, the data on them are "anonymized" by broad banding certain variables such as occupation and, especially, the area codes. The point of the broad banding is a recognition of the fact that, in principle, if very different in practice, it may be possible to identify an individual from a unique combination of characteristics.

Names and addresses of sub-samples of persons from population censuses and government surveys are sometimes used as a basis for follow-up enquiries provided these enquiries are carried out by OPCS or their agents. Even then there is a danger of public misunderstanding and disquiet on the question of confidentiality. For example the names and addresses of a sample of persons identified in the 1971 Census as having nursing qualifications but not employed as nurses currently, were passed to the Social Survey Division of OPCS for a follow-up study. The fact that the survey organization and the census organization were both parts of the same Office was not always understood by persons outside and allegations of breaches of the confidentiality undertakings given at the time of the 1971 Census, continued to be made from time to time long after the event. The same case was also cited, although for different reasons, in the report of the Data Protection Committee set up in July 1976 under Sir Norman Lindop. That report, published in December 1978 (Cmnd 7341), contained the following passage.

"Where data are collected for statistical purposes the information is sometimes also used for the construction of registers for future statistical enquiries...follow-up surveys, if unexpected by the data subject, can cause understandable and needless alarm. This is illustrated by the concern which was expressed when a number of women who had disclosed in the 1971 Census that they were former nurses were subsequently...asked to give further information about their reasons for leaving the profession. ...In our view, the fact that information will be used to construct registers for further enquiries would normally be made clear at the time the information is collected."

In fact by the time of the 1981 Census it was decided that no follow-up enquiries of the census should take place unless these were approved at the time the main census arrangements were approved by Parliament.

So far, all that has been said in this paper with regard to the release of individual records on surveys has related to coded and broad banded data on computer files. The actual documents - survey questionnaires and census forms - are not released outside the Office at the time they are collected or for a long time afterwards. What happens is that all census forms and a few survey forms are kept under secure conditions for 100 years (census forms) and 75 years (survey forms) at which time they are transferred to the Public Record Office, under the terms of the Public Records Acts 1958 and 1967, from where they are made available for public inspection. Most survey forms are not kept but are destroyed after about five years. The reason for keeping forms is for genealogical/medical research purposes (census forms) and for general historical and sociological research (census and survey forms) since the original forms may contain more detail on, for example, people's occupations or attitudes to various things, than was coded and stored on computer file. They are part of the heritage of the nation although it must be conceded that nevertheless making the forms available to the public at all could be construed as a technical contravention of the confidentiality pledges since these, at the time they were given, were unconditional with regard to time. However, so far, this matter of release after 100 years has not sparked any controversy - although the mere fact of retention at all has (and there is nearly always a considerably lobby which presses) in the interest of data protection, for census forms to be destroyed. For example, the lobby backing the inclusion of a question on ethnic origin in the next population census, headed by a Parliamentary Committee, is especially forthright in its view that forms should be destroyed after the next census.

2. THE MAIN ISSUES

Data protection legislation is now an accomplished fact in the UK, as in many other countries. However, a number of issues relating to confidentiality of census and survey data still remains. Some of the main ones seem to be as follows.

- (i) Is formal data protection legislation sufficient to secure public trust?
- (ii) Should the legislation relate only to computerized records and should individuals be allowed to inspect and correct where necessary, the data holdings which relate to them? Does it matter, for data protection purposes, whether statistical records are complete and accurate?

- (iii) Are there positive benefits to the collector and user of data for the data to be protected by legislation?
- (iv) Is there a legitimate case for public use sample tapes (or indeed public use tapes for a whole census), and if there is, is broad banding of variables justified or necessary? What constitutes identifiable information?
- (v) Should names and addresses be collected at all?
- (vi) What restrictions on use of data are tolerable in the cause of data protection?
- (vii) Is the introduction of random error in small area statistics justified in the cause of data protection? Does this destroy statistical integrity and corrupt truth in any important way?
- (viii) Should census forms and survey questionnaires be kept for historical purposes, or should they be destroyed?
- (ix) How much is it necessary to tell informants about the nature of the uses to which the data are to be put, bearing in mind that these uses may be so wide ranging (especially for censuses and multi-purpose surveys) as to be almost impossible to foretell and that, in some cases, total foreknowledge may so influence the responses as to be destructive to the aims of the enquiry?
- (x) What kind of publicity is required, and what proportion of the census or survey budget should be spent on this, to convince an often sceptical public that confidentiality is in fact maintained?

THE GOVERNMENT STATISTICAL SERVICE CODE OF PRACTICE
ON THE HANDLING OF DATA OBTAINED FROM STATISTICAL ENQUIRIES

1. These guidelines on the handling of data obtained from statistical inquiries by the Government Statistical Service apply equally to information obtained from compulsory (statutory) or voluntary statistical inquiries.
2. Statistical inquiries are identified as such to the respondent, either on the forms to be completed or by the interviewer. Respondents are informed when response to an inquiry is compulsory under statute.
3. Confidentiality during the collection, storing, transferring and processing of information about identifiable statistical units is safeguarded at all times. If such information is transferred:
 - the department that originally collects or, if an outside organization carries out the collection, sponsors the collection of information, satisfies itself with the safeguards applied to the data both by itself and by those to whom the data may be transferred; and
 - the transferring department obtains prior written assurance from the receiving department or outside organization that it will apply the safeguards in this code, and also accept any limitations that may be imposed concerning the use of further transfer of the information.
4. Information about identifiable statistical units is not used for other than statistical purposes or transferred to another department or outside organization for such purposes unless:
 - this is provided for by law and no undertakings have been given to the contrary; or
 - in voluntary inquiries either the respondent was so informed when the information was collected, or has subsequently given consent in writing.
5. Information about identifiable statistical units is not transferred to another department for statistical purposes, or to organizations and *bona fide* researchers outside government departments for statistical purposes, unless:

- this is provided for by law and no undertakings have been given to the contrary; or
- in voluntary inquiries either the respondent unit was so informed when the information was collected, or has subsequently given consent in writing; or
- prior written authorization has been given by a minister in the department in possession of the information, and the transfer is not forbidden by law, and no undertakings have been given to the contrary.

6. Where it is not forbidden by law and where no commitments have been entered into to the contrary, a department may transfer anonymous information about statistical units to another department or to organizations and *bona fide* researchers outside government departments.

7. All reasonable care is taken that information about identifiable statistical units, or tables or other analyses from which such information can be deduced, is not published or made generally available or otherwise transmitted, except as specified elsewhere in this code, unless the respondent concerned has given consent.

8. Where information about identifiable statistical units may be of interest to future historians and researchers a department may preserve the information in an archive responsible for the preservation of public records, such as, in England and Wales, the Public Record Office or such other place of deposit appointed by the Lord Chancellor; or in Scotland, the Scottish Record Office. Arrangements for the long-term preservation of such information in England and Wales are made in accordance with the Public Records Acts of 1958 and 1967, while in Scotland there are analogous non statutory arrangements. The preservation in such an archive of information about identifiable statistical units only takes place under the following conditions:

- if the information was collected under statute, only as provided for by statute;
- otherwise, only if the preservation is not forbidden by law or where undertakings have not been given to the contrary, and where the department, in the light of any appropriate consultation, no longer considers the information to be sensitive.

DISCUSSION

Dr. VAN HOESEL (*University of Leiden, Nederland*) opened the discussion:

He remarked that his institute in the University of Leiden was both a producer and a user of statistics. However, because most of the opinions expressed at the Seminar had been those of producers, he would look at the subject from the point of view of the user. He thought that the regulations applying to data files in the UK were extremely user-unfriendly. Most files of data collected for statistical purposes contained a large amount of information which would be of value to social scientists but existing regulations precluded all possibility of extracting it. This represented a considerable waste of public money. Although he would like to see them improved still further to permit direct access in certain circumstances, better facilities for users existed in the Netherlands.

He agreed that good data protection measures were also in the interest of users, since by increasing public confidence, they improved the quality of the statistics. However, too much emphasis on protection simply increased suspicion and was counter-productive. There were too many groups, public and private, surveying the UK population and the market was being spoiled as had been the North Sea by overfishing. He thought a code of conduct for data collection was needed, which might be administered by the central statistical service. He suggested the use of TV spots to introduce important surveys, covering the relevance, conduct and data production arrangement.

He was not in favour of destroying important data or part of the data such as names and addresses as the possibility should exist for the scientific community to check on the results produced by statisticians. Names and addresses were in any case needed for recurrent surveys.

In summary he proposed:

- (1) to give users the facility to squeeze all relevant information from data files
- (2) to restore the market of respondents by stopping over-exploitations
- (3) not to destroy data.

Mr. REDFERN *asked* whether one of the reasons for the passage of data protection legislation in the UK was to permit the control of trans-national data flows and to conform to the Council of Europe conventions on this point.

Mr. HARRIS *remarked* that many of the countries entering into data protection legislation had taken account of the Council of Europe guidelines or OECD codes. The same might be true of a code of conduct. He asked whether there had been any experimentation in the UK in the lines of asking respondents whether they objected to the information they had given being supplied, with identification removed, for other government or research purposes.

Mr. EMBLETON *asked* whether any requests had been made in the UK for the introduction of public-use samples. He asked also, whether census forms were released after 100 years and whether this did not comprise a technical breach of the confidentiality regulations.

Mr. BARNES, *replying to the debate*, said that they did try to perceive the needs of the users of statistics as well as those of the producers. The balance in the UK was in favour of the producer but he remarked that it was, perhaps, no coincidence that a very successful census had taken place in the UK in contrast to the experience of some other countries with more liberal arrangements for public-use files. The situation as regards public-use tapes was under continuous review.

On the subject of TV publicity, Mr. Barnes said that in the UK they preferred a low-key approach to surveys which were about to be launched and had found this the most successful over a long period.

On page 169 there was an account of an examination of processing and storage arrangements for the census carried out by the British Computer Society. This provided a valuable independent check of the type Dr. Van Hoesel had referred to.

In reply to Mr. Embleton he said that a meeting of the Economic and Social Research Council had, indeed, recently considered the question of public-use tapes and it remained to be seen whether a request for such tapes would emerge. As far as he knew the current Census Act, that of 1920, did not cover the point of releasing census forms in 100 years time. The issue had provoked no public discussion at the 1981 census.

They had not tried asking, in the course of a survey, whether the respondent objected to his information being passed on. The nearest they came to it had been to ask whether the respondent would object to being included in a later survey; about 85% did not object.

THE DEVELOPMENT OF DATA PROTECTION IN THE UK:
ITS IMPLICATIONS FOR OFFICIAL STATISTICS

T.J. ORCHARD
Central Statistical Office

SUMMARY

After twelve years of consultation and discussion the UK has legislation to control the use of computers to store and process information on living individuals. The Data Protection Act of 1984 clearly defines terms such as personal data, processing, data user and data subject; lists eight data protection principles; sets out the responsibilities and powers of a Data Protection Registrar; specifies the contents of a Register; and states the rights of individuals.

The main concern of the Registrar is that the activities of data users do not harm any data subject. Since statistics present information about groups of people rather than individuals it is inconceivable they will do any harm. The legislation therefore takes a fairly relaxed line with data used for statistics and it seems unlikely that the activities of government statisticians will be restricted by it.

Personal data used for statistical purposes must be registered but, provided the statistics are not made available in a form which would allow any living individual to be identified, they are exempt from the provisions giving individual data subjects the right of access and they also do not have to be completely accurate.

0. INTRODUCTION: THE HISTORY

The United Kingdom now has a Data Protection Act "to regulate the use of automatically processed information relating to individuals and the provision of services in respect of such information".

The legislation comes at the end of a period of discussion and consultation on computers and privacy going back to

1967. The first steps towards data protection legislation were taken by individual Members of Parliament, who introduced four Private Members Bills between 1967 and 1972. None of these succeeded because they did not have the support of the government but there was genuine public concern about the threat to privacy from the increasing use of computers to store data on individuals and this pressure led to the setting up of an official committee to examine the problem. This was the Younger Committee which published its report in 1972 and laid the foundations for today's legislation. Indeed the eight data protection principles of the Act are based on the ten principles given in the Younger Committee Report.

When the Younger Committee Report was published the government said it would produce a White Paper setting out its views. This White Paper (Cmnd 6353) and its supplement (Cmnd 6354) were finally published over three years later, at the end of 1975. It announced the government's intention to produce legislation. The next step came six months later, in July 1976, when the government appointed a Data Protection Committee under the chairmanship of Sir Norman Lindop to "advise the government on the permanent control machinery needed to secure that all existing and future computer systems holding personal information, in both the private and public sectors are operated with appropriate safeguards for privacy; and to consider and refine the objectives to be incorporated in legislation establishing permanent safeguards".

The interests of statisticians were represented on the Lindop Committee by a university professor of statistics. The committee also consulted very widely and obtained evidence from professional bodies such as the Royal Statistical Society as well as the Central Statistical Office and government departments. In this way the views of statisticians were fully taken into account with the result that the legislation now has a fairly liberal approach to data used for statistical purposes and recognizes statisticians' long tradition of protecting the confidentiality of information on individuals. The Lindop Committee report specifically said there seemed to be no need for subject access to data used for statistics and that data files could be linked and data could be transferred if

there were adequate safeguards to protect the privacy of individuals. They also recommended that individuals should be told if data being collected for statistics may also be used to construct registers for further inquiries.

They considered the benefits of a Universal Personal Identifier and the threat this posed to privacy. They suggested that the Data Protection Administrator should be vigilant to prevent a drift towards using a particular identifier widely and that a UPI should not be introduced without there being legislation for it.

At the time the Lindop Committee was working on its report, the Council of Europe, the OECD and the EEC were also looking into problems associated with computers and privacy. Their deliberations culminated in the Council of Europe Convention (for the protection of individuals with regard to automatic processing of personal data) and the OECD Guidelines (governing the protection of privacy and transborder flows of personal data).

There was little progress in the UK on legislation because the government changed in 1979 but in 1981 the Home Secretary announced the government's intention to introduce legislation, which would enable it to ratify the Council of Europe Data Protection Convention. In April 1982 they published a White Paper (Cmnd 8539) explaining their proposals and shortly after this the Data Protection Bill was first published.

In the UK, proposals for new legislation are discussed, and amended, by both Houses of Parliament. This process often takes a long time and it was not completed before the General Election in 1983 so the Bill was lost. After the election the government re-introduced the Bill and it finally received its Royal Assent in July 1984, twelve years after the Younger Committee published its report.

It will take some time both to establish the office of the Registrar, to set up the register and for data users and bureaux to adapt their systems to the requirements of the legislation - particularly so far as providing for subject access is concerned. The data protection scheme will therefore be implemented in stages. Following the appointment of the Registrar, in the autumn of 1984, there will be a period of six months when the Registrar will be

able to receive applications for registration (and will not be able to refuse registration) and during which it will not be an offence to operate unregistered. After this there will be an eighteen-month period in which it will be an offence to operate unregistered, but during which the Registrar will not be able to exercise his powers so as to secure observance of the data protection principles. The Act will thus not be fully operational until the autumn of 1987.

1. THE COVERAGE OF THE ACT

The Data Protection Act covers data about identifiable living persons which are held and processed by equipment operating automatically, and where the processing is carried out by reference to the data subject. By "identifiable" is meant not only information which identifies an individual by name but also information containing other personal identifiers such as national insurance or membership numbers where the key to translate the codes into names is held by the data user, even if it is not stored in the computer. But a person who merely stores data which have yet to be processed and which he does not intend to process is excluded from the definition of data user. So too is anyone who may have access to data by virtue, say, of a computer terminal and is capable to extracting the information constituting the data (one element of "processing") but who has no control over the contents and use of the data. That is not to say that the Act does not apply at all to such a person because he is not a data user himself. If he is a servant or agent of a data user he will have certain responsibilities arising out of the Act.

The Act defines a number of key terms and it is necessary for each of these to be taken together with the rest in order to see the provisions in their entirety before deciding whether particular data files come within the scope of the Act.

data - information recorded in a form in which it can be processed by equipment operating automatically in response to instructions given for that purpose.

personal data - data consisting of information which relates to a living individual who can be identified from the information, including any expression of opinion about the individual (for example "is a poor credit risk"), but not any indication of the intentions of the data user in respect of that individual (for example "I intend to promote X").

data subject - a natural individual who is the subject of personal data. Data on businesses are therefore outside the Act except in so far as the data may also be about a natural person e.g. one-man business.

data user - means a person who holds data, to be processed by equipment operating automatically in response to instructions given for that purpose and who also, either alone or with others, controls the contents and use of the data.

computer bureau - a person carries on a computer bureau if he provides other persons with services in respect of data. Such services are said to be provided if a person, as agent for other persons, causes data held by them to be processed by equipment operating automatically in response to instructions given for that purpose, or if he allows other persons the use of equipment for so processing information.

processing - in relation to data this means amending, adding to, deleting or re-arranging the data or extracting the information constituting the data and, in the case of personal data, means performing any of these operations by reference to the data subject. This does not apply to any operation performed of preparing the text of documents (i.e. word processing activities).

disclosing - in relation to data, includes disclosing information extracted from the data.

Although it is not defined in the Act, "for statistical purposes" means processing to produce statistics which are not made available in a form which identifies any data subject.

2. THE DATA PROTECTION PRINCIPLES

At the heart of the Data Protection Act are eight general principles governing the collection and use of personal data, the quality of those data, the purposes to which the data shall be put, the right of data subjects to obtain access to data relating to them, and the need for appropriate security measures to protect personal data. The principles are similar to those of the Council of Europe Convention of Data Protection (one object of the legislation is to enable the United Kingdom to ratify the Convention), and the Registrar will be charged with overseeing the observation of these principles. All of the principles will apply to personal data held by data users but only the eighth principle will apply in the case of computer bureaux. Schedule 1 of the Act lists the principles and also provides details of how the principles should be interpreted, in some cases by defining the meaning to be attached to terms used, and in others by detailing factors which will be considered when determining whether the principles have been contravened. The eight principles and some points concerning their interpretation are:

- 1) "The information to be contained in personal data shall be obtained and personal data shall be processed, fairly and lawfully".

An important factor to take into account in determining if the data were obtained "fairly" is to determine if the person supplying the information was misled or deceived as to the purposes for which it is to be held, used or disclosed.

Information is also "fairly" obtained when the supplier is either authorized or required to supply it under any enactment. So collection of statistical information under the Statistics of Trade Act 1947 will for example be deemed to be a "fair" collection.

Of specific interest to statisticians is a further provision later in the Schedule which means that personal data used for historical, statistical or research purposes (and where no damage or distress is likely to be caused to any data subject) are not regarded as being obtained "unfairly" where such use was not disclosed when

the data were obtained, as for example in the case of information originally collected for administrative purposes. Data collected for statistics can also be used for administrative purposes without this use being disclosed, provided the use is described in the register and the Registrar does not decide that the use would make the original collection unfair.

- 2) "Personal data shall be held only for one or more specified and lawful purposes".

To comply with this principle such purposes must be specified in the register and that includes statistical purposes. For example specifying the purposes directly to data subjects will not be sufficient in itself to meet this principle.

- 3) "Personal data held for any purpose or purposes shall not be used or disclosed in any manner incompatible with that purpose or those purposes".

To comply it means that personal data can only be used for those purposes which have been registered and similarly, disclosure is only permitted to those described in details stated in the register entry.

- 4) "Personal data held for any purpose or purposes shall be adequate, relevant and not excessive in relation to that purpose or those purposes".

The Act provides no guidance on what is meant by adequate, relevant or excessive. Formal guidance would presumably be difficult since none of the terms are absolute concepts, each needing to be judged in context. It is possible that the Registrar may provide some elaboration.

- 5) "Personal data shall be accurate and, where necessary, kept up to date".

Any question whether or not data are inaccurate depends on whether they are incorrect or misleading as to any matter of fact, and whether a claim by a data subject for compensation as a result of the inaccuracy is likely to succeed in the courts. In the case of personal data held

solely for statistical purposes it seems impossible to imagine any purely statistical use causing damage to be suffered by any data subject because the data used were inaccurate.

Data which are themselves inaccurate when received by a data user, either from the data subject or a third party, are not regarded as inaccurate providing: they accurately record the information received, carry an indication that they have been so obtained, and if the data subject has notified the data user that he regards the information as incorrect or misleading, they also carry an indication to that effect.

- 6) "Personal data held for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes".

There is a specific relaxation here for statistical purposes. Where personal data are held for historical, statistical or research purposes and not used in a way that damage or distress is caused the data may be kept indefinitely.

- 7) "An individual shall be entitled -

- (a) at reasonable intervals and without undue delay or expense -
(i) to be informed by any data user whether he holds personal data of which that individual is the subject; and
(ii) access to any such data held by a data user, and
(b) where appropriate, to have such data corrected or erased".

By undue expense is meant no more than the maximum prescribed fee. In deciding whether access has been given at reasonable intervals it will be necessary to consider the nature of the data; the purposes for which the data are held; and the frequency with which data are altered. Thus where the data are particularly sensitive, are used for a purpose which is potentially prejudicial to data subjects, or are frequently changed, the Registrar might judge it appropriate for access to be granted more often than where none of these considerations applies. The correction or erasure of personal data is appropriate only when necessary for ensuring compliance with the other data protection principles.

But this principle will not apply to most statistical data. The European Convention on Data Protection with which the UK Act seeks to comply, acknowledged the

special case of statistical and research data and granted an exemption from subject access in certain circumstances. This exemption appears in the Act as follows:

"Personal data held only for - (a) preparing statistics; or (b) carrying out research, are exempt from the subject access provisions; but it shall be a condition of that exemption that the data are not used or disclosed for any other purpose and that the resulting statistics or the results of the research are not made available in a form which identifies the data subjects or any of them".

This allows personal data to be transferred from one data user to another for statistical purposes without losing the exemption from subject access.

- 8) "Appropriate security measures shall be taken against unauthorized access to, or alteration, disclosure or destruction of, personal data and against accidental loss or destruction of personal data".

This applies to all data users and also to computer bureaux which provide services to data users for the processing of data. Two factors which will be considered when determining whether this principle has been contravened are: the sensitivity of the data in question and the harm which would result from a breach of the security measures; and the circumstances surrounding their storage (including the reliability of staff having access to the data).

3. RIGHTS OF DATA SUBJECTS

Data subjects will generally enjoy certain rights subject to certain exemptions as described in the next section. The main right is that of access to personal data but in addition the legislation provides in certain circumstances for compensation for inaccuracy, loss or unauthorized disclosure and it also provides for rectification and erasure of data.

The right of access to personal data means that an individual is entitled to be informed by any data user whether the data held include personal data of which that individual is the data subject and, at the same time, to be

supplied with a copy of the personal data in an intelligible form. The data user must provide this information within forty days of receiving a valid request. The data need not be that which existed when the request was made but nothing can be done to the data which would not have been done if the request for access had not been made.

The right to compensation for inaccuracy applies where damage has been caused as a result of inaccurate data held by a data user, but in proceedings against a data user it will be a defence if the data user can prove that in all the circumstances he took reasonable care to ensure the accuracy of the data. In this respect data which accurately record information obtained from the data subject or a third party will not be regarded as inaccurate simply because the information itself is inaccurate or misleading provided that when the data are held or disclosed the data indicate that the information was so obtained. Thus data users will not be obliged to verify the accuracy of every item of information which they record provided they record the status of the information. But where a data subject notifies the data user that he regards the stored information as incorrect or misleading, the data user will lose his immunity against court action for damages unless he includes a "dispute marker" both as part of the data held and also as part of any information disclosed.

Again where an individual suffers damage as a result of the loss, destruction or unauthorized disclosure of personal data by the data user he will be entitled to compensation through the courts for this and any distress caused. As in the case of compensation for inaccuracy it will be a defence if the data user can prove that he had taken such care as in all the circumstances was reasonably required to prevent the loss, destruction, or unauthorized disclosure in question.

The right to rectification and erasure through the courts is in addition to the Registrar's powers to order rectification or erasure in connection with the seventh data protection principle, that is only where it is necessary for ensuring compliance with the other principles. A court is able to order rectification or erasure on the application of a data subject where the data relating to him are inaccurate within the meaning of the Act or where he has suffered damage by

reason of the unregistered disclosure of personal data in circumstances entitling him to compensation, and where there is a substantial risk of further unauthorized disclosure.

4. EXEMPTIONS

One exemption has already been referred to, namely the exemption from giving subject access in the case of personal data held for preparing statistics or carrying out research (see under the seventh principle above).

More generally the main exemption is in the field of national security where personal data are fully exempt from the operative provisions of the Act. If an exemption is necessary for these purposes they are neither required to be registered nor subject to the Registrar's powers. There are however a number of other exemptions. Personal data held for the purposes of prevention or detection of crime; the apprehension of prosecution of offenders; or the assessment or collection of any tax or duty may be exempt from the subject access provisions. Similarly personal data may be exempt from the non-disclosure provisions if failure to disclose is likely to prejudice any of these purposes. The Secretary of State has powers to place restrictions on the right of subject access to personal data concerning health and social work.

Personal data held solely for payrolls and accounts (that is for the purposes of calculating pay, pensions and other monies payable in respect of employment; and for the purposes of keeping accounts of money paid and received and of goods or services supplied and received) are exempt from registration provided certain conditions are met and data subjects do not have a right to subject access or to seek compensation for inaccuracy, loss or unauthorized disclosure or to obtain rectification or erasure; but the exemption only applies where the data are used solely for the specified purposes and where disclosures of the data are confined to the circumstances set out in the Act. The same exemptions apply to personal data held for certain domestic or other limited purposes that is personal data held by an individual and concerned only with the management of his personal, family or household affairs; or held by an unincorporated members club and relating only to the members

of the club; or held by a data user only for the purpose of distributing articles to the data subjects and consisting only of their names and addresses.

The requirement to provide a copy of personal data within forty days of a request is modified for examination marks to be forty days from the date of publication of the marks, provided this is not more than 5 months later.

5. REGISTRATION

The register to be maintained by the Registrar will provide one of the means by which the United Kingdom will be able to ratify the Council of Europe Convention. It will do so by meeting the obligation laid down in Article 8a of the Convention to ensure that persons should be able to establish the existence of an automated data file, its main purposes and the identity and location of the controller of the file. As a consequence the public will therefore be able to consult the register and discover the uses to which automated personal data are put by data users. The other main use of the register is as a tool for the Registrar in his function of overseeing compliance with the legislation and, in particular, observance of the data protection principles.

All data users who hold personal data will be required to register with the Registrar. Each registration will comprise:

- (a) name and address
- (b) a description of the personal data to be held, and of the purpose or purposes for which the data are to be held or used
- (c) a description of the source or sources from which the data user intends or may wish to obtain the data
- (d) a description of any persons to whom the data may be disclosed
- (e) the names or a description of any countries or territories outside the United Kingdom to which the data may be transferred
- (f) one or more addresses where requests from data subjects requiring access should be sent.

In addition, for each registration a fee will be payable. There are no details at the moment but the Registrar will specify the form in which applications for registration can be made and also the level of detail required. Because of the part it plays in the scheme, it will be a criminal offence to operate as a data user without registering.

Applicants for registrations who hold personal data for more than one purpose will be entitled to either submit a single application or to make separate applications, when they wish to do so, in respect of any or each of those purposes. In the former case the application might make separate mention of each purpose or summarize these purposes into a more general, composite purpose. However the Registrar will be able to refuse an application when he considers the ensuing register entry would provide insufficient information about the matters to which it relates or where he considers a breach of the principles is likely. Such a refusal may result in more information being supplied or additional applications for registration being made.

In deciding whether to submit a single application or a number of separate applications, applicants are likely to have in mind both the fact that a separate registration fee will be payable for each application and also the fact that a separate application for subject access will have to be made in respect of each register entry. It will therefore not necessarily be in the interests of a user holding data for widely differing purposes to make a single registration since, on an application for subject access, he may have to search his entire range of systems for data relating to the person concerned, and for which only one fee would be paid.

The implication here for statistical users in particular, where typically a number of surveys or projects are likely to exist at any one time, is that it will not be essential to separately register each individual survey or project. On the other hand there may well be considerations of the kind already mentioned that would encourage a data user - particularly one with a number of large systems - to make more than one application. One other factor might be a data user's wish to separate the registration of those of his systems where a right to subject access will exist from those where it will not.

6. THE POWERS OF THE REGISTRAR

Under the Data Protection Act it is an offence to hold and process personal data without there being an entry in the register for this. The first power of the Registrar is thus to refuse an application for registration when he thinks the data protection principles would not be complied with.

It is also an offence to hold, use, disclose or transfer data except as described in the entry in the register; and computer bureaux must protect the data as stated in the

eighth principle. If the Registrar thinks a registered user has contravened or is contravening the data protection principles in a way that may cause any person damage or distress he can serve the user with an Enforcement Notice. This specifies the steps to be taken, within a stated time, to comply with the principles being contravened. For example inaccurate data may have to be corrected or erased.

If the Registrar thinks an Enforcement Notice will not secure compliance with the data protection principles he can serve the user with a Deregistration Notice. This will remove the offending entry from the register and force the user to stop holding or processing the data.

The third type of notice the Registrar can serve is a Transfer Prohibition Notice. This is served on a user to prevent the transfer of data from the UK to a State which is not bound by the European Convention.

Data users have the right to appeal against the notices, and refusal of registration, to a Data Protection Tribunal. This tribunal will consist of a chairman and an equal number of people representing data users and data subjects. It is expected this will only be one of each.

When offences have been proved to have been committed a court may order that the data connected with the offence should be forfeited or destroyed.

7. RESPONSIBILITIES OF THE REGISTRAR

The Act lays down the general duties of the Registrar:

- He should set up and maintain a register and provide facilities for making the information in the entries available for inspection by members of the public at all reasonable hours and free of charge.
- He will, on payment of a fee, supply any member of the public with a duly certified copy in writing of the particulars in any entry.
- He should promote the observance of the data protection principles by data users and persons carrying on computer bureaux.
- He should consider complaints that the data protection principles are being contravened when these are made without delay by a person directly affected and will tell

the complainant the result of his consideration and what action he proposes to take.

- He should encourage trade associations or other bodies representing data users to prepare, and to disseminate to their members, codes of practice for guidance in complying with the data protection principles.
- He must prepare an annual report to lay before each House of Parliament.

8. IMPLICATIONS OF THE DATA PROTECTION ACT FOR OFFICIAL STATISTICS

The Act has only just received Royal Assent so we can only guess what effect it will have on the activities of the Government Statistical Service. It is thought that users of personal data for statistical purposes

- will have to register their data, the purposes for which the data are used and the persons to whom the data may be disclosed or transferred. The descriptions will have to be precise; "for statistical purpose" is unlikely to be sufficient to include the use of data to produce registers for subsequent studies.
- will not need to provide subject access
- are unlikely to have to ensure that their data are completely accurate. (This is not stated explicitly in the Act but when he exercises his powers the Registrar must take account of the harm done by non-compliance with the data protection principles. Inaccuracies in data users for statistics cannot harm individuals.)
- will not be prevented from freely transferring anonymized data and producing Public Use Samples.
- will have to ensure that individuals cannot be identified from the statistics but will not have to take any more measures to suppress table entries with small numbers than is the normal practice now.
- will not need to inform data subjects when data collected for administrative purposes are subsequently used to produce statistics.
- will not be able to allow data collected for statistical purposes to be used for administrative purposes.
- will only be able to transfer personal data to countries which have signed the Council of Europe Convention.

It thus seems unlikely that the new Data Protection will place undue restriction on the activities of government statisticians in the UK. Indeed there will be some benefits, for example the Data Protection Registrar will be able to reinforce the assurances given by the Registrar General that the information collected by population census will be confidential, will only be used for statistical purposes, and will not be transferred to another government department.

In the absence of a Universal Personal Identifier, there is little scope for linking data files in the UK but the practice of collecting statistical information in conjunction with that required for some administrative function is quite common. Typically both administrative and statistical data are collected on the same document and, after coding and capturing data, the records of the statistical data file have document numbers to allow coding errors to be corrected by reference to the original forms. If the department chooses to register its administrative data and statistical data separately, and this could be done even if they were stored together, the statistical data would be exempt from the subject access provisions. This would not be true if both statistical and administrative uses are covered by a single entry in the register but there is no way to avoid registering the statistical data. However if the statistical data are copied to another user who cannot identify individuals from the document numbers on the records, the data would not be personal data and would not have to be registered by this user.

DISCUSSION

Prof. ZIGHERA (*Université Paris X-Nanterre*) opened the discussion:

He regretted that, although both the OECD and the Council of Europe had concerned themselves with the question of data protection, the political authorities of the European Communities had appeared less interested by this class of problem. He had a number of questions of detail:

What was the situation referred to on page 180 in which a person would store data without ever processing it?

The purposes of data files needed to be specified when they were being established. If it was proposed to use them for a new purpose, how could this be dealt with?

The holder of a data file was obliged to provide to interested people copies of the entries concerning themselves. In certain situations this could lead to a flood of requests for information which was too large or too costly to be managed. What would happen in these circumstances;

He noted, with concern, proposals to suppress information for small groups of individuals by combining it into larger groups. He pointed out that, in certain modern methods of analysis involving classification by many variables, such small groups were extremely important although they would in no way be identified in the final results.

Prof. LOSANO (*Università di Milano*) remarked as follows:

He expressed his appreciation for Mr. Orchard's explanation of a legal text which continental jurists found very complicated. He admitted being surprised by the author's statement that the UK Data Protection Act would, in substance, resemble similar laws already in force in continental Europe. The problems arose not so much from the legal activities connected with the implementation of this law (it was widely known that the British legal system was different from the continental one), as from two specific points, viz. the very broad powers of the data guarantor (or Registrar) and the applicability of the law to only part of the United Kingdom.

The Registrar had very wide-ranging and ill-defined technical duties, so much so that mapping these out would be possible only after the Data Protection Act had been in force for a few years. This personality - along with the Data Protection Tribunal - nevertheless held power of decision for all technical problems. Thus, an appeal to an ordinary judge (as provided for in Section 14) could be made only in respect of legal problems, but could not call into question the data processed automatically once these had been verified by the Registrar and the Data Protection Tribunal. He was therefore to all intents and purposes a special judge; this was an optimum solution from the technical point of view but one which in Italy was simply out of the question because its Constitution - shaped by specific historical events - prohibited the institution of special judges.

The final section of the Data Protection Act excluded the application of these rules in the Channel Islands, which had a long tradition of legal autonomy *vis-à-vis* the United Kingdom. No other European law on privacy contained a clause of this type. This could lead to the Channel Islands - already considered by many as a tax haven - becoming in the future an informatics haven as well.

Professor Losano was aware that the two examples quoted were closely bound up with Anglo-Saxon legal traditions and therefore had their own undoubted consistency and acceptability within that context. His concern was that these specific aspects would in future be turned against the United Kingdom to deny her access to personal data stored in a foreign country: the latter could base its refusal to make the data available on the law which prohibited (except when expressly authorized) the transfer of personal data to a country which offered no legal guarantees equivalent to those it could itself provide. This could prove a major hurdle to the compilation of statistics at European level.

Mr. HARRIS (*EUROSTAT*) remarked as follows:

A number of Member states would not have been happy to see the European Commission involving itself in the question of data protection legislation. He could not say offhand whether, in fact, such involvement would be in accordance with the Treaties establishing the Community. It was to be regretted that neither the OECD nor the Council of Europe had been able to accept the invitations which they had received to the present Seminar.

Mr. SUDFELD (*Statistisches Bundesamt*) asked:

With regard to the statement of purposes for which a data file was being established, in what degree of detail should these purposes be spelled out.

The President:

Referring to the sentence on page 191 "will not be prevented from freely transferring anonymized data and producing public use samples", he asked whether these were anonymized samples and if so how could one use anonymized samples for a survey? Also the later phrase "will not need to inform data subjects when data collected for administrative purposes are subsequently used to produce statistics" appeared to contradict the principle that the purposes of a data file should be declared initially.

The author replied to the various questions raised as follows:

"Files created but never used" referred to files of back-up data retained for security purposes only and these were excluded from public-access provisions. The entry in the Register in respect of a data file could be amended by the addition, to the stated list of purposes, of another purpose not originally foreseen. The "run on the bank" situation was not, he thought, likely to arise but in any case a "search fee" could be charged for each item of information sought, thereby restraining excessive demands.

The Registrar had already been appointed and was working on the setting up of his office and was discussing with organizations the procedure he would have to follow. The timetable had not been set down for the implementation of the legislation. It was thought that preliminary work would be completed by mid-1985. At that time a date would be fixed from which people would have two years in which to register their data.

The author did not think there was any risk of the Channel Islands' becoming a data haven. The Act could be extended, by Order in Council, to the Channel Islands. As regards the powers of the Registrar, he could not fine individuals but could initiate prosecutions. His powers really consisted in

the issuing of notices. For example he could prevent data being transferred or could deregister data users, thereby preventing them from operation.

He did not know in what level of detail uses of data files must be described: this had yet to emerge, depending on the attitude of the Registrar.

The procedures for transforming anonymized data and producing public use samples already existed with regard to, for example, the census of population. The procedures could continue provided they were registered as uses of the data file. If administrative data were subsequently used for statistics, an amendment of the register would suffice; there would be no need to notify the data suppliers.

DO STATISTICIANS NEED A CODE OF CONDUCT?

R. JOWELL
Social & Community Planning Research

SUMMARY

Unlike their colleagues in other fields, statisticians have long resisted adopting a professional code. They should now reconsider their position. But they should not repeat the mistakes of other codes. Instead they should devise a new model - an educational code - that recognizes the inherent conflict between different statistical goals and does not attempt to prescribe blanket remedies. A possible model for such a code (the draft ISI statement) is attached as an Appendix.

0. INTRODUCTION

When statisticians, researchers or data analysts get together they are always in danger of reinforcing their frequently stated position that rules and regulations imposed upon them by governments are draconian and mostly unjustified. After all, they protest, statisticians are concerned only with the population's (or sub-population's) aggregate characteristics, never with individuals and their identities. The names of people or organizations who supply characteristics. Having served their purpose, the names are then ignored or forgotten. To those whose job it is to collect, analyze and preserve data, this approach to individual identities is and always has been one of their professional norms, part and parcel of the statistical tradition. Why then, they repeatedly ask, do others implicitly suspect their professional practices and seek to restrain them in various inconvenient and counterproductive ways? Or, as the programme notes for this seminar put it, has not "the distrust of government ... gone too far?"

In this paper I argue that statisticians share some responsibility for the introduction of laws which now affect their work, since they have, over the years, been remarkably insensitive to the growing fears of legislators and civil libertarians in many countries. Instead of bolstering their professional practices, conceding the dangers, attempting to

erect their own barricades against possible dangers and potential malpractices, statisticians have, on the whole, largely denied the existence of a problem and, when finally they began to address the issue, they did so too defensively or, at least, too complacently.

Unlike other disciplines and professional groups (not only doctors, lawyers, architects and the like, but also sociologists, anthropologists, psychologists and many others), statisticians have not succeeded in codifying their professional values and norms. Until the American Statistical Association produced a rather thin interim code (1980), statisticians had implicitly or explicitly rejected the notion that professional norms ought to be written down and subjected to scrutiny. Indeed, it took over 30 years of repeated attempts for American statisticians to produce their code.

Meanwhile, the International Statistical Institute (ISI) had been trying vainly on and off for a similar period to produce its own statement of professional ethics. Now, as the ISI's one hundredth anniversary approaches, an agreed draft statement (Appendix 1) looks as if it may finally be adopted. As chairman of the committee charged with devising that statement, I would not, however, be unduly surprised if, in the event, the statement fails to clear its final hurdle at the next ISI Session. Caution (or prudence?) may again prevail.

What follows are extracts from and amendments to a paper given at the ISI plenary session in 1981, at which I was invited to discuss the case for an international professional code for statisticians (Jowell, 1981). As will be seen, after examining the advantages and disadvantages of various types of codes, I was and am persuaded that a code would be useful, but only if it is one that seeks to inform and guide practice rather than one which vainly attempts to regulate or control it. Statistical work is much too diverse to be subjected to universal rules and edicts.

I should stress, however, among this audience, that what follows refers to all statistical data collection and analysis, not specifically or even principally to government statistics.

1. THE CASE FOR A CODE

It would be comforting to believe that the surge of interest in ethics and code formulation among statisticians during the 1970s (an interest generally shared, incidentally, by social scientists), stemmed from a new concern about improving our service to society, an altruistic acknowledgement that others were entitled to discover what we did and how we did it, a determination to open to public scrutiny the high ethical ideals to which we aspired. In reality, however, the interest was more self-serving, its principal objective appearing to be to deflect the growing suspicion among legislators and the public in many countries that statistics and improved methods of data manipulation threatened privacy and civil liberty. The interest coincided with the introduction of data protection and privacy laws in Europe and America, some passed with the acquiescence of the statistical community, some without. The problem for the statistical community was that it did not have the opportunity to offer self-regulation as a realistic alternative to legislation because it had failed to adopt models or codes that would convince others that it had a well-formulated and robust set of protective mechanisms.

To assert that statisticians worked within widely accepted professional constraints was not enough; documentary evidence was required and was simply not available. Unlike other professions and disciplines, we had no codes or statement, merely protestations that what we did was exemplary and threatened nobody. The result was that we were unconvincing, and we remain so.

Collective self-interest is usually the principal motive for the development of a professional code. In our case, for instance, a code might serve to enlighten those who for too long have been allowed to characterize statistics as a mischievous and meddlesome discipline that harms rather than promotes society's interests; it might demonstrate the profession's concern about spurious uses of data and unwarranted intrusions into private domains; it might serve as a defence against improper pressure from clients, funders, employers or legislators.

There would also be wider motives for adopting a code, perhaps the most obvious of which is the creation of a stronger professional identity among statisticians, despite

their diversity of interests, loyalties and activities. Indeed that is the traditional function of a code: to symbolize and consolidate a group's professional identity.

But if we do seek to promote greater professional identity among statisticians (with or without a code) we need to decide what we mean by the word "professional". The word has several meanings - in English, at any rate - two of which concern us here. On the one hand, it refers to the handful of highly regarded occupations traditionally referred to as "the professions", which derive their status partly from their well-established and estimable codes. The Hippocratic Oath, for example, is widely known for the stringent moral standards it invokes for medical practitioners. Indeed, it elevates a doctor's duty to serve the community almost to the level of a creed, implying that medicine is at least as much a cause as a career. The main practical force of such an Oath nowadays seems to be to advance both the self-image and the public image of its signatories.

If our primary motive for developing a code were to achieve this kind of status enhancement or occupational *aggrandisement* we would almost certainly be destined to fail.

On the other hand, members of all occupational groups, humdrum or elite, can be described as being "professional" in their approach to work. Here the word conveys skill, efficiency, a commitment to high standard, probity and, above all perhaps, a sense of pride in the work itself and in the occupational group to which it attaches. If professional identity implies a commitment to these characteristics it is clearly worth pursuing.

The collective and routine pursuit of high standards is, however, a very difficult cause for a code to advance. Even given the unlikely prospect of swift agreement on what constitutes high statistical standards, we would still be faced with the intractable problem of trying to reconcile individually acceptable but collectively conflicting standards. Take, for example, the pursuit of greater accuracy, or the propensity to undertake more and more refined analyses of available data. Such admirable motives may well be responsible for some of the most insensitive (even unethical) behaviour towards the subjects of research. The need for representativeness, for instance, is the usual

justification for undue pressure or deception to secure high response rates in sample surveys. The desire for penetrating analyses may lead to linkages of data that threaten the privacy of those to whom they relate.

No code could legislate on the priorities inherent in such practices. Nor could it even suggest universal rules of "good practice" since too many statistical exercises present unique and complex ethical and technical problems that a generic code could neither predict nor cater for.

Nonetheless, by exposing and publicizing professional dilemmas, by explaining ethical and technical conflicts, by providing a context within which newcomers and outsiders could understand the difficulties, a code would probably afford a modicum of protection to those who would otherwise be adversely affected by our potential excesses. In this respect professional interest and public interest coincide.

So a code of practice would, I suggest, be valuable as long as it did not (as most codes admittedly do) attempt to oversimplify and thereby deliberately falsify the world it referred to, as long as it recognized the inherent conflicts between goals and accepted the need for uneasy compromises.

As Ladd has pointed out (1980 p. 155), "being a professional does not automatically make a person an expert in ethics, even in the ethics of that person's own particular profession". Ladd goes on to argue that organized rules of ethics are an intellectual and moral absurdity. As will become apparent, I share Ladd's aversion to the imposition of rules and sanctions to govern most of our professional responsibilities and regard the proper functions of a code as being explanatory and descriptive, not authoritarian or prescriptive.

2. WHAT TYPE OF CODE?

Most professional codes - or the provisions within them - can be caricatured as belonging to one of two broad classes: those that involve lofty ideals (which can be referred to as aspirational codes); and those that legislate on minute aspects of professional conduct (which can be referred to as regulatory codes). Diener and Crandall (1978) draw the same distinction but refer to the two classes as wisdom ethics and content ethics. The two types of provisions are not mutually exclusive and, indeed, many codes contain both. They are, however, based on different premises.

Aspirational provisions are expressions of often unattainable ideals; they are guides rather than edicts. Consider, for example, the doctrine of informed consent, which broadly refers to the right of potential research subjects not only to refuse to participate but to know the material facts about the study before making their decision. An aspirational provision on informed consent might say that it is "the only honourable basis on which research among human subjects can be undertaken". It might go on to talk about the "dignity and individuality" of research subjects, and so on. It is widely known within the profession that informed consent is frequently neither sought nor obtained. Yet that sort of provision could still be adopted as an expression of ideals to which we would try to conform in the absence of overriding technical (or even other ethical) considerations.

Regulatory provisions, on the other hand, are rules to govern behaviour, based on the premise that there are universal (and enforceable) models of appropriate practice. A regulatory provision on informed consent would probably start its life unambiguously, for example that "informed consent must always be obtained when research among human subjects is being undertaken". But, to make it workable, the rule would almost certainly end its life with the insertion of a qualifying phrase, such as: "unless special circumstances make this undesirable in a particular case". Typically, therefore, regulatory codes tend to be ambiguous on important issues, reserving the less ambiguous provisions for minutiae.

The case against an international regulatory model for statisticians is overwhelming. Not only are there obvious cultural and political differences between countries, but there are also major variations in practice and convention. A set of universal rules would be very different to formulate, still more different to implement. Even if these problems could be solved, a regulatory code would inevitably be so qualified that its utility would be destroyed. It would contain little more than a collection of truisms.

An international aspirational code could avoid these disadvantages, but its utility would be even more different to discern. Rhetorical or sanctimonious calls for moral fibre, altruism and high endeavour are hardly likely to induce more than a momentary inspirational glow. Worse still, they are likely to be transparently different from the world we actually inhabit and are of little practical

value to those who legitimately look to a code for information about, and justifications for, professional norms or ethical conventions.

My bias must by now be apparent: if statisticians are to consider the adoption of professional code, they should reject both the aspirational and regulatory models. They should, instead, adopt a third model, which can be referred to as an educational code. Based on the twin premises that most ethical issues defy unambiguous regulation and that ethical decisions are matters for the individual rather than the group, an educational code would seek to describe and explain professional norms, expose inherent conflicts, and give guidance on possible approaches to their resolution. Its aim would be to ensure that individual ethical decisions are informed by professional experience, not governed by professional authority.

An educational code would not be entirely new. Indeed, it could be described as an aspirational code with a major new element. Yet none of the codes I have located sets out specifically to structure its provisions to illuminate issues rather than to pronounce upon them. Most take as their implicit starting-point the need to control malpractice. An educational code would start from the premise that deliberate malpractice is almost impossible to control, even by law, and that the main function of a code is to enable the diligent professional to better understand the ethical components of his or her work. A strong indictment of the international statistical community is that it provides little or no systematic guidance to its newcomers as to how they might approach the ethical problems and dilemmas we all know they are likely to face sooner or later in their careers.

Such a code could naturally not be enforced, nor could sanctions be imposed for breaches of its provisions. Consideration could, however, be given to inviting the main funders of statistical research (government, industry, foundations) to insist that recipients of grants or contracts should be signatories of the code. That would ensure at least that deviations from the guidelines were the result of individual deliberation rather than of innocence or ignorance.

3. THE COVERAGE OF A CODE

Statisticians, in common with most other groups, are employed in a variety of institutional settings: within government, industry, commerce, consultancy, universities, research institutes. They are required to perform a variety of roles: to offer advice, to collect and assimilate data,

to detect and interpret relationships, to identify and predict trends, to design experiments, to develop analytical tools. A professional code would need to be framed with that diversity in mind.

At the same time, statisticians share common characteristics and concerns, principal among which is, perhaps, a concern with the tools of their trade or methodological issues. They also share the problem of being faced with at least four competing allegiances:

to the individual sources of their data (respondents);
to their employers or funders;
to their peers (at their workplace, in the wider statistical community and in other disciplines);
to society at large.

It is with these four overlapping responsibilities that a professional code should largely concern itself.

In this paper I can barely touch on the variety of ethical considerations in each relationship. Those interested in more detail can refer to the original paper, and to the references cited within it.

4. STATISTICIANS AND RESPONDENTS

Different writers classify the issues relating to the ethical treatment of respondents in different ways. Bower and de Gasparis (1978) list six issues; Diener and Crandall (1978) list eight. The issues can, however, be collapsed into four categories: intrusion and privacy, informed consent, anonymity and confidentiality, potential harm to respondents.

4.1 Intrusion and privacy

"No one shall be subjected to arbitrary interference with this privacy ... Everyone has the right to the protection of the law against such interference..."

UN Declaration of Human Rights
(As quoted in Dalenius, 1977).

Do sample surveys "arbitrarily" interfere with privacy? To some extent they clearly do: samples are selected without the consent of their members, who then receive unsolicited letters, visits or telephone calls from determined

interviewers. These may not be serious intrusions as long as we do not define privacy as the right to be left alone. And according to the (British) Committee on Privacy (1972 p. 10), that interpretation would go "far beyond any right which the individual living in an organized society could reasonably claim".

On the other hand, excessive intrusion through surveys may occur, and researchers must always be sensitive to the possibility of oversurveying certain groups of areas. "Most people see decent obscurity as an important part of their personal freedom" (Barnes, 1979 p. 56).

A more important dilemma for research, so the argument runs, is that the individual's "right to privacy" always needs to be balanced against society's "right to know". A society needs accurate information about its collective characteristics and behaviour so that inequities can be exposed, changes monitored and policies subjected to scrutiny. Since the individual is necessarily the initial source of that information, he or she should not be allowed capriciously to withhold it. It follows from this argument that researchers should be entitled to employ subterfuge or deception on individuals for the benefit of society as a whole.

It is this crusading view of research as an instrument of social changes that leads to the most dubious ethical decisions. It allows researchers to trivialize individual sensibilities, to regard their own work as overwhelmingly important and, by Benthamite sleight of hand, to justify almost any procedures they choose to use. Diener and Crandall (ibid, p. 55) argue convincingly against this approach: "The very fact that people value the privacy of their thoughts as well as their behaviour is reason enough for social scientists to be careful about potential invasions. We respect other people's values for ethical reasons and also for the pragmatic reason that society will censure science if it tramples cultural values".

The difficulty researchers face is that there are no ready-made criteria for determining what approaches, methods or issues are likely to be sensitive, embarrassing or offensive to respondents. Individuals and cultures vary widely in their sensibilities.

Another dimension of privacy concerns the privileged access that statisticians sometimes have to private information, such as medical, tax or police records, for sampling. In one sense this should not be the statistician's responsibility, since he or she has been granted access by the custodian of the records. Nonetheless the statistician is faced with a conflict between technical and ethical values. The ethical imperative would seem to be that the custodian should seek explicit permission from each person before allowing access to his or her record, but the technical advice would be to avoid this potential source of bias and expense.

A code could refer to these and other issues concerning intrusion and privacy. It could conceivably give guidance on some of them. But it would be incautious, if not foolish, to attempt to legislate upon them. Yet some governments have attempted to do so, perhaps because statisticians have simply failed to make the case against blanket regulation. They should have made the case at least partly through a code which demonstrated and explained why almost every instance of potential intrusion requires its own solution and that no general remedy is appropriate or desirable.

4.2 Informed Consent

The doctrine of informed consent was first implicitly invoked in the Nuremberg Code (1947) as part of the judgment of the war crimes tribunal on defendants accused of crimes involving experiments on human subjects. It was later explicitly incorporated into the World Medical Association's Declaration of Helsinki (1964, revised 1975) on biomedical research. The doctrine now has a prominent place in most codes of research ethics.

The doctrine states that researchers should try to avoid both uninformed and misinformed participation by subjects in research. It also forbids actual or implied coercion. Its adoption by medical researchers marked a reaffirmation of the Hippocratic principle that the interests of the patients were paramount in any conflict between them and the interests of science.

When social scientists borrowed the doctrine, they made no such affirmation about the interests of respondents *vis-à-vis* those of science or society. They adopted the doctrine's words and equivocated on its substance. That is still the case. Evidence of this equivocation can be found in much of the literature on research ethics and in many

professional codes. Once again, it stems mainly from a widespread reluctance to admit that there is no generic answer to the question of whose interests should prevail. Indeed regulatory codes cannot admit that possibility, and aspirational codes need not admit it.

So most codes seem to lurch from strong endorsements of the principle that respondents' interests are paramount to strong endorsements of scientific predominance. And they achieve all this movement without even a passing reference to the uncomfortable journey.

In reality, both consent and coercion can be informed, uninformed or misinformed. Censuses, for instance, are undertaken under conditions of informed coercion. In classifying studies on a continuum from informed consent to uninformed coercion, it is surprising to discover how many fail to achieve the condition of informed consent.

The desire for decisiveness in most codes prevents a clear and open discussion of these complicated issues. In reality, I would argue, informed consent is not a doctrine at all. It is a qualified belief in openness and honesty that has to be balanced against other considerations, among which is the sometimes countervailing belief in accuracy and discovery. That it is not to argue that the ideal of informed consent should be dropped from codes or guidelines. It does suggest, however, that we should reject the pious pretence that informed consent is the cornerstone of all research practice. In reality, Bok's "principle of veracity" (1978 p. 30) that "truthful statements are preferable to lies in the absence of special consideration", comes closer to describing prevailing practice. The role of a code in this context should be to highlight rather than to conceal reality and, at best, to influence it. In the end, the individual researcher must "learn to live with an uneasy conscience but continue to be worried by it" (Barnes, 1981).

4.3 Anonymity and confidentiality

Statistical data are, on the whole, unconcerned with individual identities. They exist to answer questions such as "how many?" or "what proportion", not "who?". They must be distinguished from administrative data whose main concern is to identify individual characteristics. But this distinction is confounded by the fact that statistical analysis is often based on data initially collected for

administrative purposes; and, even when it is not, the individual (or group or organization) is usually the initial source of data.

Data protection laws are aimed primarily at administrative rather than statistical data. The report of the (British) Committee on Data Protection (1978), for instance, states positively that "the essential characteristic of statistical data is that they are concerned with groups of individuals and not with individuals themselves ... It follows that the use of personal information for statistical purposes carries few risks for privacy."

But the Committee was actually referring to confidentiality rather than privacy. In reality, there can be no absolute safeguards against breaches of confidentiality - which I define as the release of identified data, or of data whose source can be inferred, when such disclosure contravenes an implicit or explicit obligation to the source. Many methods exist for preventing such breaches, from simple security measures to ever more sophisticated data processing techniques. One of these methods is anonymity - which I define as the condition whereby data travel incognito. Its virtue as a security system is that it helps to prevent unwitting breaches of confidentiality. But it is far from foolproof.

The likelihood of identities being unwittingly disclosed is influenced by several factors, among which are the size of the population, its heterogeneity, the size of sample and the depth of information about each case. Curiously, some data protection laws exempt from control small-scale psychological and qualitative studies simply because they do not involve automated data processing. This reflects an unfortunate and longstanding emphasis on the computer's unique ability to threaten privacy rather than on its impressive potential for protecting it. Once data have found their way into computers they are usually fairly remote from their source (sometimes very remote, as in the case of "transborder flows"). In general, the more remote they are, the less serious are the consequences of inadvertent disclosure, although that does not apply, for instance, when the source is, say, a multi-national company.

People or organizations are mostly asked to participate in statistical inquiries as if each inquiry was an end in itself, a self-contained exercise designed to advance knowledge within specified boundaries.

This belief, associated with the knowledge that participation is confidential and anonymous, encourages respondents to divulge more or less private information about themselves. A difficult question in this context is whether or not the spirit of this participation should allow statisticians subsequently to link identified data from one survey to identified data from another or to identified administrative data. But to attempt a generic answer to that question (as some have done) is clearly counterproductive. Circumstances will inevitably influence whether or not secondary use of data can be justified. In any event, only if systems are relatively secure can statisticians confidently assert to respondents that their interests will be protected against unwitting misuse. And only then might statisticians be able to persuade civil libertarians that the benefits or archived data far outweigh the risks.

Flaherty (1979, p. 307) has argued persuasively that an "ultimate goal of public policy in every country should be to encourage custodians to disseminate data and researchers to use it". Provided that the individual is adequately protected, wider access to data will surely serve rather than threaten the cause of civil liberty and open government.

4.4 Potential harm to respondents

Respondents may be harmed by their participation in research either as individuals or as members of a group. As individuals they may be subjected to undue stress, loss of self esteem or psychological injury. As members of a group they may suffer as a result of being stereotyped by research findings. Both Bower and de Gasparis (ibid.) and Diener and Crandall (ibid.) include excellent discussions of these risks, and I will not attempt to replicate them here. They also include vivid accounts of studies in which researchers have, to put it mildly, overreached themselves. Whatever the advantages may be of conducting certain inquiries, it should always be reiterated that science has no special entitlement to measure all phenomena. Important as it is, the advancement of knowledge is not in itself sufficient cause for violating people's values, whether or not they are acutally harmed in any tangible way.

It is idle to attempt to legislate, as some codes do, against causing harm or distress to respondents. Even seemingly innocent questions about domestic arrangements,

job history and income may inflame or upset people in certain circumstances. It is much more helpful to attempt, as Levine (1975) has done, to produce guidelines for assessing in advance both the probability and the likely severity of harm in various research settings.

The risks of collective stress or harm are still more intangible. Information provided and collected in good faith may be used against respondents' interests, not individually but as members of society. Such circumstances may derive from purely descriptive tags that turn into negative stereotypes or offensive labels, such as: "District X has the highest incidence of crime in the country", or "Women over 60 are the most likely victims of street crime". Or they may derive from social action - based directly on research findings - that ultimately harms the interests of a group to which some respondents belong.

Statisticians may reasonably claim that, as individuals, respondents will not be victimized (or, for that matter, rewarded) as a direct result of participating in research. Such a claim is in any case implied in a pledge of anonymity. But they can rarely claim that respondents will be unaffected by the publication of statistical findings. To do that would be to disparage the value and influence of statistical work.

5. STATISTICIANS AND EMPLOYERS OR FUNDERS

Deming's personal Code of Professional Conduct (1972) offers an admirable model of the obligations that statisticians and their funders or clients should meet in their dealings with each other. It is explicit in its division of responsibilities and uncompromising in its defence of the statistician's right to determine statistical matters. I suspect, however, that only a statistician of Deming's standing could get away with such a code; the rest of the profession remains subject to far more pressure or undue influence from funders or employers than is, perhaps, healthy for the maintenance of high standards. Unlike Deming, some statisticians see themselves as belonging essentially to a service industry where the tradition of the customer's infallibility is rarely questioned.

The reduction in funds for statistical work in many countries may place severe strains on standards of research. Even marginal diminutions in levels of quality control, or in sample sizes, or in budgets for research time, can have a cumulatively damaging (and possibly irreversible) impact on

statistical practice. Moreover, in trying to sustain or restore programme budgets, statisticians may be tempted to exaggerate the explanatory powers of their data, or to accept contracts or grants that embrace dubious methodologies. The growth of these practices would harm the interests of funders and statisticians alike.

A major professional concern should be the reinstatement of the role of the statistician as a designer of data collection methods and an interpreter of data, rather than as a slavish provider of data to predetermined specifications. A code might conceivably be a vehicle for reaffirming the principle (or aspiration?) that statistical research is concerned as much with ideas as with data.

Relationships with funders involve mutual responsibilities. The funder is entitled to expect statisticians to possess attributes such as probity and objectivity, a command of their discipline, and candour in relation to costs, the limitations of their data and the availability of alternative methodologies. The statistician is entitled to expect funders to possess most of these qualities too, but, in addition, to observe the boundaries of the statistician's technical and ethical domains, and to respect the integrity of the data.

Statisticians and their employers have a rather more complex relationship. Some of the same responsibilities apply but are complicated by the special factor of status. A junior statistician may believe, for instance, that he or she is being required to use methodology that is questionable, either from an ethical or from a technical standpoint. Resignation is an ultimate option, but hardly to be recommended in most circumstances. In the absence of reaching agreement on an alternative methodology, what options remain? The statistician can refuse to carry out the assignment, go along with it (perhaps on the grounds that it is someone else's responsibility), or "blow the whistle", thereby involving others in the argument. A code, even an educational code, could sometimes help to prevent such conflicts from becoming crises, simply by illustrating how similar problems have been faced and resolved before. But it could not be used successfully to arbitrate on fine ethical and technical choices.

6. STATISTICIANS AND THEIR PEERS

The obligations and rituals of what may be called professional citizenship are at the heart of many codes of practice: determination of authorship, criticism of learned work, rules against advertising or competition, conventions

of publication, establishment of review committees, and so on. Indeed, to judge from some codes, the professional often inhabits a world into which outsiders rarely intrude. So dominant are intra-professional concerns that other considerations are only grudgingly granted entry.

Other codes attempt to fulfil the role of moral tracts, exhorting their members to display qualities such as honesty, courtesy, consideration and propriety. It would admittedly be more comfortable if all statisticians were, say, considerate and courteous, but from a professional point of view it is far more important that they should be competent and scrupulous. The advocacy of any of these qualities in codes of practice is absurd and demeaning.

Nonetheless, a statistician who blatantly breaches professional norms - say by fiddling data (and being found out), or by gratuitously inflicting harm on subjects - performs a potential disservice to others in the field. Ironically, it is partly for this reason that self-regulation of professions tends to be ineffective. Censure cannot erase the event and may well pollute the environment within which the remainder of the profession is attempting to operate. Thus, public criticism by fellow members is only rarely made. Such closing of ranks is probably self-defeating in the long run. In any case, it is intellectually and ethically hazardous for a profession to behave as if it were an extended family, the members of which should jump to each other's defence come what may.

If we refer back to the purposes of an educational code, the scope of its provisions of professional relationships may become clearer: a code should inform the uninitiated, describe professional norms and highlight potential conflicts. So, in our case, a code should confine itself to norms such as the responsibility of statisticians to describe the limitations of their data, to disclose sufficient details of their methodology to permit informed academic scrutiny, to allow access to raw data within the constraints of confidentiality, to illuminate likely inaccuracies in their data, and so on.

7. STATISTICIANS AND SOCIETY

The phrase "information is power" usually refers to unpublished data held within large corporations or by governments. These data, it is argued, are equivalent to a secret weapon in the hands of "the establishment" that can be used at any time against the public interest. The best

defence, some claim, is to insist on openness and access. And there is certainly a strong case for all data collected at public expense to be publicly available.

As always, however, there are counter-arguments. Since statistical data are concerned with groups and subgroups rather than with individuals, there is always a danger that a particular group will be stereotyped by those who deliberately or inadvertently confuse imperfect associations with perfect descriptions. Nonetheless, this danger should not be overstated and used as an excuse for suppressing innocuous material. I am convinced that failure to publish unidentifiable data can only exceptionally be justified on the grounds of public interest. The exceptions may embarrass those academics who see themselves as disinterested scientists seeking to expose truth (or an approximation of it) without fear or favour. But, as Price (1979, p. 57) points out, "it is no longer possible for scientists to believe that their only ethical obligation is to the advancement of knowledge; the impact of scientific developments on society, and the dependence of research on public support, make the ideal of the ivory tower indefensible". Statisticians must surely always be deeply concerned about the impact of their work on society. On occasion (see Jahoda, 1981), that concern may lead them to resist publication, or at least to delay it.

In their relationships with society, statisticians also have to face three common public stereotypes of their profession. First, that they can prove anything, and frequently do, irrespective of their data. Second, that they are conspirators (with government or industry) in an attempt to invade the privacy or harm the interests of particular groups or individuals. Third, that they are glorified technicians, unaware of their influence on events, who treat the public as just another "object of measurement". Statisticians have to guard against these caricatures and to demonstrate that they are untrue. They have to show (through their code and their professional demeanour) that they are at least as concerned with society's values and sensitivities as they are with the impartiality and accuracy of their data.

8. CONCLUSIONS

I have argued that many ethical goals frequently come into conflict with others and, as frequently, with technical goals. A regulatory code for statisticians would therefore face formidable obstacles. If it tried to regulate only on

seemingly uncontroversial issues, it would be very brief and undemanding; it might also, by omission, appear to endorse all practice to which it did not refer. On the other hand, if its regulations were more wide-ranging it would require vigilant enforcement (which is demonstrably impracticable) or risk being widely, and often justifiably, ignored.

An aspirational code would face fewer obstacles, by referring only to ideals and broad moral duties, it would circumvent problems of definition and implementation. But with these omissions it would be unlikely to have any enduring impact.

Both these models would have another serious drawback: they would be uninformative. Statistical practice involves inherent clashes of values, often between finely differentiated options. Any code that failed to recognize these conflicts would be unrealistic and ineffectual.

I have therefore argued for an educational code covering relationships with respondents (or other sources of data), with funders, with peers and with society at large. Its aims would be to promote an appreciation of ethical and professional issues, to inform ethical choices and to document professional wisdom and experience. Its rationale would be that ethical judgements are best resolved by individual conscience rather than by collective edict. Its function would be to provide a framework and guidelines within which individual statisticians could make informed choices. Its only requirement would be that its signatories should read it and refer to it when faced with an ethical problem or dilemma. It would have to be revised at regular intervals.

Such a code is appended. It is the draft code of the International Statistical Institute due to be considered for adoption next year.

As will be seen, it consists of a codification of broad ideals under various headings, followed by a short discussion of how and why those ideals are difficult to fulfil in all circumstances. The statement does not attempt to conceal the fact that the ideals conflict with one another; on the contrary it tries to highlight the conflicts and to demonstrate that different routes would (and should) be taken in different circumstances. It is appended only as a model of one possible approach to breaking down the traditional resistance of statisticians to the codification of their professional norms. The ISI will in due course make its own decision on whether it is an appropriate model.

Whatever reactions there may be to the details of the statement, however, we must ask whether the time has not come for statisticians to provide tangible affirmation that they do have well-established and fairly universal professional values that are capable of being understood and appreciated by others.

REFERENCES

- AMERICAN STATISTICAL ASSOCIATION (1980) *Interim Code of Conduct for ASA: General Guidelines*. ASA Washington DC.
- BARNES, JA (1979) *Who Should Know What? Social Science, Privacy and Ethics*. Penguin Books.
- BARNES, JA (1981) *Ethical and Political Compromises in Social Enquiry*. Proceedings of the Society of Applied Anthropology's 41st annual meeting, Session 215 (to be published).
- BOK, S (1978) *Lying: Moral Choice in Public and Private Life*. Pantheon.
- BOWER, RT and de GASPARIS, P (1978) *Ethics in Social Research: Protecting the Interests of Human Subjects*. Praeger.
- COMMITTEE ON PRIVACY (1972) (*Younger Committee*) *Report*. HMSO, London. Cmnd. 5012.
- DALENIUS, T (1977) *Computers and Individual Privacy: Some International Implications*. Bulletin of the ISI 67 (1) (203-211).
- DECLARATION OF HELSINKI (1964, revised 1975) as reprinted in Duncan, HS Dunstan, GR and Welbourn, RB (eds) (1981) *Dictionary of Medical Ethics*. Revised and enlarged version, Darton, Longham and Todd, London.
- DEMING, WE (1972) *Code of Professional Conduct*. International Statistical Review 40 (2) (214-219).
- DIENER, E and CRANDALL, R (1978) *Ethics in Social and Behavioural Research*. University of Chicago Press. Chicago and London.
- FLAHERTY, DH (1979) *Privacy and Government Data Banks: An International Perspective*. Mansell, London.
- JAHODA, M (1981) *To Publish or Not to Publish?* Journal of Social Issues 37 No. 1 (208-219).

- JOWELL, R (1981) A professional Code for Statisticians: Some ethical and technical conflicts. Proceedings of the International Statistical Institute, 43rd Session.
- LADD, J (1979) *The Quest for a Code of Professional Ethics: An Intellectual and Moral Confusion*. Paper prepared for the AAAS Workshop on professional ethics - November 1979.
- LEVINE, RJ (1975) *The Role of Assessment of Risk-benefit Criteria in the Determination of the Appropriateness of Research involving Human Subjects*. Paper prepared for the National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research, US Department of Health, Education and Welfare.
- NUREMBERG CODE (1947) as reprinted in Duncan, HS, Dunstan GR and Welbourn, RB (eds.) (1981) *Dictionary of Medical Ethics*. Revised and enlarged version, Darton, Longman and Todd, London.
- PRICE, DK (1979) *The Ethical Principles of Scientific Institutions*. *Science, Technology and Human Values* 26 (46-60).

NOT FOR CITATION OR QUOTATION

DRAFT ISI DECLARATION ON PROFESSIONAL ETHICS

(For discussion prior to and at the 45th
Session of the International Statistical
Institute, Amsterdam, 1985)

BACKGROUND NOTE

This document arises from a decision of the 42nd Session of the ISI in Manila, 1979, to set up an Ethics Committee with the responsibility of formulating a declaration on ethics for possible adoption at the Centenary Meeting in Amsterdam in 1985.

Following the Committee's appointment by the Bureau, a drafting sub-committee was formed whose meetings were funded by the Nuffield Foundation in Britain.

This declaration is based on the approach advocated and supported at a plenary meeting of the 43rd Session of the ISI in Buenos Aires (A professional code for statisticians: some ethical and technical conflicts - Jowell, 1981). The Meeting of the 44th Session in Madrid 1983, where it was broadly endorsed. Various suggestions from those meetings have been taken into consideration.

We are grateful to the British Social Research Association's Ethics Committee for their work on the bibliography and for their useful comments on our various drafts.

We commend the declaration for adoption by the General Assembly and for subsequent distribution within the profession.

R. Jowell (Chairman))	
H. Muhsam)	Drafting
R. Padiou (co-opted))	Group
E. Rapaport)	
E. Berumen Torres (co-opted))	
W.E. Deming)	Consulting
A. Donda)	Group
G. Motsemme (co-opted))	

ISI DECLARATION ON PROFESSIONAL ETHICS

PREAMBLE

Statisticians work within a variety of economic, cultural, legal and political settings, each of which influences the emphasis and focus of statistical inquiry. They also work within one of several different branches of their discipline, each involving its own techniques and procedures and its own ethical approach. Many statisticians work in fields such as economics, psychology, sociology, medicine, whose practitioners have ethical conventions that may influence the conduct of statisticians in their fields. Even within the same setting and branch of statistics, individuals may have different moral precepts which guide their work. Thus, no declaration could successfully impose a rigid set of rules to which statisticians everywhere should be expected to adhere, and this document does not attempt to do so.

The aim of this declaration is to enable the statistician's individual ethical judgements and decision to be informed by shared values and experience, rather than to be imposed by the profession. The declaration therefore seeks to document widely held principles of statistical inquiry and to identify the factors that obstruct their implementation. It is framed in the recognition that, on occasions, the operation of one principle will impede the operation of another, that statisticians - in common with other occupational groups - have competing obligations not all of which can be fulfilled simultaneously. Thus, implicit or explicit choices between principles will sometimes have to be made. The declaration does not attempt to resolve these choices or to allocate greater priority to one of its principles than to another. Instead it offers a framework within which the conscientious statistician should, for the most part, be able to work comfortably. Where departures from the framework of principles are contemplated, they should be the result of deliberation rather than of ignorance.

The declaration's first intention is thus to be informative and descriptive rather than authoritarian or prescriptive. Second, it is designed to be applicable as far as possible to different areas of statistical methodology and application. For this reason its provisions are fairly broadly drawn. Third, although the principles are framed so as to have wider application to decisions than to the issues it specifically mentions, the declaration is by no means exhaustive. It is designed in the knowledge that it will require periodic updating and amendment by a standing committee. Fourth, neither the principles nor the commentaries are concerned with general written or unwritten

rules or norms such as compliance with the law or the need for probity. The declaration restricts itself as far as possible to matters of specific concern to statistical inquiry.

The text is divided into four sections, each of which contains principles or sets of principles followed by short commentaries on the conflicts and difficulties inherent in their operation. The principles are interrelated and therefore need to be considered together; their order of presentation should not be taken as an order of precedence.

At the end of each section, a short annotated bibliography is provided* for those who wish to pursue the issues or to consult more detailed texts.

1. OBLIGATIONS TO SOCIETY

1.1 Widening the scope of statistics Statisticians should use the possibilities open to them to extend the scope of statistical inquiry, and to communicate their findings, for the benefit of the widest possible community.

Statisticians develop and use concepts and techniques for the collection, analysis or interpretation of data. Although they are not always in a position to determine the scope of their work or the way in which their data are ultimately disseminated and used, they are frequently able to influence these matters. (See Clause 4.1).

Academic statisticians enjoy probably the greatest degree of autonomy over the scope of their work and the dissemination of their results. Even so, they are generally dependent on the decisions of funders on the one hand and journal editors on the other for the direction and publication of their inquiries.

Statisticians employed in the public sector and those employed in commerce and industry tend to have even less autonomy over what they do or how their data are utilized. Rules of secrecy may apply; pressure may be exerted to withhold or delay the publication of findings (or of certain findings); statistical series may be introduced or discontinued for reasons that have little to do with technical considerations. In these cases the final authority for decisions about an inquiry may rest with the employer or client. (See Clause 2.3).

* The bibliography has not yet been compiled.

Professional experience in many countries suggests that statisticians are most likely to avoid restrictions being placed on their work when they are able to stipulate in advance the issues over which they should maintain control. Government statisticians may, for example, gain agreement to announce dates of publication for various statistical series, thus creating an obligation to publish the data on the due dates regardless of intervening political factors. Similarly, statisticians in commercial contracts may specify that control over at least some of the findings (or details of methods) will rest in their hands rather than with their clients. The greatest problems seem to occur when such issues remain unresolved until the data emerge.

- 1.2 Considering conflicting interests Statistical inquiry is predicated on the belief that greater access to well-grounded information will serve rather than threaten the interests of society. Nonetheless, in planning all phases of an inquiry, from design to presentation of findings, statisticians should also consider the likely consequences for society at large, groups within it, respondents or other subjects, and possible future research.

No generic formula or guidelines exist for assessing the likely benefit or risk of various types of statistical inquiry. Nonetheless, the statistician has to be sensitive to the possible consequences of his or her work and should, as far as possible, guard against predictably harmful effects. (See Clause 4.4).

The fact that statistical information can be misconstrued or misused is not in itself a convincing argument against its collection and dissemination. All information, whether systematically collected or not, is subject to misuse. And no information is devoid of possible harm to one interest or another. Individuals may be harmed by their participation in statistical inquiries (see Clause 4.4), or group interests may be damaged by certain findings. A particular district may, for instance, be negatively stereotyped by a statistical inquiry which finds that it contains a very high incidence of crime. A group interest may also be harmed by social or political action based on statistical findings. For instance, heavier policing of a district in which crime

is found to be high may be introduced at the expense of lighter policing in low crime districts. Such a move may be of aggregate benefit to society but to the detriment of some districts.

Statisticians are not in a position to prevent action based on statistical data. They should, however, attempt to preempt likely misinterpretations and to counteract them when they occur. But to guard against the use of their findings would be to disparage the very purpose of much statistical inquiry.

1.3 Pursuing objectivity While statisticians operate within the value systems of their societies, they should attempt to uphold their professional integrity without fear or favour. They should also not engage or collude in selecting methods designed to produce misleading results, or in misrepresenting statistical findings by commission or omission.

Science can never be entirely objective, and statistics is no exception.

The selection of topics for attention may reflect a systematic bias in favour of certain cultural or personal values. In addition, the employment base of the statistician, the source of funding and a range of other facts may impose certain priorities, obligations and prohibitions. Even so, the statistician is never free of a responsibility to pursue objectivity and to be open about known barriers to its achievement. In particular, statisticians are bound by a professional obligation to resist approaches to data collection, analysis, interpretation and publication that are likely (explicitly or implicitly) to misinform or to mislead rather than to advance knowledge.

Annotated Bibliography: To follow.

2. OBLIGATIONS TO FUNDERS AND EMPLOYERS

2.1 Clarifying obligations and roles Statisticians should clarify in advance the respective obligations of employer or funder and statistician; they should, for example, refer the employer or funder to the relevant parts of a

professional code to which they adhere. Reports of the findings should (where appropriate) specify their role.

- 2.2 Assessing alternatives impartially Statisticians should consider the available methods and procedures for addressing a proposed inquiry and should provide the funder or employer with an impartial assessment of the respective merits and demerits of alternatives.
- 2.3 Pre-empting outcomes Statisticians should not accept contractual conditions that are contingent upon a particular outcome from a proposed statistical inquiry.
- 2.4 Guarding privileged information Statisticians are frequently furnished with information by the funder or employer who may legitimately require it to be kept confidential. Statistical methods and procedures that have been utilized to produce published data should not, however, be kept confidential.

An essential theme underlying each of the above principles is that a common interest exists between funder or employer and statistician as long as the aim of statistical inquiry is to advance knowledge. (See Clause 1.1). Although such knowledge may on occasions be sought for the limited benefit of the funder or employer, even that cause is best served if the inquiry is conducted in an atmosphere conducive to high professional standards. The relationship between funder or employer and statistician should therefore be such as to enable statistical inquiry to be undertaken as objectively as possible (See Clause 1.3) with a view to providing information or explanations rather than advocacy.

The independent statistician or consultant appears to enjoy greater latitude than the employee-statistician to insist on the application of certain professional principles. In his or her case, each relationship with a funder may be subject to a specific contract in which roles and obligations may be specified in advance (see Deming 1972). In the employee's case, by contrast, his or her contract is not project-specific and generally comprises an explicit or implicit obligation to accept instructions from the employer. The employee-statistician in the public sector may be restricted further by statutory regulations covering such matters as compulsory surveys and official secrecy. (See Clause 4.4).

In reality, however, the distinction between the independent statistician and the employee-statistician is blurred by other considerations. The independent statistician's discretion to insist on certain conditions is frequently curtailed by financial constraints and by the insecurity of the consultant's status. These problems apply less to the employee-statistician, whose base is generally more secure and whose position is less isolated. The employee (particularly the government statistician) is often part of a community of statisticians who are in a strong position to establish conventions and procedures that comfortably accommodate their professional goals (See Clause 1.1).

Relationships with funders or employers involve mutual responsibilities. The funder or employer is entitled to expect from statisticians a command of their discipline, candour in relation to limitations of their expertise and of their data (See Clause 3.1), openness about the availability of more cost-effective approaches to a proposed inquiry, discretion with confidential information. Statisticians are entitled to expect from the funder or employer a respect for their exclusive professional and technical domain and for the integrity of the data. Whether or not these obligations can be built into contracts or written specifications, they remain preconditions of a mutually beneficial relationship.

A conflict of obligations may occur when the funder of an inquiry wishes to ensure in advance (say in a contract) that certain results will be achieved, such as particular finding or a minimum response level in a voluntary sample survey. By agreeing to such a contract the statistician would be pre-empting the results of the inquiry by having made implicit guarantees on behalf of potential subjects as to their propensity to participate or the direction of their response. To fulfil these guarantees, the statistician may then have to compromise other principles, such as the principle of informed consent. (See Clause 4.2).

Above all, statisticians should attempt to ensure that funders and employers appreciate the obligations that statisticians have not only to them, but also to society at large, to subjects, to professional colleagues and collaborators. One of the responsibilities of the statistician's professional citizenship, for instance, is to be open about methods in order that the statistical community at large can assess, and benefit from, their application. Thus, insofar as it is practicable,

methodological components of inquiries should be free from confidentiality restriction so that they can form part of the common intellectual property of the profession. (See Clause 3.2).

Annotated Bibliography: To follow.

3. OBLIGATIONS TO COLLEAGUES

- 3.1 **Maintaining confidence in statistics** Statisticians depend upon the confidence of the public. They should in their work attempt to promote and preserve such confidence without exaggerating the accuracy or explanatory power of their data.
- 3.2 **Exposing and reviewing methods and findings** Within the limits of confidentiality requirements, statisticians should provide adequate information to colleagues to permit their methods, procedures, techniques and findings to be assessed. Such assessments should be directed at the methods themselves rather than at the individuals who selected or used them.
- 3.3 **Communicating ethical principles** To conduct certain inquiries statisticians need to collaborate with colleagues in other disciplines, as well as with interviewers, clerical staff, students, etc. In these cases statisticians should make their own ethical principles clear and take account of the ethical principles of their collaborators.

Each of these principles stems from the notion that statisticians derive their status and certain privileges of access to data not only by virtue of their personal standing but also by virtue of their professional citizenship. In acknowledging membership of a wider statistical community, statisticians owe various obligations to that community and can expect consideration from it.

The reputation of statistics will inevitably depend less on what professional bodies of statisticians assert about their ethical norms than on the actual conduct of individual statisticians. In considering the methods, procedures, content and reporting of their inquiries, statisticians should therefore try to ensure that they leave a research field in a state which permits further access by statisticians in the future. (See Clause 4.1).

Statistical inquiries are frequently collaborative efforts among colleagues of different levels of seniority and from different disciplines. The reputations and careers of all contributors need to be taken into account. The statistician should also attempt to ensure that statistical inquiries are conducted within an agreed ethical framework, perhaps incorporating principles or conventions from other disciplines, and that each contributor's role is sufficiently defined. The World Medical Association's Declaration of Helsinki (1975), for instance, gives excellent guidance to statisticians working in the field of medicine.

A principle of all scientific work is that it should be open to scrutiny, assessment and possible validation by fellow scientists. Particular attention should be given to this principle when using computer software packages for analysis by providing as much detail as possible. Any perceived advantage of withholding details of techniques or findings, say for competitive reasons, needs to be weighed against the potential disservice of such an action to the advancement of statistical knowledge.

One of the most important but difficult responsibilities of the statistician is that of altering potential users of their data to the limits of their reliability and applicability. The twin dangers of either overstating or understating the validity or generalizability of data are nearly always present. No general guidelines can be drawn except for a counsel of caution. Confidence in statistical findings depends critically on their faithful representation. Attempts by statisticians to cover up errors (see Ryten, 1981), or to invite overinterpretation, may not only rebound on the statisticians concerned but also on the reputation of statistics in general. (See Clause 1.2).

Annotated Bibliography: To follow.

4. OBLIGATION TO SUBJECTS*

4.1 Avoiding undue intrusion Statisticians should be aware of the intrusive potential of some of their work. They have no special entitlement to study all phenomena. The advancement

* This section of the declaration refers to human subjects, including individuals, households and corporate entities. For a set of guidelines on animal experimentation, for instance, see the Swiss Academy of Science (1983).

of knowledge and the pursuit of information are not themselves sufficient justifications for overriding other social and cultural values.

Some forms of statistical inquiry appear to be more intrusive than others. For instance, statistical samples may be selected without the knowledge or consent of their members; contact may be sought with subjects without advance warning; questions may be asked which cause distress or offence; people may be observed without their knowledge; information may be obtained from third parties. In essence, people may be inconvenienced or aggrieved by statistical inquiries in a variety of ways, many of which are difficult to avoid. (See also Clause 1.3).

One way of avoiding inconvenience to potential subjects is to make more use of available data instead of embarking on a new inquiry. For instance, by making greater statistical use of administrative records, or by linking records, information about society may be produced that would otherwise have to be collected afresh. Individual subjects should not be affected by such uses provided that their identities are protected and that the purpose is statistical, not administrative. On the other hand, subjects who have provided data for one purpose may object to its subsequent use for another purpose without their knowledge. (See Clause 4.3 iii). This is particularly sensitive in the case of identified data. Decisions in such cases have to be based on a variety of competing interests and in the knowledge that there is no "correct" solution. (See Clause 4.4).

As Cassell (1982) argues, people can feel wronged without being harmed by research: they may feel they have been treated as objects of measurement without respect for their individual values and sense of privacy. In many of the statistical inquiries that have caused controversy, the issue has had more to do with intrusion into subjects' private and personal domains, or by overburdening subjects by collecting "too much" information, rather than with whether or not subjects have been harmed. By exposing subjects to a sense of being wronged, perhaps by the method of selection or by causing them to acquire self-knowledge that they did not seek or want, statisticians are vulnerable to criticism. Resistance to statistical inquiries in general may also increase. (See also Clauses 3.1, 4.3c), 4.5 and 4.6).

4.2 Obtaining informed consent

Statistical inquiries involving human subjects should be based as far as practicable on the freely given informed consent of subjects. Even if participation is required by law, it should still be as informed as possible. In voluntary inquiries, subjects should not be under the impression that they are required to participate. They should be aware of their entitlement to refuse at any stage for whatever reason and to withdraw data just supplied. Information that would be likely to affect a subjects's willingness to participate should not be deliberately withheld, since this would remove from subjects an important means of protecting their own interests.

The principle of informed consent from subjects is necessarily vague, since it depends for its interpretation on unstated assumptions about the amount of information and the nature of consent required to constitute acceptable practice. The amount of information needed to ensure that a subject is adequately informed about the purpose and nature of an inquiry is bound to vary from study to study. No universal rules can be framed. At one extreme it is inappropriate to overwhelm potential subjects with unwanted and incomprehensible details about the origins and content of a statistical inquiry. At the other extreme it is inappropriate to withhold material facts or to mislead subjects about such matters. (See Clauses 4.3d) and 4.4). The appropriate information requirement clearly falls somewhere between these positions but its precise location depends on circumstances. The clarity and comprehensibility of the information provided are as important as the quantity.

An assessment needs to be made of which items of information are likely to be material to a subject's willingness to participate. The following items are among those that need to be considered.

- i) Purpose of study, policy implications, etc.
- ii) Identity of funder(s)
- iii) Anticipated uses of the data, form of publication etc.
- iv) Identity of interviewer/experimenter and organizational base
- v) Method by which subject has been chosen (sampling frame, etc.)
- vi) Subject's role in study
- vii) Possible harm or discomfort to subject

- viii) Degree of anonymity and confidentiality
- ix) Proposed data storage arrangements, degree of security, etc.
- x) Procedures of study (time involved, setting, etc.)
- xi) Whether participation is voluntary or compulsory:
 - a) if compulsory, potential consequences of non-compliance
 - b) if voluntary, entitlement to withdraw consent (and when that entitlement lapses)
- xii) Whether material facts have been withheld (and when or if such facts will be disclosed).

In selecting from this list, the statistician should consider not only those items that he or she regards as material, but those which the potential subject is likely to regard as such. Each party may well have special (and different) interests. As a means of supplementing the information selected, the statistician may choose to give potential subjects a declaration of their entitlements (see Jowell, 1981) which informs them of their right to information but leaves the selection of extra details in the subject's control.

Just as the specification of adequate information varies, so does the specification of adequate consent. A subject's participation in a study may be based on reluctant acquiescence rather than on enthusiastic cooperation. In some cases, the statistician may feel it is appropriate to encourage a sense of duty to participate in order to minimize volunteer bias. The boundary between tactical persuasion and duress is sometimes very fine and is probably easier to recognize than to stipulate. In any event, the most specific generic statement that can be made about adequate consent is that it falls short both of implied coercion and of full-hearted participation.

On occasions, a "gatekeeper" blocks access to subjects so that statisticians cannot approach them directly without the gatekeeper's permission. In these cases, statisticians should not devolve their responsibility to protect the subject's interests onto the gatekeeper. They should also be wary of inadvertently disturbing the relationship between subject and gatekeeper. While respecting the gatekeeper's legitimate interests they should adhere to the principle of obtaining informed consent directly from subjects once they have gained access to them.

The principle of informed consent is, in essence, an expression of belief in the need for truthful and respectful exchanges between statisticians and human subjects. It is clearly not a precondition of all statistical inquiry. Equally it remains an important and highly valued

professional norm. The acceptability of statistics depends increasingly not only on technical considerations but also on the willingness of statisticians to accord respect to their subjects and to treat them with consideration. (See Clause 4.1).

4.3 Modifications to informed consent Where technical or practical considerations inhibit the achievement of prior informed consent from subjects, the spirit of this principle should still be adhered to, for example:

- a) **Respecting rights in observation studies** In observation studies, where behaviour patterns are recorded without the subject's knowledge, statisticians should take care not to infringe what may be referred to as the "private space" of an individual or group. This will vary from culture to culture. Where practicable, statisticians should attempt to obtain consent post hoc. In any event, they should interpret behaviour patterns that appear deliberately to make observation difficult as a tacit refusal of permission to be observed.
- b) **Dealing with proxies** In cases where a proxy is utilized to answer questions on behalf of a subject, say because access to the subject is uneconomic or because the subject is too ill or too young to participate directly, care should be taken not to infringe the "private space" of the subject or to disturb the relationship between subject and proxy. Where indications exist or emerge that the subject would object to certain information being disclosed, such information should not be sought by proxy.
- c) **Secondary use of records** In cases where subjects are not approached for consent because a statistician has been granted access, say, to administrative or medical records or other research material for a new or supplementary inquiry, the custodian's permission to use the records should not relieve the statistician from having to consider the likely reactions, sensitivities and

interests of the subjects concerned, including their entitlement to anonymity. Where appropriate, subjects ought to be approached afresh for consent to the new inquiry.

- d) **Misleading potential subjects** In studies where the measurement objectives preclude the prior disclosure of material information to subjects, statisticians should weigh the likely consequences of any proposed deception. To withhold material information from, or to misinform, subjects involves a deceit, whether by omission or commission, temporarily or permanently. Such manipulation will face legitimate censure and should not be contemplated unless it can be justified. Instead, consideration should be given to informing subjects in advance that material information is being withheld, and when or if such information will be disclosed.

A serious problem arises for statisticians when methodological requirements conflict with the requirement of informed consent. Many cases exist in which the provision of background information to subjects (say, about the purpose or sponsorship of a study), or even the process of alerting them to the fact that they are subjects (as in observation studies), would be likely to produce a change or reaction that would defeat or interfere with the objective of the measurement. These difficulties may lead statisticians to waive informed consent and to adopt either covert measurement techniques or deliberate deception in the interests of accuracy.

The principles above urge extreme caution in these cases and advise statisticians to respect the imputed wishes of subjects. Thus, in observation studies or in studies involving proxies, the principle to be followed is that mere indications of reluctance on the part of an uninformed or unconsenting subject should be taken as a refusal to participate. Any other course of action would be likely to demonstrate a lack of respect for the subject's interests and to undermine the relationship between, say, proxy and subject on the one hand, and between statistician and subject on the other.

Statistical inquiries involving deliberate deception of subjects (by omission or commission) are rare and extremely difficult to defend. Clear methodological advantages exist

for deception in some psychological studies, for instance, where revealing the purpose would tend to bias the responses. But, as Diener and Crandall (1978) have argued, "science itself is built upon the value of truth"; thus deception by scientists will tend to destroy their credibility and standing. (See Clause 3.1). If deception were widely practised in statistical inquiries, subjects would in effect, be taught not to "trust those who by social contract are deemed trustworthy and whom they need to trust" (Baumrind, 1972).

Nonetheless, it would be as unrealistic to outlaw deception in statistical inquiry as it would be to outlaw it in social interaction. Minor deception is employed in many forms of human contact (tact, flattery, etc.) and statisticians are no less likely than the rest of the population to be guilty of such practices. It remains the duty of statisticians and their collaborators, however, not to pursue methods of inquiry that are likely to infringe human values and sensibilities. To do so, whatever the methodological advantages, would be to endanger the reputation of statistics and the mutual trust between statisticians and society which is a prerequisite for much statistical work. (See Clause 3.1).

In cases where informed consent cannot be acquired in advance, there is usually a strong case, for the reasons above, for seeking it post hoc. Once the methodological advantage - of covert observation, of deception, or of withholding information - has been achieved, it is rarely defensible to allow the omission to stand.

4.4 Protecting the interests of subjects
Neither consent from subjects nor the legal requirement to participate absolves the statistician from an obligation to protect the subject as far as possible against potentially harmful effects of participating. The statistician should try to minimize disturbance both to subjects themselves and to the subjects' relationships with their environment. Statisticians should help subjects to protect their own interests by giving them prior information about the consequences of participating. (See Clause 4.2).

Harm to subjects may arise from undue stress through participation, loss of self-esteem, psychological injury or other side effects. Various factors may be important in assessing the risk-benefit ratio of a particular inquiry, such as the probability of risk, the number of people at

risk, the severity of the potential harm, the anticipated utility of the findings, few of which are usually quantifiable. (See Levine, 1975).

The interests of subjects may also be harmed by virtue of their membership of a group or section of society (see Clause 1.2). So statisticians can rarely claim that a prospective inquiry is devoid of possible harm to subjects. They may be able to claim that, as individuals, subjects will be protected by the device of anonymity. But, as members of a group or indeed as members of society itself, no subject can be exempted from the possible effects of decisions based on statistical findings.

When the probability or potential severity of harm is great, statisticians face a more serious dilemma. A statistician may, for instance, be involved in a medical experiment in which risks of some magnitude to subjects are present. If volunteers can be found who have been told of the risks, and if the statistician is convinced of the importance of the experiment, should he or she nonetheless oppose the experiment in view of the risks? In these circumstances, probably the best advice is to seek advice - from colleagues and others, especially from those who are not themselves parties to the study or experiment.

4.5 Maintaining confidentiality of records Statistical data are unconcerned with individual identities. They are collected to answer questions such as "how many?" or "what proportion?", not "who?". The identities and records of cooperating (or noncooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged.

4.6 Preventing disclosure of identities Statisticians should take appropriate measures to prevent their data from being published or otherwise released in a form that would allow any subject's identity to be disclosed or inferred.

There can be no absolute safeguards against breaches of confidentiality, that is the disclosure of identified or identifiable data in contravention of an implicit or explicit obligation to the source. Many methods exist for lessening the likelihood of such breaches, the most common and potentially secure of which is anonymity. Its virtue as

a security system is that it helps to prevent unwitting breaches of confidentiality. As long as data travel incognito, they are more difficult to attach to individuals or organizations.

There is a powerful case for identifiable statistical data to be granted "privileged" status in law so that access to them by third parties is legally blocked in the absence of the permission of the responsible statistician (or his or her subjects). Even without such legal protection, however, it is the statistician's responsibility to ensure that the identities of subjects are protected.

Anonymity alone is by no means a guarantee of confidentiality. A particular configuration of attributes can, like a fingerprint, frequently identify its owner beyond reasonable doubt. So statisticians need to remove the opportunities for others to infer identities from their data. They may decide to group data in such a way as to disguise identities (see Boruch & Cecil, 1979) or to employ a variety of available measures that seek to impede the detection of identities without inflicting very serious damage to the aggregate dataset (see Flaherty, 1979). Some damage to analysis possibilities is unavoidable in these circumstances, but it needs to be weighed against the potential damage to the sources of data in the absence of such action. (See Finney, 1984).

The widespread use of computers is often regarded as a threat to individuals and organizations because it provides new methods of disclosing and linking identified records. On the other hand, the statistician should attempt to exploit the impressive capacity of computers to disguise identities and to enhance data security.

Annotated Bibliography: To follow.

*

* *

DISCUSSION

Mr. PEARE (*EUROSTAT*) opened the discussion:

He said that it had become very obvious during earlier discussion that the ethical aspects of the subject of the seminar was a major preoccupation of participants. He commented on the paper from two different viewpoints, firstly, as a member of that branch of the statistical profession known as official statisticians. It had been suggested that a sense of professional ethics had led official statisticians to dissent from changes made in the

role of official statistics in the UK. The existence of a code such as this could strengthen the hand of official statisticians in resisting what they regard as unwarranted (but legal) interference in their professional activities. Most of its proposals would probably be acceptable to the majority of official statisticians. In particular he welcomed the first sentence of section 1.2 "Statistical inquiry is predicated on the belief that greater access to well-grounded information will serve rather than threaten the interests of society". Together with the first sentence of section 4.5 "Statistical data are unconcerned with individual identities" he believed that this constituted the core of a declaration of statistical ethics which should reassure worried respondents, at least as to the motives of statisticians.

Reassurance must, of course, go beyond motives and deal with the practical steps to be taken to ensure that these intentions are realized. The first three chapters seemed to represent admirable steps, of general applicability towards achieving this aim. Similarly section 4.1 which exhorted statisticians to be aware of the intrusive potential of some of their work seemed essential. However, sections 4.2 and 4.3 caused some difficulties. Professor Jowell commented on the necessarily vague nature of the principle of informed consent. At a minimum, he felt that the qualification in section 4.3a which suggested that the "private space" of individuals would vary from culture to culture should be applicable throughout these two sections or to be more radical and that these two sections would be better placed as examples of how to achieve the exhortation of section 4.4. To answer the question posed by Professor Jowell in the title of his paper from the viewpoint of an official statistician he thought the answer should be yes and further that the proposed code was very close to a suitable code for such statisticians.

His second viewpoint was that of a member of the statistical profession in the larger sense. Professor Jowell rejected analogies with codes of ethics for other professions such as medicine and law. However, he and his colleagues had implicitly accepted an invalid analogy between statistics and these other professions. Virtually all medical and legal practice was in the hands of members of those two professions but this was far from being the case for statistics. Even within government service, essentially statistical operations were being carried out by persons who would not regard themselves as statisticians and outside government service this phenomenon was widespread. This was not something to be discouraged but much of the adverse publicity which from time to time attached itself to

statistics arose from statistically unprofessional behaviour by persons who would not regard themselves as statisticians. Professor Jowell suggested that "the profession should take responsibility for its members' conduct". It would be difficult to dissent from this proposition but it did seem valid to ask whether the profession should also take responsibility for the conduct of non-members. The declaration on professional ethics had a feel of being drafted by statisticians for statisticians (Professor Jowell's earlier paper given at the ISI conference in Buenos Aires in fact had the title "A professional code for statisticians; some ethical and technical conflicts") and he wondered if it would receive wider attention were more emphasis placed upon it as a code for "statistical practice" rather than as a code for statisticians?

Mr. BEGUE (*Institut National de la Statistique et des Etudes économiques*) said:

He wondered whom this code was destined for. Would it be issued to the general public which would immediately forget it? Would it be quoted if some major issue arose? This was akin to what statistical institutes did already.

Prof. LOSANO (*Università di Milano*) said:

There was no questioning the usefulness of a professional code of conduct of the kind proposed by Mr. Jowell. This would provide the statistician with a basic handbook in the event of doubts in the execution of his duties. A code of conduct entailed, however, two unavoidable risks which must be borne in mind when it was drawn up. On the one hand, how effective a professional code of conduct proved depended above all on the extent to which its clauses involved sanctions, not legal sanctions in this case, but "social" or "group" sanctions. The code could otherwise end up as what the author himself called an idealistic code. On the other hand, setting up a system of even minimal sanctions entailed bureaucratizing the profession; it would lead to the creation of an order (of lawyers, notaries, journalists, business consultants, etc.) in which a disciplinary council would apply social sanctions to anyone infringing the professional code of conduct. In the climate of high unemployment, this trend could lead to the creation of closed corporations whose code of conduct would be stringently adhered to but access to which would be heavily bureaucratized. This observation related specifically to item 4.3a, which mentioned surveys conducted on the basis of observations of subjects without their knowledge. When this part of the code of conduct was being drawn up, due account should be taken

of the fact that certain national legislation prohibited this type of direct survey, at least for certain categories of subjects. For instance, Article 4 of the Italian Law known as the "Statuto dei Lavoratori" which covers workers' rights, etc. (and to which he referred in Section 2 of his report) explicitly prohibited all forms of monitoring of workers except in specific cases to which the trade unions had agreed. Any attempt at post hoc consent of the type presented in the proposed code of conduct would, in Italy spark off a serious trade union dispute.

Mr. KEIDING (*University of Copenhagen*):

He referred to the situation in regard to the Helsinki declaration on medical experimentation on humans. In continuation of Prof. Losano's remarks about the need for sanctions, the Helsinki declaration did play a role in medical research and was being enforced, mainly through review by ethical committees. He considered that there was a parallel between medical research and statistical privacy. Thus the British had few problems while in Denmark and other Scandinavian countries a very determined effort by the medical profession succeeded in preventing a major conflict arising on medical research. However, in the German Federal Republic a grave conflict between the legal profession and the medical researchers all but stopped randomized clinic trials in the early 1980's. He wondered whether the data protection personnel should also have some involvement with the proposed code of conduct.

Mr. QUATRESOOZ (*ESOMAR*):

He explained that he was representing ESOMAR, the European Society for Opinion and Marketing Research which had over 2000 members. This society had published codes of ethics and of fair practice governing opinion surveys. These were, of course, binding on the members only but they had persuaded the managing director of major market research organizations to undersign the codes. Only these companies which had signed were mentioned in the ESOMAR handbook. A commission met annually and sometimes expelled members who did not observe the codes. Membership of ESOMAR was important to clients when commissioning surveys. The code was also taken into account in the courts when conflicts arose.

Mr. JOHNSTON (*United Nations Statistical Office*):

He said that the international statistical community was greatly indebted to Prof. Jowell for having undertaken this very difficult and complex task, to which the paper made a great contribution. He wished to restrict his comments to the perspective of the present seminar, namely privacy of data processing in statistics, with special reference to the issues concerning population censuses. Prof. Jowell had certainly recognized the difficulty of designing one code to fit a tremendous variety of circumstances and situations, but it was nonetheless important to ask how well the draft code would do the job. He raised the following questions:

Section 4.1 on "avoiding undue intrusion" made no mention of the inevitable intrusion which occurred if data response was obligatory, as in a census, nor did it offer any rationale in defense of such intrusion. In section 4.5 it was stated that "statistical data are unconcerned with individual identities," yet identification data were almost always collected in statistical inquiries. Did this not seem contradictory to the public? Surely, data protection officials or a concerned public were not going to be convinced of the soundness of statistical ethics by such general and apparently contradictory positions as these. Likewise in section 4.6 the statisticians might be called on to state more clearly the extent and nature of confidentiality, even to provide examples of how it was ensured. The public could well wonder what a blanket statement of confidentiality really meant.

He asked whether other complementary and perhaps more practical steps could be taken to define statistical ethics, within the larger context of defining and identifying statistics as a profession more explicitly. For example, statistical education could give explicit attention to ethical issues. Likewise, at a technical level, there would be much more explicit attention, from both professional and official perspectives, to ways and means of ensuring confidentiality as a practical matter, in the face of new computer technologies.

The author replied to the discussion:

He remarked that the code of ethics was intended to cover statistical practices and was not confined only to

professional statisticians or ISI members. He agreed that one of the purposes of the declaration was to convince the public. However, the main purpose was to influence statisticians. In many cases the educational programme of statisticians ignored questions of research ethics. He did not agree with Prof. Losano that the code was ineffective because sanctions were impossible. The ISI certainly could not apply sanctions but it hoped to persuade national statistical institutes to adopt and promote the code. He was grateful for the suggestion that data protection agencies should be consulted about the code: this had not yet been done. So far only statistical groups had been consulted and the response had been overwhelmingly favourable. He was familiar with the ESOMAR code, which he considered excellent but which was mainly for commercial market research purposes and was too narrow to apply to statistical work in general.

With regard to compulsory surveys, such as the census, the code laid down that the obligation to inform the respondent was in no way diminished. He recognized that the code was extremely limited on particular ways of ensuring confidentiality and the various drafts of the code had become successively shorter on this point. Practices favoured by one country were rejected by another. However, annexed to this code would be an extensive bibliography enabling people interested in this aspect to study the subject in depth and to see how ethical dilemmas had been dealt with.

Mr. THYGESEN (*Danmarks Statistik*) subsequently furnished the following comments in writing:

If we are not to have a common code of conduct it is essential that his code should be beneficial to the case of statistics: It would help to reassure the general public that the production of statistics is not a dangerous and harmful activity. The difference between statistics and administrative control should be spelled out very clearly. The present draft puts too much emphasis on the dangers of statistics and the many precautions that have to be taken by statisticians. This could very well have the opposite result.

The code should only contain considerations and recommendations that statisticians can agree upon across the borders. The organization of society and of statistics differs greatly from country to country. Therefore, the code should concentrate on a few problems that are common to most types of statistical activity.

DATA PROTECTION AND THE STATISTICAL COMMUNITY

D.H. FLAHERTY

University of Western Ontario, London, Canada

SUMMARY

The statistical community tends to react negatively to the suggestion that it should be subject to data protection. Statisticians argue that the confidentiality of the data with which they work is adequately protected by existing laws and practices. Public reaction to census controversies suggest, however, that there is much scepticism about the ability of governments and their statistical agencies to protect personal information properly. The data protection movement is a response to this concern. Thus, it is argued, the statistical community should, in its own best interests, recognize the political ramifications of their data collection activities and adopt a more constructive attitude in their relations with data protection agencies.

0. INTRODUCTION

The precise focus of this essay is the application of data protection principles and practices to the statistical community. But in order for my particular perspective and, indeed, biases to become clear to you, it is necessary to begin with a personal retrospective on my career as a privacy advocate.

From 1974 to 1978 I was the co-principal investigator of an international research project on the problem of balancing the need to protect personal privacy and the need for access to government microdata for research and statistical purposes. Although I had some prior experience with the study of privacy issues, I had never before worked with national statistical agencies or on privacy and data protection issues outside North America. The research goal was to look at the implications for personal privacy of the collection and storage of personal information by government statistical agencies and the subsequent desire of the research community to use some of the same personal information, sometimes in identifiable form, for research and statistical purposes.

The results of my research were published in 1979 as a series of case studies of five countries in a volume entitled "Privacy and Government Data Banks: An

International Perspective (1)". In retrospect, I regard this volume as too much of an apology for the point of view of national statistical agencies. I identified strongly, as I still do, with the societal importance of the work of government statisticians and researchers and in effect wrote a brief for the desirability of what they were doing. If anything, I regarded the activities of data protectors, in such countries as Sweden, as overly intrusive in the statistical field and, perhaps, even somewhat unnecessary. If I was then too much under the sway of statisticians and researchers, similar allegations can now be made about my current identification with data protectors at the national and state level.

The research project described above enabled me to observe the development of data protection legislation in Sweden, West Germany, the United Kingdom, Canada, and the United States, especially from the point of view of statistical activity. In 1979-80 I decided to begin another comparative project on the implementation of such "privacy" laws during the process of moving from principles of data protection to actual practices. Since 1981 I have been studying the small bureaucracies created to oversee the implementation of data protection laws in Sweden, West Germany, France, the United Kingdom, Canada, and the United States. At this point in my research, I am a strong admirer of the data protection initiative, but somewhat sceptical about its long term impact on government personal information practices, because of the strong countervailing powers seeking to promote efficiency and cost reductions in the public sector (2).

Since 1981 I have focused primarily on problems of data protection in administrative data bases, such as those maintained by the police. I have not had significant contact with statistical agencies, except with Statistics Sweden, where the monitoring of the relationship between this organization and the Data Inspection Board (DIB) is a good bellweather for how data protection is working in practice. But, because of my 1979 book, I have continued to be interested in the relationship between data protection and statistics, including the census controversies in West Germany and Sweden in 1983.

In this essay I propose to comment on statistical activities from another perspective than the one I adopted in my 1979 book. I now take a more clearly pro-privacy stance in my

research, in part because privacy advocates are a tiny minority in every country, and privacy interests need to be articulated on a continuing basis. I think I now have a clearer perspective on how statistics should or could fit into data protection activities and vice versa. Since I am not a government official but an academic, you will also notice my willingness to offer perspective comments about how other people should run their agencies.

I propose to touch on the following issues in a somewhat comparative, but not comprehensive, perspective: 1) some observations on recent interactions between statisticians and data protectors in several countries; 2) why there continue to be negative reactions from the general public to statistical activities; and 3) how statistical agencies should fit into the overall scheme of data protection. In general it is my view that certain national statistical agencies have not been very sophisticated or sufficiently responsive to important national initiatives in data protection: my intention is to take a middle ground between statisticians and data protectors and to cast a few barbs in both directions. You may all then join together in making similar comments on the limits of the academic mentality.

1. WEST GERMANY

My earlier Privacy Project dealt directly with the regulation of the Federal Statistical Office (SB) and the Hesse Statistical Office by the emerging data protection institutions, but I have not had any direct contact with West German Statisticians since 1978. A new federal law on statistics has been enacted and is in force (3). Section 11, paragraph 5 of this new law stipulates under what circumstances and in what form statistical offices may disseminate personal data. The Conference of Federal and State Data Protection Commissioners set up a working group to develop the application of data protection principles in the context of the statistics laws, especially as they relate to the transfer of personal information among the various statistical offices (4). In his 1980 annual report Professor Hans Peter Bull, who served as the Federal Data Protection Commissioner from 1978 to 1983, concluded that the Federal Statistical Office in Wiesbaden had made significant steps forward in implementing data protection (5).

Yet, as you all know, an enormous public controversy broke out in West Germany early in 1983 over the census of population scheduled for late April. This episode illustrates two themes: 1) the continuation of public anxieties about personal privacy; and 2) the failure of the Federal Statistical Office to take adequate account of public sensibilities and the implications of data protection in carrying out its preparations for the census, which in West Germany primarily means crafting a new special law for this particular data collection activity.

The Federal Constitutional Court in effect postponed the West German census on April 13, 1983 on ground of potential invasion of privacy, despite the passage of the special 1983 census law of March 25, 1982 by a unanimous legislature and the general support of major data protection officials for the census (6). The government had committed itself to going ahead with the census on March 29, 1983, and the Office of the Federal Data Protection Commissioner engaged in public relations in defence of the plans for the census. The suit arose when several citizens challenged the constitutionality of the census law on the grounds that it constituted an invasion of privacy under Articles 1 and 2 of the Basic Law (or constitution) of 1949, especially the paragraph in the statute dealing with the use of data to be collected. The Federal Constitutional Court received written comments and oral statements from the representatives of statisticians and from leading data protectors themselves. The initial case for statistics was not very well prepared by the federal Ministry of the Interior, it is suggested, and statisticians fell back on their usual arguments for carrying out their traditional activities. Certain of the data protection commissioners used the first hearing as an opportunity to explain to the highest West German court their continued concerns about certain aspects of the census law, especially the continuation of the practice of using census returns to correct local population registers (7).

Continued public anxieties about even such a relatively innocuous and desirable activity as the taking of a census are a reminder of how much people continue to worry about the erosion of their personal privacy in an era of information technology. (I am well aware that the points will be most familiar to you, yet they need to be re-emphasized since they create the political culture in which statistical agencies have to operate, whether they like it or not.) Prof. Spiros Simitis, the Hesse Data

Protection Commissioner, has described the census controversy as the first mass movement for data protection in West German history. There were literally hundreds of phone calls to his office demanding limits on government information collection. Significant segments of the population rose up against the proposed census, because of its general fears about the use of computers in data processing.

Even though the case for society's need for a population census can be made successfully, even though the tradition of confidentiality at statistical agencies can be demonstrated, and even though some of the West German resistance to the census was motivated by political considerations and irrationality, the census provides a recurrent opportunity in every country for people to oppose the collection of personal information by all government agencies. The census is one of the very few universal data collection activities and is often carried out on a compulsory basis. It is well known that the censuses between 1969 and 1971 in North America and Western Europe helped to spark the data protection movement in the first place. In late March 1983 a public opinion poll in West Germany indicated that 52% of the population mistrusted the census questions and that 25% of the 25 million German households would not complete the form (8). It was not enough for a new Minister of the Interior, Herr Zimmerman, to argue that people should trust the government, since Western populations are clearly not in a frame of mind to do so. The Federal Statistical Office was equally unprepared for the extent of public resistance, despite the number of warnings it had received, from data protectors in particular, about potential problems.

It is ironic that the census controversy, which exposed the inadequacies of data protection itself to public and judicial scrutiny, had the result of formally creating a new basis for data protection as a constitutional right. On December 15, 1983 the Federal Constitutional Court in Karlsruhe ruled unanimously in the census case that West Germans have a constitutional right to self-determination about their own personal information (9). Although this right is not absolute in the face of competing societal values, information use, the court ruled, has to occur on the basis

of established laws and in response to the overwhelming public interests of the community. The court's positive ruling on the census, despite the unconstitutionality of some of the provisions in the 1982 census law, is only a minor part of this momentous decision of Judge Bends and his eight colleagues.

West German data protectors are not intensely preoccupied with interpretation of the implications of this decision for their work. The Federal Constitutional Court has in effect given constitutional status to data protection agencies. Their activities have been given strong reinforcement in the public sector by the court's decision, which has also routed the (few) critics and opponents of data protection activities. In April 1984 the Federal Justice Minister announced that he saw a need for a comprehensive overhaul of both the federal Data Protection Law and the 1982 census law. It is expected that, at the very least, the statisticians will now have to separate the collection of data for statistical purposes from the collection of information used for administrative purposes, a lesson that it seems rather late in the day to have to learn, especially given prior experience in other leading Western countries.

West German statisticians at both the state and federal level now have all the more reason to pay very careful attention to the formal and informal advice of data protection officials, even though they have already functioned for years on the basis of special sectoral data protection laws. Although a citizen may have an absolute obligation to furnish data for the census, this can only occur for specific purposes specified by law and with adequate measures to protect against the unlimited collection, storage, use and transfer of personal data. In other words, West German statisticians have learned the hard way, and at substantial cost, that data protection is for real and that it responds to meaningful individual concerns.

A final irony of the 1983 census controversy is that it occurred during the same period when Minister of the Interior Zimmerman was deciding not to renew the five-year term of Professor Bull as Federal Data Protection Commissioner, despite his alliance with the Minister in public defence of the census plans and census law. Bull's successor, Dr. Reinhold Baumann, a career civil servant, took office in the midst of the census controversy in May 1983. If there had been fears that he might reduce the commitment and trust of his office to data protection, his

baptism of fire in the census controversy almost certainly guaranteed that this would not be the case. Moreover, his desire to make less use of the media than Bull dissipated in the course of a year of heavy involvement with the media.

2. SWEDEN

There is more experience with the interaction between statistics and data protection in Sweden than in any other country. The Data Inspection Board (DIB), which began to function in 1974, was the first national data protection agency. The Swedish experience to data illustrates several relevant points. Certain data protectors do have a tendency to take an inordinate interest in statistical activities as opposed to becoming intimately involved in the much more sensitive and important issues of data protection that exist in administrative uses of data. They tend to be shocked at the amount of personal data held by a national statistical agency and poorly informed, at least at the start, about the difference between statistical and administrative uses of personal information. If most data protectors have to choose between doing battle with statisticians, as opposed to the police or a national security agency, for example, the choice seems obvious in terms of relative political power. It is also possible, based on the Swedish experience, that data protectors will have their own views on the conduct of statistics, which may come as something of a surprise to professional statisticians. The views of Jan Freese, the Director General of the DIB, are a case in point; he has had both a peculiar preoccupation with Statistics Sweden and strong views on the conduct of statistics (10).

The Swedish experience leads to the perhaps elementary point that national statistical agencies have to develop their contacts with data protectors over time and with some political sophistication. Edmund Rapaport, a senior official of Statistics Sweden, has worked hard, and with considerable success, to professionalize the relationship between his agency and the DIB. One result is that the main contacts between Statistics Sweden and the DIB are regularized at the working level and thus defused of emotional content. The Swedish model illustrates how, after initial problems, the relationship between statistics and data protection can become professional and relatively harmonious; discussion of a notable exception, planning for the 1985 census, follows below.

The DIB and Statistics Sweden have certainly worked out a mutual accommodation of their respective responsibilities (11). Censuses have been carried out with much less controversy in Sweden, since the DIB came into existence. The statistical agency can now tell the general public that the data protectors have put their good housekeeping seal of approval on census questionnaires and census practices. The successful conduct of the 1980 census of population again points to the positive benefits of consulting with the DIB about such an activity. In 1978 DALK, the commission on revision of the data protection law, ruled against the automatic exemption of statistical registers from licensing provisions of the 1973 Data Act (12). The amendments to the Data Act, effective July 1 1982, appear not to have changed the status of Statistics Sweden under the legislation.

Nevertheless, some outstanding issues remain, which illustrate, among other things, the level of detailed problems at issue between statisticians and data protectors. Statistics Sweden wants to prolong the time period, with the permission of the DIB, for which the statistical agency is permitted to keep primary data on individuals on magnetic tapes. For one of these applications, which is very sensitive because it concerns tax assessment data, some prominent researchers and the two largest Swedish trade unions have written directly in support of the agency's applications (13). In a related area, the Official Secrets Act of 1 January 1981 still permits the disclosure of primary statistical data on individuals to courts of law under special circumstances; Parliament has requested the government to set up a commission of inquiry on this particular question, which deviates from the customary practice of building an absolute barrier around statistical data.

At this point in my discussion of Sweden, some of you may be wondering how I plan to explain the huge fight that broke out in the summer and fall of 1983 about the conduct of the 1984 census of population and housing, especially since it led to some heated exchanges and differences of opinion between Statistics Sweden and the DIB. Perhaps there are no simple explanations, but simply some lessons that the statistical community can learn. First, I would emphasize the issue of international contagion, especially within Western Europe. Newspapers like to write about problems with censuses, wherever they may occur. Whether consciously or sub-consciously, the West German debates of the spring of

1983 fed the Swedish outbreak. The press also capitalized on the proximity of the year Orwell made famous. Secondly, it did not help to calm matters that Professor Sten Johansson, a sociologist at the University of Stockholm and a leading Social democratic thinker, became Director General of Statistics Sweden on August 1, 1983. Johansson, as a leading social researcher and user of data, had long been forthrightly critical of the utility and even need for the DIB.

Lest you think it irrelevant to focus on personalities in this fashion, I want to emphasize that the issue is of the utmost relevance to understanding the relationship between data protection and the statistical community. Much of my current research concerns the ways in which data protection can be made as effective as possible, the personalities, political adroitness, and degree of commitment of data protectors and their staff members are critical ingredients in achieving success. Jan Freese stands out among data protectors internationally because of his longevity in the field, his influence, and his capacity as a publicist. He has given data protection a higher profile in Sweden than in any other country, but he also has what can only be called a special preoccupation with statistics. Thus when Freese and Johansson locked horns over the 1985 census, the stage was set for what in fact became a series of head-to-head debates across Sweden through the fall and winter of 1983-84, which were even exported to other countries in speeches and magazine articles. They debated not only the proposed census, but whether or not Statistics Sweden even needed to be subject to the Data Act, the statute which created the DIB, and whether the DIB should even exist, given the competing problems in Swedish society.

The case of Statistics Sweden and the 1985 census is especially interesting, because the agency is making fascinating efforts to take the census primarily on the basis of existing personal registers rather than requiring the population to undergo the periodic burden of filling out extensive mandatory forms. The Swedish refer to such a census as FOBALT. As a "paradise for personal registers", to quote Freese, Sweden is probably among the very few countries in which such a scheme could even be contemplated. Yet when Statistics Sweden proposed such a census to the

government in January 1983, it was responding to governmental and parliamentary instructions dating back to 1979. It was not until July 1983 that Svenska Dagbladet, the conservative Stockholm newspaper, discovered the 1985 census proposal and raised a big fuss about it, once again demonstrating the vital role of the press in generating both good and bad debates about data protection.

A difference of opinion exists as to whether the DIB was really opposed to the plans for FOBALT before the public eruption in the summer of 1983. Freese has stated publicly that he was surprised at the outbreak of the debate (14). But thereafter, he argued against the census process on several grounds, especially the fact that administrative files would have to be enriched with other personal data in order to make FOBALT possible. He argued that the traditional manual way of taking the census was cheaper, which is an example of the extent to which data protectors can choose to second guess almost any aspect of the conduct of statistics. Freese, in particular, wanted individual members of the population to give informed consent for the record linkages necessitated by FOBALT, a proposal that Statistics Sweden opposed on the grounds of cost (15).

Fortunately, statistical and data protection decisions and opinions in Sweden are subject to the decisions of the government, which decided in December 1983 that the next population census would be modelled on the traditional 1980 census. FOBALT will be delayed at least to 1990. It is again relevant to note that Freese's own general political views are not as close to the current government's as those of Johansson.

Freese views the debate over the census as "a symptom concerning the computerized information society as a whole. The hurricane [debate] made the DIB feel stronger. And perhaps the DIB tried not only to protect the citizen's privacy but also the statisticians" (16). He also remains opposed to any revisions to the DA, as proposed by a government statistics planning committee, which would allow Statistics Sweden to process any personal information, including record linkages, without the permission of the DIB. Freese has called this idea "incredible" in light of the public debate about FOBALT. Given the continuing public anxieties about both privacy and statistical activities, it

would be very unwise, in my view, for national statistical agencies in any Western country to try to function outside the protective umbrella of national data protection laws.

3. CANADA

The most relevant issue in Canada in recent years has been the thwarted effort to breach the absolute (and essential) wall of confidentiality erected around Statistics Canada by section 16 of the Statistics Act. It illustrates, to my mind, the continuing pressure in every country for access to all kind of data bases for new administrative purposes, despite the existence of general data protection laws, such as the federal Privacy Act of 1982 in Canada. One of my reasons for raising this issue here is to encourage further discussion of strategies by which data protectors and government statisticians can resist such efforts successfully.

On May 18, 1983 Solicitor General Robert Kaplan, a Cabinet Member in Pierre Trudeau's long-lived Liberal government, introduced in the House of Commons Bill C-157 to establish the Canadian Security Intelligence Service (CSIS). A storm of controversy arose over this proposed legislation, which created a civilian security agency completely separate from the Royal Canadian Mounted Police (RCMP). As part of sweeping investigatory power the the CSIS, Bill C-157 contained an extraordinary override provision in clause 22(1), whereby "notwithstanding any other law", the Director of the new Security Service could seek a judicial warrant to "obtain any information, record, document or thing..." With such a warrant a representative of the CSIS could "search for, remove or return, or examine, take extracts from or make copies of or record in any other manner the information, record, document or thing". No other country I have studied, to my knowledge, has a comparable provision.

The Canadian press quickly realized, and Kaplan acknowledged, that Bill C-157 would allow the Security Service access to data from the censuses of population that are intended to be absolutely confidential under the federal Statistics Act (17). The Act states that only employees, sworn to secrecy under section 6, may examine census data and that disclosures by such persons of any identifiable information are prohibited (18). The Chief Statistician of Canada indicated that such a provision would be contrary to the base interests of Statistics Canada, which had not been consulted on the matter. The McDonald Commission, which had conducted a wide-ranging inquiry into abuses of power by the

Security Services of the RCMP, had previously recommended that census information should remain off limits to the new civilian Security Service under any circumstances (19).

Like many other Canadian newspapers, the Toronto Star on June 5, 1983 editorialized against making census data available to the Security Service, on the grounds that it would undermine the accuracy and honesty of responses from individuals to census questions: "1984 is only six months away anyway; Kaplan should not rush it". Given the historical tradition of absolute confidentiality for census data, it is not surprising that Bill C-157 produced such an outcry in the press; it is perhaps more surprising that the framers of the Bill thought they could obtain such a provision, although they had a natural inclination to equate census data with many other types of sensitive personal data held by the federal government, such as the tax returns of individuals.

Critics of the proposed practice had to carry their fight to the federal Cabinet, which had to decide on changes proposed to Bill C-157. Although an argument could be made that the Security Service would not find access to data at Statistics Canada a very important source of information, (which makes it all the more odd to seek such access in the first place), the statistical agency does possess, for example, the latest population census ("information on the social and economic characteristics of every person in the country"), a Personal Income Tax Data Base ("all persons who file income tax returns for each tax (calendar) year"), an Integrated Vital Statistics Data Base ("integrated information obtained from the universal system for registration of births, marriages, deaths and still births"), and specific data bases on 200,000 registered nurses, elementary and secondary teachers (all, except Québec), all university faculties, and all students currently enrolled in a university (20). The amount of personal information held by Statistics Canada, and the compulsory methods often used to collect it, explains the necessity and desirability of the historic development of an absolute prohibition against the disclosure of identifiable personal information under the secrecy clause of the Statistics Act (21).

It is most likely that the Canadian Federal Government responded on this specialized issue more to pressure from statistics Canada and the Cabinet Member responsible for it, the Minister of Supply and Services, than to public or media

pressure, although media attention and support was probably essential to sensitizing politicians to the issue. The question of creating a civilian Security Service was very controversial in general. The issue of access to census data appears not to have arisen in Parliamentary committee discussions of Bill C-157, prior to the re-drafting of the Bill by the government to eliminate access to census data. The new version, Bill C-9, introduced on January 18, 1984, declared that a warrant would not be issued in respect of data subjects to the Statistics Act (22). In the Globe and Mail, Jeff Sallot explained that "Statistics Canada officials lobbied successfully behind the scenes in recent months to keep their records confidential. The statisticians argued that the reliability of the information citizens provide would decline if there were no iron clad legal guarantee that it would remain off limits to other government agencies" (23).

One significant point that may be drawn from this incident is that the confidentiality of census data is perceived to be a political issue, since the census process generates private data that people do not want known outside of the statistical agency. It is also very important to recognize that Martin B. Wilk, the Chief Statistician, managed to sell the message internally to relevant Ministers, the Cabinet Committee on Security and Intelligence, and ultimately the Cabinet as a whole, that data held by Statistics Canada has to be handled and protected differently from all personal information held by the federal government. This should be recognized, in my view, as a considerable success for the statistical community in avoiding one more attempt to gain access to statistical data from administrative purposes.

4. CONCLUSIONS

I have no difficulty in agreeing with the general view that strong traditions of concern for confidentiality exist in national statistical agencies. As noted earlier, specialists like yourselves hold this perception but, because of general scepticism about governments, it is not shared by the general public to the same extent. For such a reason, statistical agencies have to be continually vigilant about how their data collection activities in particular may be

perceived by the general public and by politicians from the perspective of continued public concerns for the protection of privacy in an information society. Statisticians need to keep up the fight to maintain a clear functional separation between research and statistical and administrative uses of personal information.

I believe that data protection measures are necessary for all personal information systems in the public sector, including data collected and used for statistical purposes. Given public sensitivities about statistics, data protection should be a significant way to reassure respondents that their privacy interests are being carefully and adequately protected.

A primary goal of general data protection is to produce laws, principles, and information practices suited to each type of personal information system. In fact, the confidentiality provisions in most statistical and census laws are good examples of the kind of specialized measures for data protection that inevitably become the second generation in any data-evolving protection system for the public sector.

Detailed practices for data protection need to be worked out for statistical data from the general principles of fair information practices that can be found in the typical legislation. As illustrated by the Swedish experience, it does take some time for statisticians and data protectors to learn to work with one another: if it is not a marriage, it is at least a form of living together. Statisticians, if I may be forgiven for saying so, are much like academics in their lack of desire for outside supervision of any sort. Such an attitude towards data protection is not constructive. The existence of massive amounts of personal data used for both research and statistical purposes worries people; they ask legitimate hypothetical questions about what could happen, under certain circumstances, to promises of confidentiality and to statutes guaranteeing absolute confidentiality (24).

In my (academic) judgement, statisticians need to be or to become good politicians, in the sense of having political judgments, in order to live in comfortable coexistence with data protectors, the general public, and professional politicians. It is not enough to trumpet one's current and past virtues in order to maintain public confidence in statistical activities; not the least of the reasons why is

that the general public has as little appreciation of statistical work as they have of research activities at universities. My advice is the same for data protectors: confrontational measures are a recipe for disasters. Data protectors have to make the effort over time to understand the world of statisticians and their normal practices; statisticians have to try to justify practices in handling personal information, such as the use of imputation, which may be hard to explain, to rationalize, or to sell successfully in a world more sensitive to the preservation of personal privacy than the past has been (25).

One can predict problems for national statistical agencies when any country adopts data protection legislation for the first time. The United Kingdom is the latest prospect for the bemusement of experienced observers of survivors of previous wars. Problems will occur in the start-up phases, but experience in other Western European nations indicates that they are survivable, especially given some good judgement and sound common sense on both sides. Since statisticians can take some comfort that they are in fact starting from a base of fair information practices that has long been in existence, the likelihood of long-term battles with data protectors seem remote.

Perhaps my final words should be for data protectors, who have to learn to be both political and professional in their dealings with the statistical community. The most experienced data protectors, such as the West Germans, have developed staff specialization in areas such as statistics and good working relationships with the regulators. The professionalization of the relationship can only help both sides. There are a lot of difficult detailed issues at stake in the relationship between data protection and statistics. My interests as an historian, for example, are directly challenged by the movement to anonymize census returns after a period of time or to destroy personal data. Although I am not enough of a specialist to have the answers to specific questions, I am fully persuaded that they can be settled in a rational manner by a process of private discourse between

statisticians and data protectors. It is also inevitable, in our democratic societies, that irreconcilable points of view have to be fed into the political arena for final decision. It is a matter of opinion whether it is more difficult to explain the needs of statistics or of data protection to politicians - both camps are representative of minority interests for a good part of the time. Thus the incentives are for statisticians and data protectors to settle their specific problems within the broad context of the legislative framework governing their respective spheres of activity.

Endnotes

* I wish to acknowledge the generous financial support for my research from the Ford Foundation on the Social Sciences and Humanities Research Council of Canada. I am also most grateful to the numerous persons who have allowed me to interview them and thus guided my research findings.

(1) London, Mansell, 1979.

(2) See David H. Flaherty, "Limiting Governmental Surveillance and Promoting Bureaucratic Accountability: The Role of Data Protection Agencies in Western Societies", a Paper prepared for delivery at the 1984 Annual Meeting of the American Political Science Association, Washington, D.C., August 31, 1984. The current literature on this topic can be found in annotated entries in David H. Flaherty, ed., *Privacy and Data Protection: An International Bibliography* (London, Mansell, 1984; and White Plains, N.Y., Knowledge Industry Publications, Inc., 1984).

(3) *Der Bundesbeauftragte für den Datenschutz, Dritter Tätigkeitsbericht des Bundesbeauftragten für den Datenschutz* (Bonn: Bundestags-Drucksache 9/93, 1980), Section 2.6.

(4) *Ibid.*, Section 3.4.

(5) *Ibid.*, Section 3.4.1.

(6) Further details and informed discussion can be found in Frank Kuitenbrouwer, "German Census Postponed on Court Order", *Transnational Data Report*, VI, No. 4 (June, 1983), 182-83.

(7) *Ibid.*, p. 182. As I wrote in 1979: "Observers of other countries might expect the use of census data by the Gemeinden [local communities] to become a source of

public controversy..." (Privacy and Government Data Banks, p. 146). The Federal Statistical Office was relatively unreceptive to representations from the staff of the Federal Data Protection Commissioner about potential problems with data protection in the census. In its first annual report in January 1979, the BfD indicated that it did not like the use of the census to correct local population registers. (Datenschutz. Erster Tätigkeitsbericht des Bundesbeauftragten für den Datenschutz [Bonn: Bundestags-Drucksache 8/2460, 1979], section 3.3.5)).

- (8) International Herald Tribune, March 28, 1983, p. 2.
- (9) See the informed discussion of the decision, and a listing of its guiding principles, in Jost Schindel, "Germany Declares Self-Determination Over Personal data", Transnational Data Report, VII, No. 5&6 (August/September, 1984), 359-61.
- (10) See Flaherty, Privacy and Government Data Banks, pp. 113-14.
- (11) In its early years the DIB devoted an enormous amount of attention to licensing and supervising the great number of personal registers maintained by Statistics Sweden. I have described this process in some detail in Ibid., pp. 116-19.
- (12) DALK Report on Revision of the Data Act, Sou 1978: 54, pp. 352-54.
- (13) Communication, Edmund Rapaport, Statistics Sweden, February 7, 1980.
- (14) See Jan Freese's speech on "FOBALT" at the meeting of the International Federation of Information Processing Societies in Paris in September 1983; an article on the Swedish census, which he wrote for a West German magazine under the date of February 22, 1984; and "Swedish FOBALT Census Opposed", Transnational Data Report, VII, No. 4 (June/July, 1984), 200.
- (15) All of these points are developed in Ibid.
- (16) Ibid.
- (17) Bill C-157, s. 22(1); Jeff Sallot, "Spy Agency to See Secret Census Data", The Globe and Mail, June 3, 1983, pp. 1-2, The Globe and Mail regularly carried stories on Bill C-157 from May 19, 1983 until the end of June.

- (18) Statistics Act, S.C. 1970-71-72, c. 15, s. 16(1).
- (19) Commission of Inquiry Concerning Certain Activities of the Royal Canadian Mounted Police, Second Report, volume 1, Freedom and Security under the Law (Ottawa: Minister of Supply and Services Canada, 1981), pp. 165, 587; "...the tradition in this country has been very strongly in favour of complete confidentiality of census returns. The unqualified guarantee of confidentiality helps to overcome the reluctance of Canadians to respond to inquiries about personal matters for purposes which may be suspect, or at least not clearly understood, by many."
- (20) Canada, Personal Information Index 1983 (Ottawa, 1983), pp. 215-28.
- (21) See Flaherty, Privacy and Government Data Banks, pp. 178, 191-199.
- (22) Canadian Security Intelligence Service Act, S.C. 1983-84, c. 21, s. 21(3). Privacy advocates oug this measure into the tax information field in particular.
- (23) Jeff Sallot, "New security bill puts tighter rein on investigators", The Globe and Mail, January 19, 1984, pp. 1-2,
- (24) For example, the U.S. Bureau of Census continues to be attacked for its release of aggregated data on West Coast Japanese during World War Two. The most recent discussion is in David Burnham's, The Rise of the Computer State (New York: Random House, 1983), pp. 20-26.
- (25) The Swedish debate over imputation is described in Flaherty, Privacy and Government Data Banks, pp. 118-19.

DISCUSSION

Mr. SCHOETTL (*Conseil d'Etat*) opened the discussion:

He expressed general support for the views of Professor Flaherty. He referred to the fact mentioned by Mr. Begue that, in a recent survey in France, almost half of those questioned believed that censuses of population data were available to the tax authorities and police and this despite the firm legal guarantees of complete confidentiality operating over many decades of censuses in France. What was needed was not a further strengthening of the legal provisions regarding confidentiality but rather an active campaign to inform and interest the public in the census. The circle of users of census data should be extended to include writers, schools and universities, trade unions and commercial enterprises. Public curiosity should be aroused on topics involving international comparisons, e.g. numbers of women in employment. The importance to the State of good statistical data should be emphasized and contrasted with the lack of public spirit of people who refuse to supply the necessary information. It should be pointed out that the taxpayer has a right to expect government decisions based on sound information, that a sick person has a right to medical research, that the peaceful citizen has a right to crime prevention measures based on proper knowledge of the structure of society and generally that people in some way discriminated against, had the right to have their grievance demonstrated by official statistics.

Mr. SUDFELD (*Statistisches Bundesamt*) remarked:

The law covering the 1983 Census of Population in the Federal Republic of Germany was being considered, politicians, statisticians and data protection officials were one hundred per cent in favour. He considered that statisticians and data protectors were neither allies nor opponents, but simply groups with a concern in common. It

was clear that a much sharper division would have to be made in the future between statistical inquiries and the checking of administrative registers even though this might involve considerable extra expense.

The author replied to the debate:

He particularly welcomed the establishment of a Standing Committee on statisticians and data protectors in the Federal Republic of Germany and remarked that, if this committee had existed earlier, the problems with the 1983 census might have been avoided.

PROTECTION OF PRIVACY, AUTOMATIC DATA PROCESSING AND
PROGRESS IN STATISTICAL DOCUMENTATION

G. CARIANI, G. GIACUMMO, P. LA CAVA
Istituto Centrale di Statistica

0. INTRODUCTION

A feature of the present-day world is the production, dissemination and consumption of information on a scale, in a variety and with a speed hitherto unprecedented. At the heart of this phenomenon is the use of mass communication media such as the radio, the cinema, television and the computer. The demand for, and production, supply and consumption of, information have thus become part and parcel of every-day life, from whence the "right to information" has been recognized as indispensable in the technological age.

Technological progress in our lifetime has, however, introduced a new element in the field of information, namely that information can now be processed automatically, which explains the term "informatics", first coined in French as "informatique" by Philippe Dreyfus in 1962.

The advance has been such that reference is now made to the computer's "artificial intelligence". What we have in fact is an electronic extension of human intelligence, enabling man to accumulate a vast store of information in minimal space and time. What is more, this information can be transmitted immediately to requesters via a network linking the mainframe computer to peripheral terminals: this is what is known as a telematic system. The computer is thus the hallmark of today's mass technology society. The very fabric of State administration is today considered by the new generation of political scientists as a large network of communication and processing of computerized messages. A gradual "computerization" of private life has thus taken place, not only as regards the number of subjects on whom information is stored, which could in future be exhaustive, but also as regards the increasingly detailed, specific and accurate nature of this information.

Recently, however, not only have we witnessed a proliferation of State (and other) consumers or "owners" of information, but since the use of computers has become increasingly widespread the methods and the quality of data processing have fundamentally evolved, producing changes

affecting the relationship between central administration and the citizen. The new frontiers of technology determined by microelectronic and telematic systems seem destined to modify still further that relationship and social relationships in general.

Much has been and can be said about the implications which new information technologies hold for the management of political systems, with the inevitable attendant and dangerous encroachment on individual freedoms. The question which this begs, however, is whether these new technologies are a boost or a threat to democracy.

A quick look at experience in this field to date reveals undeniable interference between information technologies and the management of a political system.

Use of the computer by central government to collect and process a growing corpus of information on persons has been seen to arouse concern for the violation of privacy in that it generates an acceleration of the authoritarian processes whereby citizens can be kept under surveillance, with all that this entails as regards the organization of political life. This way of perceiving the problem is of mainly European stamp.

On the other hand, the findings of research conducted by American teams are less drastic and suggest rather an evaluation of the extent to which computers affect general government. While boosting efficiency, particularly in the traditional sectors, they alter its effects and socio-political implications. Whatever the case may be, as regards the political system specifically it is difficult to arrive at any final conclusions on the basis of the information we possess. The only certainty is that the computer is seen mainly as a means of enhancing the potential for governing complex societies, the stress - rightly so in many instances - being laid on the importance of systematic organization.

1. AUTOMATIC DATA PROCESSING AND VIOLATION OF PRIVACY

The natural tendency of present-day community-based society is for the reference data in the electronic memory to constantly increase. This artificial collective conscience of today's society in turn leads to what is in statistical terms an establishment of the lowest common denominator of

individual operations, which are thus taken out of the realm of confidentiality, while at the same time exposing the latter to the risk of violation.

With the present state-of-the-art technology, privacy can be violated on three levels:

- a) physically, by using new optical and acoustic recognition methods;
- b) psychologically, by means of the various investigation methods available today to get information from an individual which he would not under normal circumstances divulge or which he gives quite innocently, thus unwittingly divulging certain aspects of his private life;
- c) by crosschecks which can be carried out by collecting, comparing and aggregating a large number of detailed data, via the computer.

Being indirect, the latter today represents the most insidious threat to privacy. The computer is a totally depersonalized tool; thus, any individual "on electronic record" can be kept under constant and unsuspected surveillance concerning important aspects of his private life. The variety and multiplicity of these aspects would in previous times have ruled out the possibility of such a documentary check. But today, for instance, the mere purchase of an air ticket involves the recording of dates, places, travel partners, etc. And the system is expanding because of the need to use such electronic means even for booking hotel accommodation, making telephone calls, carrying out bank transactions and all other social relationships of citizens.

The private life of an individual could be kept under check in any country which keeps a health register for the purposes of a national health scheme. The same is true of tax and vehicle registers.

Further substantial violations of privacy could arise in the wake of the emerging potential of large information processing systems.

State-of-the-art systems to some extent offer, at least theoretically, linkups between available files, thus giving the potential user the ideal opportunity to piece together the various items of information relating to an individual. The sequence of the various events thus makes it possible to reconstruct in the right order, using any logical system (e.g. temporal), the "life" of subjects under observation.

At present the linkup potential between different files, even when these fall within the context of public agencies, comes up against substantial hurdles both from the point of view of compatibility between the different hardware systems used and on account of the lack of standardization in the collection of basic information. Problems also stem from the use of software which is frequently not standard.

However, these problems are by no means insurmountable.

In certain countries, for instance, some of them have already been solved by means of a personal identification number. Furthermore, there is no doubt that technology will offer increasingly sophisticated hardware with increasingly standardized applications languages which will overcome any remaining linkup hurdles.

Progress accomplished in telecommunications technology and cable transmission and the enormous capacity of large computers together point to the likelihood of substantial changes in the future lives of individuals. We are already witnessing some of these applications, noteworthy examples of which are electronic newspapers, remote medical diagnoses and consultation systems, etc. The future will bring a whole host of other possible telecommunications applications, by means of VDUs installed in our homes. Other examples in addition to those already mentioned (bank transactions and all types of booking arrangements) are home security systems, sales and opinion polls. Not much further beyond the present horizon may also be the decentralization of certain occupations which can be carried out at home.

These are just a few examples of what a telematic system can do, and in future a whole range of other services will probably be available through interactive home systems (hometec).

Information systems like this could generate such a substantial and detailed flow of information as to seriously jeopardize the privacy of any private citizen, if it is borne in mind that information of apparently no great intrinsic value can, when put together, enable any user to map out the profile - ideology included - of any citizen.

2. THE RIGHT TO INFORMATION AND AUTOMATIC DATA PROCESSING FREEDOM

The computer thus makes it possible to get hold of information about the individual and his private life and thus subject him to a new form of social domination, which

could be called "information" power. For precisely this reason, the right to confidentiality has taken on a new dimension which it could not have had a century ago when it was first introduced.

Confidentiality of personal information had until quite recently an undoubtedly negative connotation; it prevented others from invading privacy and especially from acquiring information for subsequent dissemination. In a word, the right to confidentiality was equated with the accepted definition of privacy in the sense of "to be let alone".

This concept has changed over the years and alongside the right to be let alone the right to information has become established, i.e. the right each individual has to inform and to be informed. In the wake of this development the former has evolved and has had to adapt to the latter approach in order to ensure that, for the common good and therefore the good of the individual as well, third parties could have a way of looking into his privacy. So from a negative concept and a simple prohibition, the right to confidentiality has been transformed into an active power which enables the citizen to monitor the use which others make of information concerning him. This means that the individual can no longer object to certain information concerning him being acquired by other public and private subjects and made generally known. He can, however, monitor the use which others make of such information and take steps to uphold the accuracy of the information or forestall misuse.

This right to monitor one's own personal data stored in an electronic file is part of the "right of information" and, when it comes to data processed automatically, constitutes a right to "informatics freedom": the freedom of any individual to decide when, how and to what extent information concerning him can be communicated to other people. This freedom is today acknowledged for individuals, groups associations and agencies in general.

The foregoing considerations highlight how it has become necessary to set limits to "informatics power" in order to forestall any possible violation of privacy.

This protection can be provided by guaranteeing the "security" of data processed automatically and, in particular, by introducing sound legal protection for confidentiality.

3. THE SECURITY OF AUTOMATICALLY PROCESSED DATA

The first requirement as regards the confidentiality of personal data processed automatically is their protection from persons not authorized to access, process, change and disseminate them. The first thing is therefore to establish rules concerning access to the data.

From this point of view, protection can first and foremost be physical, i.e. relating to the hardware and the dataware. The premises must be safeguarded by traditional methods and access is by advanced technological means of control and recognition. These include:

- requiring the authorized user to first insert a magnetic badge;
- to identify the person by means of an enlarged recording of fingerprints or handprints (length of fingers, translucency), or by voice pattern or signature analysis.

Protection is also logical in character, i.e. related to the item of data stored in a computer. All electronic files or data banks are structured or being structured so that protection of information is taken care of at the design state.

Modern data banks physically comprise a very large series of basic data which, to be combined into intelligible information, require sets of codes identifying the logical properties of each item of data within the bank itself. This means that only the computer can, by means of specific programs, trace all the basic data which in combination provide the answer to a given question. This naturally implies that any unauthorized access would entail knowing not only the access key to the computer but also the structure of the file, and this structure, as has already been pointed out, will have had protective devices designed into it from the outset.

Access keys were mentioned in the last paragraph. These are specific protective devices external to the data bank and which differ according to whether they are expressed in terms of software or hardware.

They may be passwords or badges. Both these protection systems are intended to restrict at various levels the right of access to the data bank and to subsequently resolve the issue of the level of restriction placed on the data in a given file.

Another solution is to authorize access to certain terminals only. The computer can thus vet the interrogator's right or non-entitlement to access the base and cut the user off should the latter attempt to identify a password or should an unauthorized badge be introduced.

Data confidentiality must also be safeguarded when files are transferred from one computer to another. Substantial progress has been made in coding: the substitution of letters by symbols has given way to the use of numerical-key combinations and in order to get round the tedious computer runs needed to encode an outgoing file or decode an incoming file, specialized microprocessors have been manufactured which, installed at the input and output stages of the computer, can encode or decode with a minimum of delay.

This gives some idea of what can be done from the informatics point of view to protect data and thus the privacy of the citizen from outside interference; in conjunction with hardware and software incompatibility between the managers of various files, this should allay certain misgivings as to the ease of unauthorized access. But it also shows that it is not on the technical side that the solution whereby the citizen can be shielded against any violation perpetrated by the managers of the data banks themselves is to be found.

As was mentioned earlier, hardware and software incompatibility between files is a common problem, but we are well aware that these problems are - not easily, admittedly - surmountable and the hurdles which have prevented electronic superfiles from being set up have not been technical ones.

4. LEGAL PROTECTION OF CONFIDENTIALITY

In the more industrialized nations some of the situations and certain of the risks of violation of privacy mentioned earlier have become a reality. Efforts have therefore been under way for some time to bring the confidentiality of personal data stored in the various data banks under legal control.

The essential instrument for safeguarding the personal data kept on natural persons (legal persons or also *de facto* undertakings) is the law. It is therefore up to the legislator to adapt legal guarantees on the inviolability of the individual's privacy to trends in the development and

utilization of the new automatic data processing systems summarily described above. He must guarantee the individual the above-mentioned right to control, i.e. the right to know what has been stored in the data bank concerning him and what use might be made of this information, a right which must take into consideration *inter alia* the facility for correcting inaccurate information and any information which could somehow or other infringe personal freedom.

In any event, it should always be left to the judicious appreciation of the legislator to establish an equitable reconciliation of differing interests, the individual's to see his own privacy safeguarded and the collective one to acquire information on persons with the aim of catering for the interests of the community at large, which is the only consideration which would justify sacrificing the former for the attainment of the latter.

It should also be noted that for obvious reasons wholesale prohibition to acquire and disseminate personal information is in today's world no longer conceivable and, in certain cases, not even necessary. There exists in our modern society certain information which, although relating to the individual, is in essence of the public domain (e.g. data in public registries which may be freely consulted by anyone) and the dissemination of which does not constitute any violation of rights and can therefore be permitted without the need for special authorization.

The above considerations militate in favour of a legal framework for the information contained in the data bases, tailored in each case according to how deeply this information penetrates into privacy and according to the different types of information which could be categorized as follows:

- a) information of a statistical/scientific nature, including data relating to health matters;
- b) information of a public nature concerning security and order within the State and information of other types;
- c) information which concerns the private sphere in a more specific manner, e.g. economic, commercial and similar data.

Where there is no possibility of identification, the information referred to in a) can be processed with more freedom than the other types of information. That referred to in b) should be covered by specific regulation guaranteeing maximum caution as regards dissemination and maximum protection for the citizen when it comes to

utilization geared to the good of the community at large, safeguarding him from misuse of personal data. Lastly, the law should provide stringent control over the management and publication of the information referred to in c).

This sector is as yet uncovered by any specific set of regulations in Italy and even the Constitution seems to have overlooked the very concept of confidentiality, except for the reference in Article 13 to personal freedom and breach of the same by "inspections" not provided for by the law. This concept of inspection may in fact cover investigation of the various physical and moral aspects which go to make up the profile of an individual.

Apart from this indirect reference to the constitutional requirement, there are various other provisions which tackle the problem from the point of view of specific and sectoral aspects. For instance, there are the rules governing official secrets (1), according to which employees and public officials must maintain the strictest reserve concerning administrative arrangements and operations or information which come to their attention in the course of their duties, and the rules governing State records (2), which prohibit for a pre-established number of years the consultation of documents relating to foreign and domestic policy or the private situation of individuals.

A first explicit formulation of the right to confidentiality is, moreover, contained in the Statuto dei Lavoratori (legal text on worker's rights)(3), where the employer is prohibited from using audiovisual devices to monitor the activities of workers and from screening their political, religious and trade union opinions or aspects of their life which bear no relation to their working activity.

-
- (1) Penal Code, Article 326 (divulging of official secrets) and Article 15 of Presidential Decree No. 3 of 10 January 1957 laying down the Consolidated Text of the Provisions on Regulations Governing Civil Servants.
 - (2) Article 21 of Presidential Decree No 1409 of 30 September 1963 on the Rules on State Records.
 - (3) Law No 300 of 20 May 1970, Articles 4, 5, 6 and 8.

Mention should also be made of a law issued by the Regional Government of Lombardy on "provisions governing the regional information system" (4), which lays down the rules governing the regional information network, covering also data banks and their legal regulation.

A first sound step to bridge this yawning legal gap has been made thanks to the more systematic provisions - albeit geared to the requirements of a specific sector of public life - contained in the new rules governing law and order (5). Although this set of provisions sets out certain valid solutions to the problem (type of data collected, criteria for access, monitoring of activities of the EDP centre), it can by no means be considered as exhaustive even for that specific sector.

In particular, Article 8 of this law provides for the establishment of an EDP centre at the Ministry of the Interior, obliging "every government body, agency, enterprise, association or individual setting up or holding for any reason magnetic files in which data and information of any type whatever are stored on Italian citizens, to declare the existence of such files to the Ministry of the Interior by 31 December 1981".

This Ministry had by that deadline received roughly 100 000 declarations concerning the existence of computerized files. This figure is symptomatic of the scale and complexity of the problem, which will have to be tackled by the legal system, bearing in mind that the number of data banks is likely to increase in the future as computers become more widespread.

It should be pointed out at this juncture that legal protection of confidentiality should also be provided for the transfer of data from one country to another, *inter alia* to ensure the necessary free flow of information across frontiers.

An initiative intended to cater for this need, at least among the Member Nations of the Council of Europe, was the Convention approved on 22 September 1980 by the Committee of Ministers to "protect individuals *vis-à-vis* personal data processed automatically". This Convention stipulated the guidelines with which all States signatory to this Convention must comply, thus at the same time ensuring equality of treatment and freedom of exchange of information.

(4) Regional Law No 15 of 16 March 1981.

(5) Articles 6 and 12 of Law No 121 of 1 April 1981.

International agreements, it should be noted, tend to extend legal protection from natural persons (individuals) to legal persons (i.e. organizations, associations or enterprises), which are also interested in the protection of their data which, although not of a personal nature, have a claim to be considered as sensitive.

Italy is not yet a signatory to this Convention, as Article 4 stipulates that each signatory State shall undertake to embody in its national laws the provisions necessary to implement the fundamental principles of data protection, as set out in the Convention. Furthermore, these provisions are to be implemented no later than the date on which the Convention becomes applicable in the country. Unfortunately, Italy is lagging behind as regards such legislation.

One may wonder at this juncture what is new in our legal system other than what has already been mentioned. It is high time that the Italian legal system fell into line with the more advanced nations. Three bills have been tabled in the last few years.

The first is the Accame Bill (6) tabled in 1981. The 20 articles of this Bill, rather than constituting an actual practical instrument, describe the guidelines of protection of privacy in the context of handling and automatic processing of data and information.

The second is the Picano Bill tabled in 1982 (7) and concerns the regulations for the protection of the right to confidentiality of natural persons *vis-à-vis* automatic data processing. It explicitly provides for a State supervisory body of the type "National Committee for Informatics and Freedom" answerable directly to Parliament.

The same Bill provides for a detailed list of physical and logical safety measures capable of safeguarding data banks.

The third and final bill is the Mirabelli Bill of 1983 (8). This is a ministerial initiative concerning the fostering and coordination of information activities in general government. Essentially, it calls for safeguarding of the rights of both natural and legal persons and for a supervisory body directly answerable to the Presidency of

(6) Bill No 2552 of 21 April 1981.
(7) Bill No 3195 of 24 February 1982.
(8) Bill No 3331 of April 1983.

the Council of Ministers, and deals with related aspects of data protection and the relevant safety measures, and reciprocity of treatment at international level.

The course to be charted is clear. It can only be hoped that the judiciary, aware of the need to bring this situation under control in good time, will get down as soon as possible to examining the problem outlined above and will promptly pass a systematic law to bring order into the sector at a very delicate phase of its expansion.

5. SAFEGUARDING OF PRIVACY AND STATISTICS

What has been said thus far grows in significance when we come to the sector of statistics in the broadest sense of the term.

As in most countries, statistical confidentiality in Italy is the subject of a special law (9) which ensures that personal data is passed on only in aggregate form or, in any event, in a manner which makes it impossible to identify the subject.

The growth of data banks has nevertheless taken us to a point which is just one step away from the fulfilment of what has long been the statistician's dream, i.e. to link up the different files, most of which are now data bases, thus achieving a completely integrated system of statistics.

As time goes by, however, this dream is slowly fading as misgivings concerning the violation of privacy percolate through public opinion. In recent years we have witnessed, particularly in the more progressive countries, the beginning of widespread rejection by public opinion of all types of statistical surveys.

The assurances given by the powers-that-be have not been enough to allay these fears and uncertainties.

Does this have to mean a retreat despite the potential offered by technology?

(9) Decree Law No 1285 of 17 May 1929, enacted in Law No 2238 of 21 December 1929.

Experience in Italy offers grounds for optimism. The census operations of 1981-1982 in that country did not elicit opposition from either the press or public opinion.

On the contrary, the change which has taken place in the configuration of the State in the wake of the implementation of the regional government system with its attendant devolution of general government and legislative functions, the stepping up of sectoral intervention by central and local government and public agencies, the need on the part of private citizens for a more detailed knowledge of the actual socio-economic situation they have to face, have over the past decade been largely responsible for a further drive to make increasingly detailed information available.

For aspects falling within its sphere of competence, the ISTAT has endeavoured to cater for such legitimate pressure through its terms of reference and using precisely the constraint of statistical confidentiality already referred to.

A recent reflection of this is the legal framework provided by the law on the implementation and financing of the 1981 general censuses (population and housing, industry and trade) and the 1982 census on agriculture (10), whereby local government authorities (regional and provincial governments and the municipalities) so requesting were given "depersonalized" data on individuals.

The data banks which the ISTAT has set up and is at present updating, partly on the basis of the census results, raise similar problems. The most sensitive aspects of this source of information, to which government departments and private citizens now have access, is precisely the establishment of the levels of aggregation of these data bases. A logical level has in fact been determined which on the one hand permits a breakdown to satisfy user requirements and on the other prevents the dissemination of personal data, even when "depersonalized".

The remaining fundamental problem is to reconcile the two aspects: the confidentiality of the information collected and the need to provide personal but non-identifiable data to as many users as possible. While it can in part be solved from the technical point of view by using security and confidentiality methods for on-line dissemination of data stored in the data base, it cannot be solved by means of the legal instruments in force. There is an undoubted need for

(10) Article 2 of Law No 864 of 18 December 1980.

special legal provisions which take account of the corpus of information which the agency possesses and which ensures that optimum use is made of this information.

These provisions should not, of course, be restricted to governing the right to confidentiality, but should embrace the very structure of the ISTAT so that it can fulfil two functions, i.e. to act as a State service producing official statistics and as a data bank at the service of both general government and private citizens.

6. CONCLUSIONS

It was seen earlier that there has over the past few years been an increasing demand for information by both public and private users, while at the same time the statistical services which could cater for this demand are coming up against a twofold obstacle: the escalating cost of statistical surveys and the need not to unduly inconvenience enterprises and individuals with *in situ* surveys.

Typical examples in this context are the general censuses. There is no need to enlarge upon the complexity of the technical, scientific, organizational and political problems which arise before, during and after the census operations. Quite understandably, therefore, there is growing support for the opinion that general censuses are today still useful in countries which are not statistically developed, but prove much less so in countries which are in a position to manage stable information networks. In addition to methodological and organizational techniques which can be used to solve such problems, an alternative to surveys of the census type may therefore be to use registers, files, lists and similar sources, usually of an administrative type, and a whole range of sample surveys. In abstract terms, if there existed comprehensive computer files of personal microdata stored on suitable DP media and updated at sufficiently regular intervals, they could become similar in nature to the censuses and thus take their place.

Moreover, if this materialized, computerized sampling systems could facilitate the setting up of a valid base on which to graft any type of sample survey designed to gain an in-depth knowledge of specific aspects of the different basic universes.

Italy, for instance, has a long-established system of registers of resident population, which are now fully computerized in many of the major municipalities. It nevertheless seems difficult for the time being to use them

as an information base for certain socio-demographic aspects of the population, since their reliability leaves much to be desired, particularly as regards information such as occupation, educational qualification, etc.

There are also other types of files such as those kept by the Ministry of Finance, concerning declarations of income by natural and legal persons, and those kept by the INPS (Italian social security organization) concerning social contributions by wage and salary earners and by self-employed persons. Both these sets of files, however, relate to only part (albeit the greater part) of the Italian population.

Be that as it may, they suffer from certain defects which are typical of any large-scale file, e.g. the effects of an inevitable accumulation of errors, high updating costs in terms of both money and time which rise more than proportionally to the increase in the number of units it contains and of data stored.

As regards information systems based on files, the ISTAT started setting up in 1970, mainly for the purposes of current surveys, a register of enterprises employing over 20 persons. In order to make good the up-dating shortfall which has inevitably built up over the years, the register has been extended to enterprises employing over 10 persons on the basis of the 1981 census of industry, trade, small industry and services.

A similar register is currently being set up on the basis of the 1982 agricultural census. This register will provide the starting point for current surveys in the years ahead.

The management of this type of information system is naturally a function of the quality of the basic files, the degree of computerization of the ancillary administrative structures and the efficiency of the national statistical network. As mentioned earlier, other public agencies and bodies hold information on individuals, households and enterprises in files which are computerized to some extent. The purposes of this documentation are mainly administrative and managerial, but utilization for statistical purposes as well cannot be ruled out.

In order to attain this objective, suitable ways should be found for standardizing definitions, classifications, etc., offering the possibility of linking up the different systems through computerized files.

A drive in this direction would undoubtedly open up a promising prospect for the future and provide a valid alternative to censuses and to many of the current surveys, while at the same time achieving substantial savings of time and money.

Implementing a solution of this kind is by no means a straightforward process in that it involves not only technical problems but also the legal problems inherent in the possible violation of the right to privacy of natural and legal persons.

So while the process of coordination and integration of information systems intended to produce statistics continues and the technical aspects are improved, it is hoped that appropriate legal steps will be taken to protect the sensitive nature of the data managed by the bigger computerized files. This may be the only way in which the use for statistical purposes of files held by the various authorities can be placed on a sound footing.

DRAFT LEGISLATION IN ITALY
ON THE PROTECTION OF PERSONAL DATA

Prof. M. LOSANO
Università di Milano

1. CURRENT SITUATION

While legislation safeguarding the confidentiality of personal data stored in data banks has existed for some time in the major industrialized nations, the situation in Italy is still *sub judice*. Regulations relating to the safeguarding of confidentiality do in fact exist, but these are to be found in laws governing other sectors; a specific law on confidentiality has not yet been passed but there are various bills, one of which is currently being examined by Parliament.

At the beginning of 1984, matters were at a standstill. As bills lapse when the Houses of Parliament are dissolved at the end of the legislative period, the Accame bill (paragraph 3) and the Picano bill (paragraph 4) have lapsed, and will have to be tabled again for debate. The text drafted by the Mirabelli Committee (paragraph 5) did not even reach the bill stage, as the dissolution of the Houses in May 1983 prevented the Minister of Justice from presenting it to them as a government-instigated bill. This bill was finally presented by the Minister for Justice, Mr. Martinazzoli, on 5 May 1984. He submitted the text drafted by the Mirabelli Committee but with some slight amendments. A few months previously another bill on this subject had been presented (paragraph 6). Other bills are being drafted and will be presented in 1985.

The first question which springs to mind is this: given the economic and political crisis which is besetting Italy in the early 1980s, do these bills have any chance of becoming laws soon? One cannot help thinking of similar cases, such as the anti-monopoly legislation imposed on Italy by the Treaty of Rome but still not implemented to this day. In the case of the law on confidentiality, however, we are faced with a different situation. States which already have legislation to safeguard privacy are anxious to prevent their regulations from being undermined by the transfer of data into data banks in foreign States with weaker, or altogether non-existent, laws to safeguard privacy. As a result, foreign legislation nearly always contains a clause prohibiting the transmission of personal data to those

States which cannot offer legislative guarantees equal to those of the State providing the data. Italy must, therefore, adopt a law on privacy in order to avoid being cut off from the transnational flow of personal data. What is more, such a law must not deviate too greatly from the models in force abroad, because discrepancies could lead to the clause prohibiting the transmission of personal data being invoked.

The delay in producing legislation has put Italy in an ambiguous position in its relations with the international organizations of which it is a member. On 22 September 1980, the Council of Europe adopted the "Convention for the protection of individuals with regard to automatic processing of personal data". Italy subscribed to the Convention on 2 February 1983 but has not yet been able to ratify it, even though the Convention compels States which are signatories to promulgate internal legislation designed to put its principles into practice. The Italian authorities have thus signed an agreement, but it is not being implemented (1).

The practical consequences of this ambiguity on the part of Italy with regard to laws on privacy are documented in a legal text promulgated by the autonomous province of Trento. This province enjoys special privileges in the legislative field which grant it wider scope for autonomy than that given to other provinces. Provincial Law No 10 of 6 May 1980 is devoted to the setting up of a provincial electronic information system. In the absence of State legislation on privacy, the autonomous province of Trento has provided itself with a legal basis referring directly to a Resolution passed on 8 May 1979 by the European Parliament, thereby bypassing the national Parliament. Article 7 of the Law states: "in order to verify the legitimacy of anything related to the collection, organization, use and dissemination of data and information, with particular concern for the protection of individual rights, the Provincial Council shall appoint a committee of three members nominated by the Board, one of whom is to represent minorities. This committee shall exercise its own form of vigilance on the basis of recommendations and directives on this subject issued by the European Parliament".

(1) On 23 September 1980, the OECD issued a recommendation entitled "Guidelines for the protection of privacy and transnational flows of personal data". The fact that Italy is a member of the OECD does not, however, raise problems similar to those caused by the Council of Europe Convention, because a recommendation, by its very nature, does not oblige the States to which it is addressed to implement specific legal measures.

This regulation leads us from international relations to domestic policies. Within any country, a law on the protection of personal data is an important guarantee of the citizen's basic liberties. This topic is so vast that we cannot concern ourselves with it here. It is nevertheless vital to point out that, even if it were not interested in providing within its own borders such guarantees of individual freedom, Italy would end up being forced to adopt legislation on privacy because it is part of a particular economic and political context and therefore has indissoluble links with other industrialized States.

Sooner or later and in one way or another, then, we in Italy will also have a law on privacy. It is necessary, therefore, to examine the regulations already in force (even if contained in laws whose purpose is not connected with privacy) and those which are proposed. Having scanned these briefly, it will also be necessary to examine the degree of coordination between existing and proposed legislation and, finally, to see whether the proposed regulations can feasibly be applied.

Since there is no law on privacy, strictly speaking, it seems pointless to dwell in detail on existing regulations and draft laws. That is why these concluding remarks can be only summary and indicative in nature, as this is inevitable in the case of a topic which has not yet been the subject of specific provisions.

2. REGULATIONS ON CONFIDENTIALITY ALREADY IN FORCE

The delay with which Italy is managing to put itself on an equal footing with other industrialized countries as far as legislation on privacy is concerned has forced Italian law-makers to anticipate some forms of personal data protection in the "Statuto dei Lavoratori" (legislation on workers' safety and rights) and, above all, in the public order reform law, which incorporates a veritable mini-law on safeguarding personal data.

This anomalous practice seems destined to spread, being one of the few loopholes available when urgent solutions have to be entrusted to parliamentary majorities that are too small (2).

Regulations of this kind are known as *normae fugitivae*, because, since they are not to be found where they should be, they escape detection and are therefore a source of potentially serious legal problems; indeed, subsequent

(2) In the field of *normae fugitivae*, in addition to the public order text there is another one on investment funds, which itself contains regulations on securities not covered by the usual legislation.

legislation frequently forgets to take account of them and duplicate laws are therefore created which are quite often not consistent among themselves. The link between a future law on information and the "fugitive" text of the public order reform law will be examined shortly. Let us turn our attention now to the regulations on workers' safety and rights (3).

a) The "Statuto dei Lavoratori"

The "Statuto dei Lavoratori" (Italian Law No 300 of 20 May 1970) prohibits, in Article 8, inquiries into workers' political, religious or trade union persuasions, and into any other facts which are not relevant to an assessment of the professional behaviour of workers. Thus, any form of cataloguing of employees, whether by means of traditional files or with personal data banks, is unlawful.

With the advent of computers, the scope of Article 4 has also been extended. This article prohibits the use of audiovisual and other equipment for the purpose of monitoring the activities of workers at a distance. This prohibition was originally devised to curtail the use of closed circuit televisions in firms. However, the way in which the article has been formulated linguistically so as to be deliberately wide means that it can be applied to other computerized activities connected with the setting up of personal data banks. For example, anyone using a processor is recognized by the program by means of an identification code. This code also makes it possible to measure the working hours of an individual and, in addition, to compare the hours worked by several people.

Nowadays, in order to avoid a clash between the technical requirements of identification and statutory regulations, a group rather than an individual code is attributed; in this way it can be shown, for instance, that a maintenance team has been operating for a certain period, but it is not possible from this information to calculate the working hours of the individual members of the team. It could be that a future law on safeguarding personal data will make it possible to relax the prohibition of this form of supervision. If that happens, it will be necessary to amend this point in the "Statuto dei Lavoratori".

(3) For the sake of completeness, I should also like to mention Law No 98 of 8 April 1974, governing telephone-tapping. This has introduced Article 615 a) into the Penal code, entitled "Unlawful invasions of privacy". The provisions, although they are concerned with confidentiality, do not seem applicable to personal data banks and are therefore not dealt with in this document.

b) The public order reform law

It will be easier to harmonize any future Italian law on confidentiality with the "Statuto dei Lavoratori" than with Law No 121 of 1 April 1981, entitled "New regulations governing the organization of public order". Articles 5 to 12 of this Law contain provisions on the same matters which are the subject of the four draft laws examined later. These articles comprise the mini-law on the safeguarding of personal data which was mentioned at the beginning of this paragraph (4). The safeguarding of confidentiality provided for under this law is, however, confined to personal data that have been processed in a computer centre belonging to the police force. Consequently, these provisions cannot be invoked when individual confidentiality is violated by data banks belonging to private firms or public bodies. As is evident, the scope of this Italian mini-law is quite restricted and, what is more, incomparable with that of foreign laws. Indeed, foreign laws always exclude from their scope data banks belonging to the police and those set up in the interests of national defence.

This law gives special powers on personal data banks to the Ministry of Internal Affairs, since it obliges public and private bodies to notify the Ministry of the existence of their data banks. In fact, a large number of Italian organizations and firms notified the Ministry as requested by the specified date of 31 December 1981. The reason why this information was collected is indicated in Article 8, where it is stated that by 31 December 1982 the government was to inform Parliament of the information thus gathered for the purposes of determining suitable legislation to safeguard the right of citizens to confidentiality.

The first census of personal data banks in Italy gave the results set out in Table 1. The overall total of approximately 100 000 personal data banks seems rather low. Each body or firm is compelled to notify individually every personal data bank set up within its own organization. A large firm may have more than 100 such banks. It is probable that, already in this census, the dichotomy between large firms on the one hand and small and medium-sized firms on the other could be seen: it is in fact probable that a

(4) These statutory provisions have to be incorporated into Presidential Decree No 378 of 3 May 1982, entitled "Regulations concerning the procedures for gathering, accessing, communicating, correcting, cancelling and incorporating data and information recorded in the computerized archives of the data-processing centre specified in Article 8 of Law No 121 of 1 April 1981".

significant percentage of data banks have not been reported, partly in the belief of those concerned that their specific case is not relevant to the purposes of the law.

This impression is borne out by the analysis provided by the Ministry of Internal Affairs, which distinguishes between data banks on the basis of the number of people on whom data are supposed to be kept: 93% of the reported banks are concerned with less than 10 000 people, while the remaining 7% are above this figure. Nearly all personal data banks in Italy (or, rather, those reported to the Ministry, which is not necessarily the same thing) are therefore used for the purposes of management, while only a small minority are large enough to be suitable for wide-scale social coverage. Hence, it would be a mistake to think that violations of privacy emanate only from large personal data banks, the nature of the data themselves and the use made of them are in fact the source of such violations. If the imbalance recorded in the ministerial report is borne out by subsequent surveys, however, it must give rise to the consideration that it would be appropriate to simplify the task of notification in some way so that firms do not become overburdened with pointless rules and the Ministry does not find itself swamped with information of little relevance.

Table 1

Personal data banks in Italy (1981)

Notifications by firms	55 556
Notifications by private individuals	2 900
Notifications by various bodies	1 180
Notifications by public authorities	728
Notifications by associations with no official status	685
Total number of notifications	61 049
Total number of personal data banks reported	approx. 105 000

Source: Ministry of Internal Affairs, 1983.

The notification procedure has to be repeated every year and the public order reform law does not envisage that Parliament may promulgate a law on privacy. The Ministry of Internal Affairs probably intends to continue supplying parliament with census results, from which it can derive pointers for future law-making. In such a case, however, the text drafted by the Mirabelli Committee (with which we will be dealing shortly) would have to be coordinated with the public order reform law. The Mirabelli bill provides for the

setting up of a type of "computer technology magistrature" under the auspices of the Presidency of the Council governing body. Personal data banks would have to be reported to this body by 31 December each year. However, according to the public order reform law, this information would have to be reported to the Ministry of Internal Affairs by the same date.

This duplication would be eliminated by the Marinazzoli bill, which provides for a single notification procedure to be made to the future surveillance authorities.

3. THE 1981 DRAFT LEGISLATION (THE ACCAME BILL)

The first complete bill on safeguarding confidentiality was presented to Parliament in 1981 by Mr. Falco Accame (5). The Socialist bill originated from an awareness that technological developments had already created an opportunity for individuals' privacy to be invaded, without there being any legislation governing this matter. The introductory report points out that this meant that Italy was in a position similar to people who had already agreed to store explosives while still studying the safety instructions on how they should be kept and used. It can be said that the situation has not changed between 1981 and the present day.

The Accame bill comprises about 20 articles and, consequently, offers some guidelines for safeguarding confidentiality rather than itself constituting an implementing instrument as such.

Article 2 contains the basic definitions, from which it can be seen that the law is to be applied to data on physical persons, when such data have been processed electronically (Article 2). Provision is, however, made for the extension of this protection to data on legal persons and to data processed manually, while the formulation of such provisions is delegated to another law.

The determination of any penalties and the definition of exceptions to the law are both deferred to subsequent laws. Given such deferments, and the general way in which the various principles which inspired the bill are set out, it seems clear that the greatest use that can be made of the Accame bill today is as a reference point for amending the other bills.

It must be pointed out that it is not clear which State body is to be responsible for ensuring that the regulations on confidentiality are complied with. From the context, it

(5) *Camera dei deputati*, No 2552. Bill tabled by Mr. Accame on 21 April 1981, entitled "Provisions to safeguard the right to privacy with regard to computerized data-processing systems".

would appear that that responsibility would fall mainly on the Ministry of Justice. The committee referred to in this bill does not have any real tasks to perform but simply acts as an adviser to the Ministry itself, with a view to applying the law more effectively.

4. THE 1982 DRAFT LEGISLATION (THE PICANO BILL)

At the beginning of 1982, a group of nine Christian Democrat Members of Parliament tabled another bill on confidentiality (6). This bill contains only 18 articles but is more comprehensive than the Accame bill. It makes clear provision for a State authority to carry out the necessary surveillance. The name of this committee is reminiscent of an analogous French institution - the National Commission for Data Processing and Civil Liberties. It is interesting to observe that this body will not be run by the Government but will be the direct responsibility of Parliament. The Committee will have a directory of data banks divided into two categories: those which need only to be registered and those which - on account of the greater sensitivity of the personal data on file - are subject to both registration and authorization.

The bill also provides for penalties against those who are responsible for data banks and who do not adhere to the Committee's rules. However, these penalties are so inconsistent as to completely nullify the law: "Failure to register on the part of an owner of data or an incorrect declaration of information for the purposes of registration will incur an administrative fine of between one and three million lire, to be increased by a third in the case of false declarations" (Article 7). It seems that the proponents of this law do not have a clear picture of the costs such a law on confidentiality imposes on firms. Faced with such light penalties, it may be more convenient for a firm to process data unlawfully and to pay a fine than to conform with the measures of protection laid down in the law.

The handling of disputes is entrusted to ordinary magistrates, "who would have to hear the opinion of the surveillance authority" (Article 13).

Finally, this bill gives a long list of physical and logical safety measures for the material protection of the personal data on file.

(6) *Camera dei deputati*, No 3195. Bill tabled by Messrs. Picano, Degan, Fiori Giovannino, Russo, Ferdiando, Abete, Rubino, Gruppo, Lo Bello and Merolli on 24 February 1982, entitled "Provisions protecting the right to confidentiality of physical persons with respect to the computerized processing of data and information".

5. THE 1983 TEXT (THE MIRABELLI BILL)

While the previous bills mentioned were introduced at the initiative of Members of Parliament, the draft legislation submitted in 1983 was to have been at the initiative of a ministry. In this case, we cannot really refer to it as a bill except incorrectly, as the early end of the legislature in May 1983 prevented it from being presented by Parliament. A committee of twenty lawyers drafted the text of 37 articles which, given its scope and systematic character, supersedes the previous ones and is the obvious reference point for any future discussion by Parliament on the topic (7). It would take too long here to fully analyse this draft law, so I will confine myself to evoking the most important points.

Firstly, in the context of international discussions on who the "person" is whose data must be protected, the Mirabelli Committee has adopted the widest definition: this law would protect data on both physical and legal persons. This solution has both advantages and drawbacks; while, on the one hand, it can be regarded as positive to extend to the maximum the protection conferred by any future law on privacy, on the other hand it is rather difficult to determine what data are "personal" when speaking about a legal person. For example, any large firm stores data on competitors, their products, their particular methods or organization, etc. In such cases what form would right of access on the part of a legal person whose data are stored by a competitor take? It is futile today to indulge in any abstract exercise predicting the future. Should the regulations be adopted, we will have to see what actually happens in practice. The fact that nothing might change can also not be excluded: any firm wanting to see the data on itself stored by another firm runs the risk of a reciprocal request and therefore of being obliged to open its own confidential archives for inspection.

As is fitting for a draft law which is government-instigated, the surveillance body would be directly responsible to the Presidency of the Council of Ministers (which will no doubt give rise to debates in Parliament). The staff of this surveillance body would be predominantly lawyers and its management would be the responsibility, primarily, of lawyers from the Court of Cassation. The practical implications of one aspect of this are not at all clear to me.

(7) "Schema di diesgno di legge concernente l'esercizio delle banche di dati personali ad elaborazione informatica" ("Outline of a draft law concerning the running of computerized personal data banks"), *La Giustizia*, II, 1982, No 12, pp. 25-29. Apart from this outline, this issue also contains an accompanying report and a commentary by Giuseppe Mirabelli.

The accompanying report stresses that the aim was to create an administrative body like all the others; as a result, the magistrates from the Courts of Cassation who are called on to manage the surveillance office will have to be granted leave of absence for the three years that their appointment lasts. The proceedings of the surveillance office can be appealed against through the administrative channels of the TAR (Regional Administrative Tribunals), while the penalties laid down under heading VI of the bill are to be imposed by the normal civil and penal magistrature. These magistrates know, however, that at the end of three years the magistrates of the surveillance office will be returning to the Court of Cassation, that is to say to the highest court of appeal of the ordinary magistrature. While not wishing to denounce this type of organizational solution, I think that all the non-legal and sociological implications should be discussed in depth.

Two vital points do not seem to have been very well developed and it is to be hoped that a debate in Parliament will clarify these matters.

In Article 12 (data custody), too little is said about the security measures which would have to be taken by a firm operating a personal data bank. This aspect of data protection is as fundamental as it is complicated; to solve all the problems involved with a single paragraph of only a few lines is possibly insufficient.

Article 3 defines the person responsible for the data, that is to say the person who must appear before the surveillance authorities to answer for any violations of the law. The committee was probably not fully aware of the implications of a regulation which would require a firm or a public office to identify such a person. As the penalties laid down in the draft law are heavy (imprisonment of between six months and three years in addition to a fine: Article 23, paragraph 1), it is easy to imagine that, before being appointed as the person responsible for the data, an individual would demand a series of guarantees in the form of a practical organization within the firm or public body for which he works. German law provides for a highly structured organization in this respect. In practice, something similar will have to be instituted in Italy, but the bill leaves it up to each organization to adopt the structure which it thinks most appropriate. In my opinion, this could lead to discrepancies in organization and to complications when the law needs to be applied.

Let us consider a case in which the organization of the firm is such that no employee feels capable of assuming the title of the one who is "responsible for the data". In the absence of such an individual, and as the regulations state that the person signing the notification form to be sent to the surveillance office is the one to be held responsible for the data, it could happen that notifications will no longer

be made, which is one of the crimes punished with the penalty stated above. In such a situation, to whom would the punishment be applied? To the manager of the EDP department? To his administrative representative? To the entire board of directors? As can be seen, there is scope here for endless disputes. What is more, disputes would be heard by ordinary magistrates whose proceedings, it must be recalled once again, are slow. In other words, there is a risk that the law will gradually cease to be applied.

6. THE FIRST BILL OF 1984 (THE SEPPIA BILL)

The work of the Mirabelli Committee had already been completed and a government-inspired bill was expected, when five Socialist members of Parliament presented a bill (surprisingly illiberal, to tell the truth) on "personal information systems" (8). This designation does not refer to "personnel information systems" (or, more briefly, PIS), which have been the subject of heated protest, particularly on the part of the trade unions in recent times (9). The sponsors of the bill meant any personal data banks by their designation, not just those referring to the staff employed by a firm or organization, which are nowadays regarded as one of the essential tools of general management in a firm or organization.

The introductory report mentions two specific reasons for concern which led those responsible to draft the bill: the regulations contained in the public order reform law (already described in paragraph 2, b), and discussion surrounding the 1981 Italian census: "Worries and suspicions were aroused even as the national census form was being drawn up, because it was felt that electronic computers would be able to investigate an individual by following his life from cradle to grave in a modern and anomalous form of Welfare State". It should be pointed out, however, that the difficulties surrounding this census were in no way comparable with those which accompanied the failure of the 1981 Dutch and the 1983 German censuses. On those occasions, the very idea of the census was rejected lock, stock and barrel and as a result could not take place. In Italy, the subject of the controversy was much more circumscribed. The 1981 census form had a box for indicating any vacant dwellings; staff employed temporarily to help citizens fill in the forms asked to be taken on permanently, took strike action and threatened to withhold the census forms and to

(8) *Camera dei Deputati*, No 1210. Bill tabled by Messrs. Seppia, Alberini, Colucci, Amodeo and Poti and presented on 27 January 1984. Rules governing the use of personal information systems.

(9) A brief account of a discussion held with Austrian, German and Scandinavian trade unionists can be found in my article entitled "Un difficile dibattito sui sistemi informativi del personale" (A difficult discussion regarding personnel information systems), "Data Manager", IX, July-August 1984, No 36, p. 26.

make public the data on vacant accommodation. The firm stand taken by the Public Prosecutor, who made it quite clear to these temporary staff what the consequences of their action could be, settled the dispute, but the general public was left feeling somewhat concerned.

While this bill originated, then, from specific preoccupations, it was also put forward as a statement of highly general principles, concentrated in merely 11 articles with a large number of internal subdivisions. The provisions viewed as a whole are probably more valid as a basis for a future Parliamentary debate than as a nucleus from which a true and proper law could be developed. I am therefore justified in confining myself here to indicating only its most characteristic element.

It is obvious from a first reading to what extent this bill owes its existence to the public order reform law. Competence in this whole field is attributed - by Article 8 - to the Ministry of Internal Affairs: "A specially designated office within the Ministry of Internal Affairs shall act as the surveillance body in charge of applying this law". This office (not the Minister for Internal Affairs, please note) "shall issue a detailed organizational and technical regulation for the application of this law", propose exceptions to the law to the Minister and draw up an annual report, which "the Minister for Internal Affairs shall convey to the House of Parliament". So the entire management of a law as tricky as this one is to be entrusted to the bureaucracy of the Ministry of Internal Affairs.

Such a choice, which, as far as I know, has no precedents in any other country, leads to serious difficulties when it comes to protecting the rights of a citizen who may, for example, have requested that an item of information referring to him be deleted or corrected. If there is disagreement between the "organization maintaining the information system" and "the subject of the data", Article 5 f (6) lays down a rule which is curious, to say the least: "Should a dispute not be resolved, the organization shall readily and clearly inform the subject of this right to make his request to the competent office of the Ministry of Internal Affairs". Here we are confronted with possibly the most serious procedural shortcoming in the whole bill. Indeed, Article 10 lays down fines of between 200 000 and 2 million lire or imprisonment from 6 months to 1 year for anybody who violates the provisions of the law on privacy, while Article 11 provides for compensation for moral and material damages, as well as reimbursement of legal costs. But how does one get from the competent office within the Ministry of Internal Affairs to the ordinary magistrates? Applying to the ministerial offices would appear to be quite superfluous, as a citizen would then have to appeal to an ordinary magistrate to uphold the rights which this bill would grant him.

These structural shortcomings - that is to say, giving a ministerial office the task of drawing up implementing regulations and the lack of any link between this office and the ordinary magistrature - make it pointless to examine the provisions on basic rights contained in this bill. From the bill it is not possible to find any indication on the way in which it would be possible to apply such provisions. At the political level, finally, it seems that any bill which would entrust the protection of the rights of individual freedom to the bureaucracy of the Ministry of Internal Affairs has very limited chances of being welcomed by the present Italian Parliament.

7. THE SECOND DRAFT LAW OF 1984 (THE MARTINAZZOLI BILL)

On 5 May, a bill on the confidentiality of data was officially presented by the Government to the Italian Parliament (10). The text of the bill drafted by the Mirabelli Committee, stymied by the early elections of 1983, thus continues to make progress slowly. The number of its articles has risen from 37 to 38, since there is now one providing for financial cover for any future law. The most significant variations, however, can be found in the regulations instituting its surveillance body.

The surveillance body must be vigilant and technically well prepared. Hence it comes as a great surprise to see the cumbersome mechanism set up by Article 5 of the draft law. While the person to be held responsible in a firm is dealt with in six lines, the surveillance body is described in detail, including a regulation in a quite lengthy legal text. Let us examine the kind of surveillance body wanted by the draft law under study.

Article 5 states that the director of such an organization should be a magistrate from the Court of Cassation, specifying, however, that he may even be a retired magistrate. This really is too much! Even if they have taken early retirement, appeal magistrates tend to be elderly people. How is such a choice compatible with the technical and managerial requirements demanded of such a technical and important surveillance body? The actual officials are to be ordinary and administrative magistrates and - an inclusion not stipulated in earlier bills - State barristers-at-law. In addition, non-managerial staff can also be enlisted (from the State administrations, including those with autonomous status, and from general government agencies, including those concerned with economic and financial affairs). It is

-
- (10) *Camera dei Deputati*, No 1657. Outline of a law presented by the Minister of Justice (Mr. Martinazzoli) in collaboration with the Minister for Budgets and Economic Planning (Mr. Longo) and the Minister for the Treasury (Mr. Gorla), on the setting up and operating of computerized personal data banks. Presented 5 May 1984.

to be hoped that this provision will be sufficient to provide the surveillance body with the necessary number of technical and secretarial staff so that it can function as smoothly as possible.

The entire staff of the surveillance office would be paid by the administrations to which they belong. This matter appears crucial enough to be repeated in two successive paragraphs with the same content. In fact, this provision makes it clear just how an office of this type could function with an annual appropriation of only 2 000 million lire, as stated in the article on financial cover.

Indeed, it must be borne in mind that a large number of staff would have to make on-the-spot investigations and would therefore receive appropriate expenses for their assignments. What is more, the particular requirements of computer centres would frequently entail working outside the normal working hours; as a result, overtime would have to be paid. The payment of expenses and overtime fees would be borne by the surveillance office, while the original administrations would pay staff salaries. If one considers the number of staff needed as indicated in the bill, this system may give rise to some doubts.

The organization would appear to be needlessly large. From Article 5 it can be calculated that the maximum number of staff is in fact 142 people, of whom 72 would have decision-making powers (see Table 2).

Even if this quota were not reached, how could one possibly organize and make a body of this size work? How would it be possible to reach quick and unanimous decisions? Perhaps it would be a good idea to look again at these figures in the light of the situation in similar offices which already exist in other countries. About 15 people work in the Office of the Federal Commissioner for Data Protection in Bonn, Germany, but German law has special features linked with the federal structure which make it hard to compare, on this point, with the Italian bill (even though it cannot be denied that the problem of coordinating control over the regions by central government does indeed exist in Italy). France - which is the closest example to us - has set up a National Commission for Data Processing and Civil Liberties whose collegiate body consists of 17 Commissioners assisted by 40 officials.

Table 2

Composition of the surveillance body (Martinazzoli bill)

Director	1
Deputy Director	1
State magistrates and barristers-at-law	20
Managers from state administrations	30
Non-managerial staff	70
Qualified experts, including outside consultants	20
Maximum number of staff	142

The system for financing this future surveillance body is even more questionable. As has already been pointed out, the provisions state that any administration from which officials have been drawn will continue to pay their original salaries. This involves placing an additional burden on other public bodies, which will hardly be enthusiastic about collaborating; if they allow a person to go on secondment, they have to make provision for a replacement and thus find themselves paying two salaries for the same post. Is it possible that the administrations from whom technical staff would be drawn (because I persist in the hope that such people would be asked to work for this office) would be prepared to allow such employees to leave when they have always been rare birds in public administration? An examination of what is envisaged by the draft law may confirm such doubts.

For the limitation of public experience, however, one must turn to the final article, which earmarks 2 000 million lire annually to run this surveillance office, drawing such money from the fund for supplementary payments of pensions to those living abroad. This figure is incompatible with the tasks and size of any future surveillance office. With such a limited budget, it is not possible to create offices for nearly 150 people, equip them with a large computer centre so that they can carry out effective controls and administer the mass of data which the law would oblige them to store. The risk is that the surveillance body, already rather unwieldy for its size, would work even more slowly because of its lack of technical facilities.

It would, however, be futile to go into further detail in analysing this government-instigated draft law. It is quite likely that it will be amended when debated by Parliament and if alternative or supplementary bills are tabled.

8. CONFIDENTIALITY OF STATISTICAL DATA IN THE MARTINAZZOLI BILL

The confidentiality of statistics in Italy is not governed by only one law; its profile merges - from frequently vague surroundings - from a series of laws issued between 1880 and the present day. In this complex mosaic it is sufficient to study the most recent addition, that is to say the regulation which even nowadays deals with the confidentiality of statistics. Article 19 of Italian Decree Law No 1285 of 27 May 1929 states: "Items of information gathered by means of surveys, and tabulated by the central statistical office, whether directly or by means of delegated bodies, are to be safeguarded by the most scrupulous secrecy and shall not be disclosed, for whatsoever reasons, except in a collective form and in a way in which no references to individuals can be made. Such information may be communicated to judicial authorities only when the request is authorized by a summons, order or injunction issued in connection with legal proceedings".

While, in the case of official secrets (Article 326 of the Penal Code) and professional secrets (Article 622 of the Penal Code) the qualification that the subject is aware of the data is essential, in the case of statistics the information collected is automatically secret. Riccardo Tomei, the author of a well-documented article on this subject, correctly observed: "With statistical confidentiality, a radical change is taking place in the form of a subjective approach, whereby anything learnt from a given subject automatically becomes secret, detaching itself from its object" (11).

To guarantee the completeness and reliability of statistical data, therefore, the basic principle is that the data collected by Istat are generally and strictly confidential. Even the tax authorities have to halt before Istat's doors (12). However, the expansion of the tasks performed by the Welfare State, the demands of an economy which is even more closely linked with international markets, and increased activities on the part of local authorities (the institution of ordinary statute regions in 1971 should not be forgotten) all require the increased circulation of data collected by the central office. As a result, the principle of confidentiality has been subjected to numerous exceptions. The law on the organization of the 1981 census laid down an explicit exception to this principle, obliging Istat to provide "data made anonymous on individual survey units to be used for processing of local interest" to those communes, provinces and regions asking for them (13). This exception was also rendered necessary because it is not clear whether the regions are able to carry out sectoral censuses. Exclusive responsibility for this rests with Istat and can be traced to Law No 120 of 18 January 1934 (for industry and agriculture) and Law No 766 of 4 July 1941 (for the population).

Nor is this the only exception to the principle of confidentiality of individual data. Each Italian commune keeps a register of its resident population; when a census

(11) Roberto Tomei, *Il segreto statistico*, "I tribunali amministrativi regionali", 1984, No 5-6, p. 184; anyone wanting further and pertinent bibliographical references on secrecy in the public administration in Italy is referred to this text.

(12) Article 51 of Presidential Decree No 633 of 26 October 1972 and Article 32 of Presidential Decree No 600 of 29 September 1973 expressly prohibit any investigation into or inspection of statistical data for the purpose of taxation.

(13) Article 2 of Law No 684 of 18 December 1980, under which the general population and housing census for 1981 and the agricultural census for 1982 were held.

is held, the Office of Registers uses census data to update its own registers (Law No 2228 of 24 December 1954). In a similar fashion, the Chambers of Commerce use data from industrial censuses to update their own registers of firms (Royal Decree No 29 of 4 January 1925). The reasoning behind these exceptions is clear: the data are not of a sensitive nature and their circulation promotes the proper functioning of the offices without harming those to whom the data refer.

While the Council of State does admit the possibility of transmitting statistical data to third parties, it still tends to refer to the original principle of strict confidentiality, "precisely because of the potential risks connected with technological developments", and does so with the aim of protecting "primary interests, which are also protected by the constitution, of private citizens" (14).

The debate on secrecy has been quite lively in recent years, not only because of the demands of statisticians and information agencies, but also because of arguments raised concerning the fight against terrorism and because of enquiries into irregularities on the part of the police forces. In 1981, there was even a conference on the problems of confidentiality in the Italian legal system (15). The political implications certainly did not facilitate discussion on statistical confidentiality. Nevertheless, the opinion that these restraints must be slackened to cope with today's economic and institutional demands must steadily gain ground from now on.

A current in the opposite direction to this trend among statisticians is the Martinazzoli bill, characterized in Section IV (Dissemination and communication of data to third parties) by greater strictness. Reading the first two paragraphs of Article 17, we can note the parallelism with the already quoted law from 1929: 1) The subject to whom data refer may prohibit their dissemination or communication to given categories of third parties. 2) Such prohibition is presumed when the data have been collected with the declared purpose of a census, an opinion poll, market research or similar". The explanatory report appended to the bill specifies that in the latter cases "in the person providing the data, the conviction has taken root (from consistent laws on censuses and on practices in other fields) that the data are going to be used and disseminated in a depersonalized way; our intention, therefore, is to safeguard this trust".

When safeguarding privacy, too, the rule which eliminates statistical secrecy applies only when a formal law exists. Indeed, the third paragraph of this same article states

(14) Opinion No 1379 of the Council of State, Section I, 19 February 1982, quoted in the article by Tomei.

(15) Cf. *Atti del convegno nazionale* (Rome, 26-28 October 1982), Padua 1983, which I have not been able to see.

that: "this prohibition has no effect when the dissemination of communication is permitted by law and within the limits set down by the law". This provision sets out to facilitate the communication of useful data to certain individuals - as can be read in the report preceding the bill - "for example when exercising the right to news".

Under the 1929 law, publication of data is permitted "in collective form, so that no reference to individuals can be made". Article 18 takes up again the spirit of this provision - backed up, as we have seen, by the Council of State - and states that: 1) Data whose dissemination is prohibited may be communicated when this is done in a way which makes it impossible to identify the subject and when the information is requested for the purpose of study, statistical research or similar". This regulation also seeks to avoid the formation of obstacles to sociological or historical research; in States which already have legislation on the confidentiality of personal data there have in fact been cases in which the authorities have refused to provide the data required, using such legislation on privacy as a pretext. In Italy, it will be necessary to see whether in practice the fact that communication is only "allowed" (but not "obligatory") will be sufficient to eliminate or at least to limit refusals on the part of lazy or apprehensive bureaucrats.

The possibility of officials being apprehensive is not so far-fetched. The Martinazzoli bill does not identify precisely (as seen at the beginning of the preceding paragraph) the person responsible for data, but it does state, in Article 19, when dealing with persons whose task it is to safeguard confidentiality, that: "The person responsible for the data banks shall be considered jointly liable with the defaulter for any compensation resulting from harm done to an interested party, unless he can prove that he did everything possible to prevent the data being communicated". This is therefore a civil penalty which is in addition to the several penal sanctions for the illicit communication of data: imprisonment of between 6 months and 2 years as laid down in Article 27.

A second exception to the prohibition of the dissemination of data is contained in the second paragraph of Article 18: "The communication of data whose dissemination is prohibited is nevertheless permitted when requested for purposes regarding the defence of the State or as evidence of crimes, provided the regulations governing this matter are observed". This definition seems to include infringements of tax laws, so that the confidentiality of personal data would seem to be more limited than that of statistical data. The "regulations governing this matter" must be those laid down

in the 1929 Law: an item of information can be passed on to the judicial authorities only if there is a formal warrant for it. For personal data obtained from the census, therefore, the same problem that arose from the 1929 Law when the 1948 Constitution of the Republic came into force crops up again: as a citizen is obliged by law to answer the questions in the census form, he can find himself in a situation where he is supplying declarations that can be used against him; this is, however, in sharp contrast with Article 24 of the Constitution, as well as with "the principle *nemo tenetur prodere contra se ipsum*, which, from the time of the English Revolution, has constituted a basis for all legal regulations in the civilized world and is thought to be in force in our own constitutional system" (16).

A Parliamentary debate on these provisions contained in the Martinazzoli bill could therefore constitute an opportunity for airing a whole series of problems connected with obsolete Italian legislation on statistical secrecy while we wait for the reform of the public statistical system to take shape.

9. TOWARDS SECOND-GENERATION LAWS

The delay with which Italian politicians are tackling the question of legislation on confidentiality is not necessarily a bad thing, even if within the EEC France has had a law on computers and civil liberties since 1978 and Germany since 1977.

Italy may be lagging behind these States with regard to legislation, but it is still in time to take account of their experiences. Safeguarding personal data is an unexplored and tricky area in which good principles will often risk being sacrificed for the demands of reality. This is particularly true when reality is a long and serious economic crisis such as our country is experiencing in 1984 without any end in sight. While the Italian Parliament is just receiving a draft law modelled on OECD proposals and which is in line with the other main European bodies of law, the latter are just about to be amended. Before the Italian bill is passed, it would be advisable to heed as much as possible innovations on the part of those who already have years of practical experience in the protection of personal data.

A change of course in the laws on the protection of personal data is certainly not due to a lessening of interest in citizens' rights. In the last few years, there has been a steady increase in the amount of concern for and interest in

(16) Sandulli and Baldassarre, *Profili costituzionali della statistica in Italia*, in "Diritto e società", 1973, p. 388.

problems linked with the spread of computer technology in daily life. Workers have learned to their cost how tightly the screws of control can be turned at the workplace if such control is entrusted to a computer. The individual citizen, as a user of public and private services, wonders with some concern what might become of the hundreds of "magnetic traces" which he leaves behind him when paying his taxes, enrolling in a union, subscribing to a newspaper, booking a flight, attending a school, participating in the public health scheme, and so on. Such fears were the source of the initial laws on safeguarding personal data. The passage of time has only confirmed them. Since the first discussions were held on these topics - around 1965 - the only thing that has changed is the technology of computers. Nowadays it is acknowledged that the laws from ten years ago were drafted to tackle a technological danger which turns out to be different from the one envisaged.

People at that time started from the premise that it would be possible to have processors at the centre of a network of terminals, via which a computer's own data-processing capacity and data would be made available over a territory which would become ever larger until the whole world was included. It was thought that the trend in computer technology for centring large computers in a few places and from there disseminating the information stored, would persist. This proved to be wrong: computers have got smaller and have become more independent, more scattered throughout the country and more difficult to coordinate and to supervise. This development in the technology, which differs from what had been expected even up to ten years ago, threatens the laws on data protection which are already in force.

Proposals have therefore been made to amend them to take account of new technological realities. Second-generation laws already conflict with those of the first generation, now superseded.

A central point in any law on the confidentiality of personal data is the provision requiring any public or private body to report to a specially set up surveillance body, describing the content and purposes of the personal data banks that it manages. It is immediately obvious that the inspiration behind this regulation is a concept of computer technology which is now outdated. A system of notification (and related controls) would have been simple if the management of computers had been concentrated in a few computer centres of ever increasing size. Instead, since

we are in a situation where personal computers are becoming ever more widespread, there is a risk that surveillance offices will be swamped with irrelevant declarations, which will prevent their staff from dealing with really essential monitoring tasks.

Of course, there are ways in which the flood of notifications could be stemmed. Very small firms and private individuals could be exempted; a simplified method of notification could be envisaged for certain categories of manufacturers, provided that the personal data which such people manage are not in fact too sensitive.

Nevertheless, the spread of computers is creating new opportunities for attacks on personal data by making it possible to gain illegal access to them. Anyone with a personal computer linked up to a telephone line can try to break down the defences of any data bank. Such an activity may be deliberately unlawful, such as in the case of procuring information to be sold, or - as has been experienced in the United States - it may be purely a prank or vandalism. Some individuals get a thrill out of sneaking entry into economic data banks and changing the account figures of individuals or firms, thereby provoking enormous rows between litigants who are acting in perfectly good faith. Some people enjoy wiping out data on file, just as they might enjoy smashing up a park bench or telephone box. Since this is a reality, the legal protection of data on citizens must be backed up by a whole series of precise security measures. Physical and logical security implies the incorporation of protective measures into buildings and machines on the one hand and into programs on the other. Such security measures are becoming increasingly indispensable, but they are hardly mentioned in Italian draft legislation.

One way of preventing the build-up of a mountain of notifications to the surveillance body would be by means of a software certificate. In other words, a company selling programs could ask the surveillance body to study its product and acknowledge that it conforms with the purposes of the law. Any firm using such a program would know that it was doing so in accordance with the law. For its part, the surveillance body would not have to ask for documentation from firms using programs which had already been approved.

This subject is not tackled in the bill, but it would perhaps be worthwhile bearing it in mind when an implementing regulation is prepared in order to facilitate the tasks of any future surveillance office.

The way in which such an office is to function is indeed the most delicate point of the entire draft law: if it is too slow in carrying out its work, the whole law will not only

be useless but could also be harmful if it causes holdups in the production of firms or impedes the work of public bodies involved in any disputes.

I should like to conclude my remarks by pointing out the dangers of a failure to implement any future Italian law on confidentiality. While the Martinazzoli bill provides for severe penalties, some points of the law (if the current formulation is kept) will be difficult to apply and the law itself will need an implementing regulation to be issued by the Presidency of the Council. These and other reasons could lead to the law on the protection of personal data not being enforced. Basically, Italy must promulgate this law if it wants to participate freely in the transnational flow of personal data. At the moment international reciprocity is based on purely formal considerations: if there is a law safeguarding confidentiality and if there is a surveillance office to take care of its enforcement, reciprocity cannot be denied. The practical problems of the time taken to settle disputes and the actual application of the rules are separate from such considerations.

As the risks raised by the use of computer technology are real, any future law on confidentiality will have to be more than just a screen behind which anything goes on.

*

*

*

DISCUSSION

Dr. DAMMANN (*Referatsleiter, Bundesbeauftragter für den Datenschutz, Bonn*) opened the discussion:

He remarked that the situation described in the paper of Professor Losano was typical of that in a number of European countries in which a general data protection law still had to be prepared. However, there was sufficient time to draft data protection laws for certain subjects. He had the impression that Italy was making a virtue out of necessity and was formulating data protection rules for specific areas which would serve as models for other countries.

It must be noted, and this applied not only to Italy, that among the motivating factors for data protection was the fear that national information processing industries might suffer in the field of international competition if no such legislation existed. The rumour would persist that there was

an exclusive club of European countries which, because of data protection arrangements, occupied a favoured position in this new and expanding market. In fact there were few countries which controlled cross-frontier data transfers and even here this was not in the form of legal restriction but simply as procedural rules supervised by data protection agencies. For the man in the street it would be regrettable if consideration of data protection were more influenced by economic factors than by questions of personal liberty in the computer era. The big difficulties and delays which data protection had encountered in some countries led one to believe that governments feared the subject almost as much as the public feared computers. The first data protection laws gave the impression that the subject had primarily a preventive function; active data protection should prevent stresses, frictions and injuries similar to those caused by damage to the environment arising in connection with files of information. Dr. Dammann remarked that Professor Losano in his presentation had proposed a second-generation data protection system and explained what he meant by this:

- (a) Data protection should not be based on the methods of yesterday involving large processing centres but rather on today's information techniques, wherever possible with an eye to their future development. Professor Losano suggested operating a modern, economic data-protection system with the help of software-certificates. The idea of data protection certificates could also be extended to hardware. These ideas were good in principle but because of particular factors, could lead only to limited gains.
 - Software, both job-related and system-related, needed to be updated frequently; the certificate must be updated likewise.
 - With regard to the capabilities of a large computer, it was very difficult to guarantee that undesirable operations were impossible on technical grounds.
 - Hardware and software were simply elements in a security system which also depended on organization, personnel and arrangement in the buildings themselves: total security was no better than the weakest link in the system.

- (b) Supervision played a vital role in the implementation of data protection. Dr. Dammann shared the concern that data protection might be condemned to failure by reason

of a bureaucratic structure in the arrangements for its supervision. In this context he referred to the changes in the technology for treating information. As automatic data processing became more and more widespread, there should be corresponding changes in the reaction of data protectionists. The supervision of data protection should not consist of the exhaustive noting, registering and sanctioning of every detail of processing procedures, but rather the selective examination of those sectors or applications which involved a heavy response burden, restriction of freedom or risk for the individual. Finally, those exercising supervisory functions should not only control and resolve difficulties but should, above all, give advice to data processing centres and thereby lead to a continuous improvement in practical data protection.

- (c) It was also true of the sanctions that what was needed was an optimum and not a maximum. Professor Losano rightly criticized the draconian penalties previewed by the Italian draft law not only for unauthorized disclosure but also the possession of data obtained improperly as well as errors arising from inadequate security measures. Had there been such laws in Germany, a large number of officials would have found themselves in prison after each data protection investigation. Dr. Dammann concluded with a question to the Italian representative: Article 17 of the Martinazzoli draft previewed statistical secrecy that should also cover scientific research and opinion surveys. Article 18, however, made an exception from the obligation of secrecy when a question of national defence or investigation of crime was involved. Were such wide-ranging exceptions really proposed? Did they apply to national statistics? Did this not involve a substantial reduction in force, which threatened confidence in the secrecy of statistical data?

Mr. COX (U.S. Bureau of the Census) made two comments from the standpoint of statistical methodology relevant not only to Mr. Cariani's paper but also more broadly:

- (1) It was found in the US Bureau of the Census that the statistical matching of data from different inquiries at different dates was very difficult in the absence of registration numbers and relying simply on names and addresses. Indeed, just comparing and matching address

lists to maintain the computerized geographic reference systems (which contained no information on persons) was an area of continuing serious research.

- (2) The US Bureau of the Census was the terminal point of a one-way-street in the decentralized US statistical system, so that all the data it released must be depersonalized. However, within the Bureau, and certainly within the US statistical system as a whole, the term "depersonalized" had no precise technical meaning. This was because measures of disclosure risk, whether *de facto* or absolute risk, were lacking. Studies were being pursued in this area. Regarding the discussion of the importance of public confidence and perception, it seemed that people in statistical offices must first convince others and be sure themselves that data on persons were secure and depersonalized. The way to do so was to develop standard, statistically defensible methods of measuring and reducing disclosure risk. Equally important, it should be possible to assess and measure the effects of such methods on data quality and completeness. Mr. Cox said that he would be most interested in mutual exchanges of information of a technical nature in the areas of data matching and disclosure avoidance.

Mr. FERREIRA DA CUNHA (*Instituto Nacional de Estatística, Lisbon*) made the point:

A bill had been submitted to the Portuguese Parliament by the Government on the protection of personal data. The delay in comparison with other European countries was due to the fact that Portugal had signed Convention 105 of the Council of Europe only on 14 May 1981 and that, in accordance with its Article 4 which required the adaptation of internal legislation, this had been done in 1982 at the time when the Constitution was being revised. The constitutional law, contained in its Article 35 provided that:

- (1) All citizens had the right to be informed of the content of computer registers concerning them and, having regard to the purpose of the information, might demand any necessary rectifications.
- (2) The access of third parties to these registers was forbidden as was linkage between registers, and cross-frontier transfers of data except in circumstances specified by law.

- (3) Computer processing could not be used for data concerning political, philosophical or religious views, nor for political or trade-union affiliations except in the case of anonymized statistical data.
- (4) The law defined the concept of personal data for the purposes of computer registers.
- (5) The attribution of a single national number to each citizen was forbidden.

The National Institute of Statistics, Lisbon, participated in the work of the inter-ministerial informatics commission which was asked by the Government to give its opinion on the draft law on data protection.

Mr. Ferreira da Cunha considered that statisticians, confronted with the new environment for the production of statistics, should make their voices heard on the subject of data protection legislation in the same way as for other proposals of a legal nature having a bearing on statistics.

Mr. QUATRESOOZ (*European Society for Opinion and Marketing Research, Amsterdam*) remarked as follows:

Having listened to the two days' discussion, it was clear that the market research organizations he represented shared many beliefs and concerns in common with official statisticians in facing the current situation. The parallelism with Mr. Cariani's paper was impressive and even striking.

Following the sequence of the paper, he referred to a few of the concerns expressed and looked at them from the somewhat specific point of view of the market researcher.

1. After describing the current EDP technological evolution, Mr. Cariani asked a question: would such an evolution reinforce the democratic process or would it work against it, weaken it?

As far as ESOMAR's activities were concerned, they believed that they served the democratic thinking and interests of the public at large. They indeed provided to the silent majority the opportunity to express itself freely and to know - in return - something about the real state of public opinion (independent of what the media and other pressure groups wished to propose as the general convictions).

For example, carrying out on-line opinion polls using Minitel and providing the information to everybody helped to protect democracy against demagogy. Their profession was operating as a bi-directional transmitter.

2. Attacks on private life were mentioned quite a few times, not only in the Italian paper but also during previous discussion. He felt they were of much less concern to ESOMAR because their activity was based on sample surveys. They were, in any case, primarily interested in the behaviour and the opinions of groups of people, not of individuals as such. As a result, their data could be made anonymous as soon as possible in the process of compilation; their code of ethics was coping with this point very precisely.
3. The right for the public to be informed was also covered in the Italian paper. That was precisely what they wished to protect and what they were ready to fight for. There must be general freedom to collect and exchange information, provided that such information was obtained according to strict professional principles and was objectively and responsibly handled. The right of bona fide research should be available to any person or organization which conformed to accepted and acceptable professional practice.

In some instances, legislation had tried to distinguish separately:

- Scientific ... equated with Academic and/or Official and
- Commercial ... i.e. profit-oriented and consequently not scientific!

ESOMAR could not agree with any system which provided some organization or bodies with the exclusive right - and the impressive power - of being the only institutions to know about, to report on (and to act on?) the state of public opinion. In any case, most of the research for public and academic bodies was in practice carried out on their behalf by private research companies and vice-versa, - academics took commercial clients.

4. As far as security/protection of the files was concerned ESOMAR believed that the correct solution to the data privacy problem lay in strong and secure protection of the address files as such. Data should be made anonymous as soon as possible. Once they had been separated from the addresses, the data should not then need to be subjected to the same rigorous security controls. The

legislation went too far when it claimed that otherwise anonymous data were still to be treated as identifiable as long as names and addresses exist anywhere in the survey organization!

5. The juridical protection of the confidentiality of personal data was definitely necessary but it should remain realistic and duly take into account other existing rights such as the freedom to conduct research and to report on it. The side-effects of data protection laws could, in the out-turn, create further dangers. It was important that the public should not get the impression that, because of such legislation, they should cease to provide any personal information whatsoever. ESOMAR experience had proven that the concerns and necessities of the profession could be duly taken into account in the development of appropriate and efficient data protection laws in countries where their representatives were involved in the preliminary discussion; the United Kingdom was a good example of that kind. In other cases, they had been threatened by legislation which was not originally directed against research but which introduced restrictive clauses asking for impossibilities, for example:

- a) The request to inform the respondent of the detailed aims and objectives of the research:

The resulting dangers of biased answers, lack of objectivity in the research, etc. from such a requirement were obvious. It would also run the risk of discouraging the public from freely volunteering information in the ways they had been quite happy to do hitherto. ESOMAR must insist that such a requirement - especially in its more extreme forms - would in practice make research less scientific. Indeed it would make some types of research virtually impossible, including certain research which the law courts and legislators themselves favoured (e.g. in cases of "passing off" substitute products, public awareness and understanding of various issues, potentially misleading advertising etc.)

- b) A requirement for the respondent's written consent to the interview:

The likely effects of overall refusal rates, differential response among different groups of the population, etc. were again obvious. The respondent's power to stop the interview at any stage should be quite sufficient protection. ESOMAR surveys had always

been of the voluntary type! They protected this right and needed it! But they feared that the applications of such strict rules would run the risk of discouraging the public from freely volunteering information in the ways they had been quite happy to do hitherto.

- c) Limitations on the transfer to the researcher of data coming:
 - from the body commissioning the research
 - from public authorities
 - from other sources such as direct-mail providing address lists and similar information.
- d) Limitations on the transfer of data cross borders, even within the same organization.
- e) Regulations governing the release of research findings or even to subject these to a certain type of "censorship".

ESOMAR could produce an even larger list of practical impossibilities. But these few examples were sufficient for illustrating the dangers their profession was currently facing in some countries ... and this quite often mainly because matters were being confused. In the current state of the art, the public was particularly apprehensive about the secondary uses governmental agencies could make of the personal information they provided.

Mr. Quatresooz remarked that the levels of response rate they achieved when inviting people to comment, voluntarily, on their reading habits or their cars were much higher than those he had heard mentioned in several instances for official Censuses. People became, for example, more reluctant when asked if their car was equipped with a radio - simply because that might have some kind of relationship with taxation concerns!

In summarizing the ESOMAR position, Mr. Quatresooz said that:

- they were open to dialogue on the subject and wished to be involved in all preliminary work on legislation in development,
- they were, in this context, delighted to have been invited to take part to this Seminar and expressed warm thanks to the EUROSTAT for this,

- they had appointed a highly specialized lawyer - Dr. Schweizer, located in Munich, Federal Republic of Germany; with this help, ESOMAR acted as a clearing house and as a consultant for national research bodies; any kind of exchange of information with organizations participating in the Seminar would in the context be welcome,
- they would continue to press for the general acceptance that self-regulation procedures were adequate safeguards.

The author replied to the discussion:

He agreed that some countries had been hesitant in introducing data protection laws in recent years. While the main motive for such laws was the safeguarding of the confidentiality of information on individuals, the importance of facilitating transfrontier data flows should not be underestimated. He noted that the most influential groups with regard to legislation were those concerned with economic interests while there was no major group representing individuals. He remarked that, as things stood, bureaucratic problems sometimes represented insurmountable obstacles for data flows. He quoted the case of data kept on a computer file in Vienna which had been sent for processing to an agency in Munich. This arrangement had to be discontinued under new data protection laws.

Professor Losano thought the gradual introduction of data protection through regulations relating to particular sectors would be beneficial, if these could ultimately be combined into a general framework of data protection legislation. However, he feared that such a development could not come about in this way in Italy as the legislation concerning data protection for the different sectors was the concern of different ministries, often with conflicting interests; there was no central authority to prepare the overall data protection legislation.

With regard to the severe penalties mentioned in the draft legislation described in the paper, there was normally little possibility of these being applied, even if the legislation were enacted. As there were no special courts in Italy, all such laws would have to be enforced through the normal courts. The delays involved were likely to be so great as to defeat the purpose of the penalties.

THE CONFIDENTIALITY OF STATISTICAL DATA

E. STILIARIS

with the assistance of other NSSG officials

I. INTRODUCTION

1. THE CONFIDENTIALITY OF STATISTICAL DATA IN THE EEC

Statisticians have always maintained that data confidentiality is a basic precondition for ensuring the accuracy and, in general, the success of statistical programmes. The guarantee of confidentiality creates an environment within which it is possible to ensure a high-level obligation to supply information, the success of a statistical programme largely depends on the voluntary cooperation of the respondents. One element which encourages this cooperation is the assurance given to the respondent that he will not be running any risk by replying correctly to the interviewer's questions in the course of a survey. The most usual and suitable way of providing this assurance is for the statistical service to ensure that the replies are kept confidential.

In the Community Member States the confidentiality of statistical data is protected by statutes which lay down special rules on statistical secrecy, i.e. the conditions under which data gathered for statistical ends may be used. Statisticians in all the Member States are obliged to observe certain provisions concerning secrecy in respect of the information which they obtain in the exercise of their profession. This secrecy principle is encountered in many professions and, particularly in the case of civil servants, is often known as professional secrecy. Professional secrecy relates only to the individual who has an adequate amount of information. On the other hand, statistical secrecy reinforces professional secrecy and, we might say, transcends it. This is because statistical secrecy must be observed not only by the individual but by the statistical service itself.

The legal provisions on statistical secrecy differ from one Community country to another. Although there are two extreme cases, most of these provisions refer briefly to a general principle of confidentiality, in accordance with which it is forbidden to divulge personal data.

The first extreme case is the Belgian Law of 4 July 1962, which not only sets out in full detail the meaning of statistical secrecy but also enumerates all the details of how the law is to be implemented. In accordance with this law, which is the only statistical law in the Community to define a statistical secret, "a statistical secret is any information which we can obtain only with the assent of the interested party in a manner permitted by law". The other diametrically opposed case is the Danish law, which remains silent on the subject. Most of the laws in the Community Member States which refer to confidentiality of data provide for the imposition of sanctions on employees who commit infringements. Finally, in cases in which personal data may be communicated to other Government services, it is clearly stated that the tax authorities are excluded.

2. THE CONFIDENTIALITY OF STATISTICAL DATA IN THE NATIONAL STATISTICAL SERVICE OF GREECE (NSSG) - MEASURES TO ENSURE CONFIDENTIALITY

As we have already said, observance of the principle of confidentiality of statistical data contributes substantially towards the creation of a climate of confidence between the Statistical Service and those who provide the data, with the result that the data collected are more correct and more reliable. The management of the National Statistical Service of Greece also realize the importance of this principle and, shortly after the NSSG was reorganized, ensured that it was enshrined in Decree-Law No 3627 of 1955.

In accordance with Article 36(1) of this Decree-Law the questionnaires and tables must state whether the information solicited may be used for fiscal purposes or in connection with military conscription. If this is not the case, the contents of the questionnaires and tables are considered confidential, and in no case may they be used by a governmental authority with a view to checking income tax or any other levy, or with a view to imposing taxation, or in connection with the regulation of any dispute between private individuals and the State or any dispute involving the courts". Article 36(2) states that "a certificate, for whatever purpose, may be granted by the Statistical Service on the basis of the replies contained in the questionnaire, only to the respondent himself, be he a legal or natural person".

Staff recruited by the National Statistical Service of Greece for the collection of data are strictly bound to statistical secrecy. Article 40 of Decree-Law No 3627 states that "any official of the Statistical Service responsible for the collection and processing of statistical information, who reveals confidential personal data to third parties or to another government service, shall be liable to

a fine of up to 10 000 metallic drachmas or to imprisonment of up to one year or to both of these penalties, and to the disciplinary penalties set out in Law No 1811/1951 on civil servants".

In order to assist the NSSG in its mission and to facilitate the work of its interviewers, the law contains special provisions which stipulate that the various respondents must provide any data required by the NSSG in the context of compulsory surveys. In particular, Article 37 of Decree Law No 3627/156 makes it incumbent on all civilian authorities at national and local government level, and on all military and church authorities, on all State organizations and their staff, to provide any information required by the NSSG in the execution of its work. Article 38 of this Law places the same duty on all federations and guilds, cooperatives, firms and organizations and on all private individuals. In both cases severe sanctions may be imposed on persons who refuse to cooperate with the NSSG. Finally, Article 39 of Decree-law No 3627 provides for the imposition of sanctions on anyone who slanders the work of the NSSG or hinders the work of those responsible for collecting statistical data.

Apart from the statutory measures for ensuring the observance of the confidentiality principle, the National Statistical Service of Greece also takes preventive measures to ensure statistical secrecy. It takes steps to convince the public that personal data will be treated as confidential. Thus the word "confidential" is printed on the questionnaires. Likewise in correspondence with respondents who are obliged to provide statistical information and in statements in the press and other mass media, it is made clear that personal data will not be used in such a way as to provide any opportunity for investigating the economic situation of the respondents and on a more general level, that their rights, privileges and benefits will not be interfered with. In addition the enumerators and interviewers are given clear instructions to point out to respondents that their answers will be treated confidentially.

Other non-statutory measures are also adopted by the NSSG in line with the nature and specific features of the survey in question. Specifically, in the Annual Industry Survey (AIS), data may be published on branches of economic activity in which there are fewer than three undertakings (firms). This is because when there is only one undertaking in a branch the data are personal and cannot be published. Similarly, when there are two undertakings in a branch, each of the two competitors can obtain data on the other by subtracting its own figures from the total. In the household budget survey and the consumer behaviour survey, the name and address of

the respondents are not written on the questionnaires and they are simply given a code number. The code number-address cross-reference key is kept by a responsible official of suitable character so as to minimize the likelihood of decoding and leakage of personal data. The general purpose of these measures is to banish any reluctance on the part of the respondents to provide personal information, once they understand that these data are recorded in anonymous documents. It is made clear to the respondents that their personal data will be published, after suitable processing, in anonymous series in tabular form. It is also pointed out that the purpose of the cross-reference key is to monitor the work of the officials/interviewers i.e. to ascertain that all the sample units were investigated as planned.

Finally it should be pointed out that only staff who are directly concerned have access to the personal questionnaires used in the various surveys and censuses and that they are instructed to take care to avoid any leakage of personal data.

Following our examination of Greek statutory measures to ensure the confidentiality of personal data and the measures taken by the NSSG with the same objective in mind, we may conclude that the secrecy of personal data is fully assured in Greece, in respect of information relating to both the private and economic sphere. In addition, the fact that up to now the Greek courts have never had to try a case involving the infringement of statistical secrecy further substantiates the above conclusion.

II. THE GREEK EXPERIENCE

1. WHAT IS GENERALLY CONSIDERED AS CONFIDENTIAL?

Different countries consider as confidential personal data relating to the private and economic spheres. Certain countries attach greater importance to personal data of an economic nature because respondents demand greater guarantees when providing such data. These guarantees take the form of an assurance that the principle of confidentiality will be observed, the purpose of which should be to allow the statistician to access the personal data of a respondent without prejudicing the respondent's rights (this applies both to individuals and firms). The statistician is interested in personal replies because his aim is to evaluate the reliability and usefulness of the statistical aggregates which are derived from the personal data. Correct personal replies lead to correct statistical aggregates, creating the basis for a proper analysis which will allow correct conclusions to be drawn.

In general, the NSSG treats all personal information as confidential. Almost all the NSSG's questionnaires in which personal data are collected bear the mention "confidential". However, there are cases in which only particular questions in a questionnaire are characterized as confidential. Specifically, in the questionnaire used in the census of industrial undertakings, the only data considered as confidential are those which relate to employment and the nature and magnitude of the energy sources used by the firm (undertaking).

2. REACTION OF THE RESPONDENTS (HOUSEHOLDS, FIRMS, ETC.) TO THE OBLIGATION TO PROVIDE DATA, WITH PARTICULAR REFERENCE TO THE GENERAL POPULATION CENSUS

The reaction of the various respondents (individuals, households, undertakings, etc.) to the obligation to provide statistical data is based on a general fear that the data which they provide for a particular end may be used in such a way as to affect their private, family, social or economic life. Primarily, this reaction is considered to be a function of the respondent's income and financial standing in general, and of his character, level of education, and the degree to which he is informed about the mission and work of his country's statistical service.

There are certain singularities in the way different categories of respondents react in Greece. The respondent's income and his financial standing in general appear to affect results differently depending on the unit under investigation (household or firm). In the case of households (household budget survey, consumer behaviour survey, workforce survey, etc.), opposition is more marked in the households of the well-to-do, who are reluctant to cooperate with the Statistical Service's staff. When they do agree to cooperate they tend to conceal data. These reactions are due to the households' fears and suspicions that - despite assurances to the contrary - the data they provide for statistical purposes may come to the knowledge of the tax authorities as well. On the other hand, less wealthy households show no marked reluctance to provide data and the Statistical Service has easier access to personal data from those households. This is because many of these households are not obliged to submit income tax returns or are not able to conceal data from the revenue authorities (wage and salary earners).

When the unit under investigation is the firm or undertaking (survey of retail sales, annual industrial survey, survey of products, labour cost survey, etc.), reluctance to provide

data is greater in small firms than in large firms in the same economic branch. This reluctance mainly concerns data on the firm's turnover, but also on employment, the level and nature of the energy sources used, etc. Most firms or undertakings which owe money to the tax authorities, the insurance agencies and the energy supply services belong to the category of small firms. Thus they are particularly worried lest the data they provide should be leaked to the above authorities. On the other hand large firms and undertakings generally have a competent employee whom the interviewer can approach to ensure that the questionnaire is filled in. Thus he can work fast and efficiently. However, it should be noted that the comparative readiness of large firms to cooperate with the NSSG does not necessarily mean that they are completely frank in their answers. Here too the businessman may fear that the data he provides for statistical purposes will be revealed to the tax authorities, and so he may give incomplete data or at least those he gave to the tax authorities.

In conducting surveys in the primary sector of the economy (livestock surveys, fruit tree surveys, structural surveys in agriculture and livestock keeping, agricultural and livestock keeping census, etc.), the NSSG does not encounter any marked reluctance to provide data and, with a few exceptions, it has a relatively good working relationship with the various respondents. The main reason for this is that respondents in the primary sector - with very few exceptions - are not obliged to submit income tax returns, and so they have no reason to fear that the principle of confidentiality will not be observed. As in the case of certain other respondents, reluctance to provide data is mainly due to lack of information on the usefulness and importance of collection statistics, and on the work, mission and duties of the Statistical Service in the collection of these data. Respondents who, after being suitably informed, grasp the purpose of the survey and who have a positive view of the planned use of the data and believe that they will not be damaged by providing the information required, will not be uncooperative during the survey and will most likely answer honestly. This makes for the most accurate possible statistics and minimizes the work of the NSSG in collecting them.

The General Population Census is the largest survey conducted by the NSSG. It is the oldest survey in Greece and has always been carried out successfully. Great importance is attached to it, both by the statistical authorities and the government, and it is the best-known NSSG survey in Greece. Census day is the day of the Statistical Service in Greece.

Thus it may be superfluous to point out that the NSSG has very rarely encountered refusals to provide data during the General Population Census. Even in the few cases in which people have refused to cooperate, fear that the principle of confidentiality of personal data would not be observed has never been mentioned as a reason.

3. RESTRICTION ON INFORMATION AS A RESULT OF CONFIDENTIALITY

Observance of the principle of confidentiality means that statisticians are faced with two conflicting choices. On the one hand they must respect the law on statistical secrecy, so as to create a climate of confidence between the respondents and the Statistical Service. On the other hand they are aware that the final objective of their work is to provide as much information as possible to the users of the statistical data, i.e. to minimize the amount of statistical data which may not be revealed for reasons of confidentiality. The greatest drawback of this principle is the impossibility of publishing personal data, i.e. the restriction of statistical information.

As mentioned above the NSSG does not allow the disclosure of personal data because it respects and observes their confidential nature. The principle of statistical secrecy has great importance for a small country such as Greece. In particular in the mining and processing industries, because of the small number of firms involved, the NSSG is often unable to publish data of vital importance to a country which is in the process of industrial development. This is because, as already mentioned, the NSSG does not publish data on firms if there are fewer than three units in an individual economic branch. This problem becomes more acute when the NSSG wishes to publish such data not at whole-country but at regional level, such as the level of the geographical area of the economic development region. In these cases, i.e. when there are one or two undertakings in a branch, the branch is merged with the closest related branch and the amalgamated data are then published. In some cases it is necessary to merge more than two branches in order to publish the relevant data. However it should be pointed out that there is no way in which data on firms in the above categories can be published at *nomos* level. Finally, let us note that the fact that there are only one or two undertakings in a branch does not absolve these firms from the obligation to provide data to the NSSG.

It should also be mentioned that apart from the restriction on statistical information and the impossibility of publishing the maximum amount of decentralized statistical

data, observance of the confidentiality principle leads to time lags in publishing the data, loss of valuable time for the NSSG, and financial burdens for the State in general. This is because the procedure of identifying the confidential data and of preparing the various groupings required in order to present the data in a manner suitable for publication is expensive both in terms of time and money.

4. PRECONDITIONS FOR THE DISCLOSURE OF CONFIDENTIAL DATA

The principle of confidentiality should protect the respondents (private individuals, households, undertakings, etc.) from any consequences which might result from the disclosure of data which concern them. In this respect the NSSG must act with great circumspection and avoid any presentation or analysis of data which is not acceptable to the parties concerned.

The NSSG can achieve this objective by assuring the respondents that no use will be made of their personal data by third parties without their personal consent.

The NSSG can adopt this course in a number of cases. Specifically, in the products survey in the secondary sector, it is permissible to publish data on undertakings, even when there are fewer than three in a branch, but only after the respondents have given their personal consent in the form of a signed letter. Otherwise publication is not permitted. Thus, when we encounter the characteristic three dots in a NSSG publication, this means that data (on products) are available but that the respondent has refused to allow them to be published.

It should also be pointed out that pursuant to Article 36(2) of Decree-Law No 3627/2956, the NSSG provides a certificate based on the respondent's personal questionnaire to the respondent only, who is free to use it as he chooses.

Finally, it should be noted that, pursuant to Article 44 of Decree-Law No 3627/1956, confidential personal data to whose publication the respondent has not consented may be provided by the NSSG only to the Directorate of National Accounts at the Ministry of the National Economy (see below).

5. OBSERVANCE OF CONFIDENTIALITY VIS-A-VIS OTHER STATE AGENCIES (EXCEPT FOR TAX AUTHORITIES) - PROBLEMS ENCOUNTERED

Observance of the confidentiality principle presupposes a certain functional independence of the Statistical Service from the administrative, judicial and police authorities, as

well as from the Government itself. The Statistical Service must not only be under an obligation but also in a position to refuse to communicate personal data to these agencies. For a government to use personal data gathered in the course of statistical surveys for administrative ends, e.g. in combating tax evasion, infringements of price controls, etc., would be a short-sighted policy which would ultimately be the end of the statistics.

Despite this fact in a number of countries the law provides for the waiving of the principle of statistical secrecy in particular cases. Generally, such exceptions should be rare and justified. They include cases in which there is a need to provide confidential data to a government service and, while it is not planned to use these data at the expense of the respondents, the general confidentiality rule forbids disclosure.

Article 44(7) of Decree-Law No 3627/1956 entitles the NSSG to provide the independent Directorate of National Accounts at the Ministry of the National Economy with any statistical information which it needs. The Decree-Law states that "with a view to fulfilling its objectives the Directorate of National Accounts may use information and personal and general statistical data collected directly by the National Statistical Service of Greece or by its various services".

6. THE NSSG AND DATA KEPT BY OTHER AGENCIES

Statistical Services are free to use personal data which have been published by other agencies, provided they identify the source, in their publications and studies. In Malte, for example, foreign trade figures prepared by the customs authorities are published by the national statistical service under the responsibility of the customs authorities.

Cooperation between the various agencies which possess statistical data should not be restricted to publication of these data, as in the above case.

Rather, this cooperation should be more far-reaching, extending to as many stages as possible, including the collection, processing, presentation and analysis of the statistical data. This is the way to improve statistical information both quantitatively and qualitatively.

Up to now it has not been possible to bring about such cooperation between the NSSG and the other bodies which keep statistical data in Greece. Shortcomings in organization, programming and coordination along with other practical

problems, such as the use of different coding systems, rule out an easy solution. However all the parties involved are endeavouring to improve the situation.

7. CONCLUSIONS FROM THE GREEK EXPERIENCE

The confidentiality of individual data relating to private and family life and of personal data of a purely economic nature is fully assured in Greece.

Cases of refusal to furnish data, be it on the part of households or firms, are relatively rare. In the case of questions of a personal nature the reactions of respondents is conditioned by the degree to which they consider the questions indiscreet (e.g. in the fertility survey, respondents frequently refuse to answer the questions put to them, because of their particular nature). In the case of questions of an economic nature the responses depend on the extent to which the respondents are convinced that the Statistical Service will ensure that their answers remain completely confidential.

Adequate and timely information of the public on the importance of collecting statistical data and on the NSSG's mission is of decisive importance in solving the problems which arise from public suspicion that personal answers will not be kept confidential.

The principle of confidentiality of personal data restricts statistical information when the NSSG wishes to publish statistical data at a level of small geographical sub-units and branches of economic activity, because in these cases the data are personal.

III. THE NEW SITUATION

1. THE NEW TECHNOLOGY AND DATA CONFIDENTIALITY

Technology advances have led to marked improvements in the provision of statistical information. Automatic data processing has greatly reduced the time required in processing of the statistics as well as the costs of processing, print-out and publication of the findings. However, the impact of technology is not confined to savings in terms of time and money alone. Both the accuracy of the statistical findings and the assurance that the respondents' personal data will be kept secret owe much to technological developments. The replacement of old personal data files by magnetic data carriers has contributed decisively to minimizing the likelihood of disclosure. This is because, apart from coded personal data, the magnetic carriers do not

contain data identifying the units interviewed (name, surname, address, etc.). In addition, during the various stages of computer processing of the respondents' personal data, only a very small number of specialized personnel have access to these data, and these personnel change from one stage to another.

In the context of the new technology, the statistical services of various countries proceeded to create special data base management systems which have been geared to meet statistical requirements. These systems were first developed in the United States and Canada, where they have been operating very successfully. In Europe too, programmes for the organization and functioning of data base management systems are now being systematically developed, and data banks of this kind are already operating in most Community Member States.

There are many convincing arguments for creating information systems based on operating statistical data banks. They include the need to make the best possible use of data derived from a wide range of statistics, the exploitation of chronological series of statistical data which have been stored in a large variety of magnetic carriers and the need for rapid production of statistics both at domestic and international level. In a computer centre with a large range of files it is very difficult to provide timely and accurate information to all users of statistics without a statistical data bank. One great advantage of such a data bank is that it provides greater assurance that respondent's personal data will not be revealed, i.e. that confidentiality will be observed, by the various users having different levels of access depending on the confidentiality level of the final data.

2. THE GROWING CONCERN OF RESPONDENTS ON THE SUBJECT OF CONFIDENTIALITY

In recent years there has been growing concern about the possibility of linking computer-based statistical data systems in such a way as to create files containing personal information on private individuals and firms - information which, accordingly, is liable to be disclosed. There are fears that by identifying private transactions the secrets of the firms or individuals may be disclosed, along with their patterns of behaviour, activities which by definition are of a private nature. An equally disturbing fact, in particular for business people, is the fear that the data contained in these personal files may be used in such a way as to prevent them from obtaining credit, from employing personnel or from achieving various other objectives.

Similarly, many firms, while they do not doubt the goodwill of the Statistical Service, do not trust its personnel and fear that the statistics may be used for industrial espionage, something which is difficult to prove. The computer is seen as a device for comparing and analysing data on various aspects of the individual's economic and social relationships and for using this information to restrict the freedom - in the wider sense of the term - of the individual or firm.

The above comments are indicative of the widespread public fear that the data people provide for statistical purposes may have adverse effects on themselves and they reflect respondents' concern for confidentiality. Accordingly, in developing statistical systems the following basic principles should be observed:

- 1) when personal data for the statistical system are requested from private individuals or firms, the respondents should be informed, as to whether they are legally obliged to provide these data or not and as to the possible consequences for themselves which may arise from providing the information required or refusing to do so.
- 2) The statistical service must assure the private individual or firm that no use will be made of their personal data without their personal consent.
- 3) The statistical service must also state that personal data on private individuals and firms will not be communicated under the terms of any compulsory legal procedure, except when the interested parties have the opportunity to inspect these data prior to their being divulged.

3. THE GROWING NEED FOR STATISTICAL DATA ON SMALL AREAS AND THE CONFIDENTIALITY OF DATA - ASSOCIATED PROBLEMS

The drift of the population to urban centres, zoning trends and the cultural, social and general economic development of many countries have created particular problems. Thus the major strategic objective of these countries is to ensure decentralized regional development in all the abovementioned fields. However, the realization of this goal demands proper regional planning, which presupposes the creation of decentralized regions for development purposes. Accordingly there is a large and urgent demand for data in this field, particularly in countries which are now endeavouring to upgrade the regions.

However, in the preceding section we saw that the principle of confidentiality of personal data places restrictions on the provision of decentralized data by the statistical authorities. These restrictions will depend on the size of the country, the specific nature of the statistics to be published (primary or secondary sector, etc.), and the level of geographical subdivision and breakdown on the subject matter. The Luxembourg tobacco industry consists of only one firm. Consequently it is impossible to publish the data on this industry as personal data, even at the whole-country level. In Greece, as we have seen, problems of this kind become particularly crucial when the NSSG wishes to publish data on processing or mining firms at the level of the regional development area, agricultural departments, *nomos* and branch of economic activity.

However, similar cases are encountered in many other countries to a greater or lesser extent. Accordingly, various methods and measures have been adopted to cope with these problems. These measures, particularly in industry, include the principle of publishing data on particular firms only with their consent, whenever the number of units in each economic branch is less than a particular limit laid down by each country (e.g. four in Luxembourg, three in Greece, etc.). When consent is not forthcoming or when there is no provision for such a measure, the statistical service proceeds to group two or more related branches and to publish aggregated data on these branches. Finally, a number of countries classify as confidential only certain questions in a questionnaire, with a view to minimizing the number of cases in which it is impossible to publish personal data because of their confidentiality.

IV. TACKLING THE PROBLEM

1. USE OF TECHNOLOGY WITH A VIEW TO AVOIDING THE DISCLOSURE OF CONFIDENTIAL DATA DURING THE AUTOMATIC PROCESSING, STORAGE AND RETRIEVAL OF DATA

As pointed out in the previous chapter, the need to create information systems based on statistical data banks, taking advantage of the opportunities offered by modern technologies, is of vital importance for improving statistical information. Of equally decisive importance is the fact that the statistical data bank provides improved conditions for ensuring the safe keeping of the statistics in question. This feature, which mainly consists of avoiding the disclosure of confidential personal data, is provided at all stages in the production of the data, such as automatic processing, storage and retrieval of the data.

Because of these advantages most Community countries proceeded to set up statistical data banks. The experiences made in these countries showed that at least four years of systematic preparation are required in order to establish a functioning statistical data bank.

The NSSG, bearing in mind the advantages of such a system and the experience of the Member States in which such banks have already been installed, proceeded to design such a system and to take the measures required with a view to establishing a statistical data bank in the NSSG's Computer Centre. Accordingly, in 1984 preparatory work was begun with a view to establishing a decentralized information system.

The following steps have been taken:

- i) A feasibility study was commissioned from a specialized software firm with a view to drawing up the technical specifications for all the computer equipment to be supplied.
- ii) It is planned to purchase six microcomputers in the near future. They will be used initially for training NSSG staff, but also for research and processing purposes and as network terminals.
- iii) The existing EDP system (NCR V8565M) is being expanded from batch mode to interactive mode, in order to train and prepare the staff to work on the new system to be installed, which will support both data bases and ensure increased confidentiality for the user.
- iv) A scientific expert will shortly be recruited to work on the data bases and network.

The establishment of an operational statistical data bank in the NSSG will bring Greece into line with the other EEC countries in the field of statistical information.

2. INTRODUCTION OF THE PRINCIPLE OF DEFINING CERTAIN QUESTIONS AS CONFIDENTIAL WITHIN A QUESTIONNAIRE

The Community Member States do not adopt a single approach as regards the definition of confidential statistical data. Certain countries are radical in this respect and define as confidential all personal data without exception. Some countries place more emphasis on personal data of an economic nature, whereas in others the law on statistics makes no reference at all to data confidentiality. Generally speaking the situation as regards the definition of confidential data differs from one country to another.

However, as far as firms are concerned the statistical services of the Member States seem to agree that the names and addresses of the firms should not be covered by statistical secrecy. This is because the disclosure of this kind of information is not such as to have adverse effects on the interested parties. The information that firm X is located at address Y not only does not damage the firm in question but advertises its existence to other firms and to consumers.

The principle of defining certain questions within a questionnaire as confidential has been adopted in Germany, in which the name and address of the firm are not considered confidential data, while of the nine questions in the questionnaire used in the census of processing industries, only two have been defined as confidential (employment and the volume of energy used).

By stipulating that only certain questions within a questionnaire are confidential, it is possible to minimize the amount of personal data which cannot be published for reasons of confidentiality. Specifically, such data are confined to those questions which the respondents are most reluctant to answer and for which they demand more assurances - assurances which are provided by ensuring that the answers to these questions are kept confidential.

Accordingly, the best way of tackling the problem of not being able to publish confidential personal data is to expand the application of the above principle to households and to other categories of respondents, while restricting statistical secrecy to the smallest possible number of questions within a questionnaire.

*

* *

DISCUSSION

Mr. EMBLETON (*Central Statistics Office, Dublin*) opened the *discussion*:

The paper presented an interesting insight to the laws and practices relating to the protection of confidentiality in Greece. Moreover a number of relevant questions had been

posed in the context of the newer processing techniques which had become available to statisticians as a result of advancement in computer technologies.

Greek law. Turning first to the position in Greece, it was clear that as for all national statistical services, the NSSG attached considerable importance to the statutory provisions and non-statutory practices relating to the protection of confidentiality of individual statistical returns. Indeed, he was struck by the similarity between the Greek and Irish positions in a number of respects. One particular emphasis seemed to emerge in the Greek arrangements and that was that confidentiality appeared to be more strongly orientated towards protection of information of an economic nature than of personal data; this, at any rate was his reading of Article 36(1) of the 1955 Decree-Law, as referred to in Section 2.

Fiscal use of data. He noted too that Article 36(1) allowed for the possibility of making data available for fiscal purposes. This was not further elaborated on in the paper. Also in Section 5 an exception was made in the case of the tax authorities in respect of the observance of confidentiality *vis-à-vis* other State agencies - here too no further comment was given. It would be of interest to learn to what extent, if any, information was collected in statistical inquiries which might be used for fiscal purposes; what types of surveys were in question and what effect had this declared use had on response rates.

Protection of NSSG. There was one provision of the 1955 Decree-Law which most of those employed in national statistical services must envy. This was Article 39 which provided for "the imposition of sanctions on anyone who slanders the work of the NSSG". To be so protected in one's capacity as an official collector and processor of statistics against defamation must indeed be unusual. His experience in Ireland suggested that on occasions the media and others were quite prepared to criticize the actions of officials and institutions and in so doing to detract from their public standing - it mattered little on some of these occasions whether or not the criticisms were valid.

Industrial surveys/restrictions on publication. An interesting aspect of Greek practice - in this case a non-statutory measure - related to industrial surveys where data might not be published on branches of economic activity where there were fewer than three concerns. This too was the

practice in Ireland but unlike Greece, there were no exceptions save where the respondent gave written permission to publish. In Greece it appeared that in some cases at least, confidentiality in this sphere did not extend to all the information collected - specifically in such cases confidentiality extended only to employment and the nature and size of energy sources used by the firm. Did this imply that turnover, for example, might be published? However, in Section 3 and Section 4 the opposite appeared to be implied with respect to stated practices of amalgamating unique activities with related branches and obtaining the written permission of the responding unit or units.

Confidentiality vis-à-vis other State agencies. Turning to the relationship with other State agencies, the access to and use of individual statistical information by such agencies raised issues in practically all countries. In Greece, specific provision was made to help the Directorate of National Accounts to fulfil its objectives. This was of course, a statistical objective and as such therefore the provision was not of an exceptional nature - it appeared to be necessary because the NSSG did not compile the national accounts.

NSSG access to data held by other agencies. With regard to NSSG's access to data held by other State agencies, Mr. Embleton endorsed the sentiments that maximum use should be made of such information to improve both quantitatively and qualitatively the corpus of statistics. In Ireland too, they had suffered from the same defects as in Greece - there was a lack of adequate coordination resulting in incompatible data sets, under-utilization of the available information, overlapping data requests, etc. In this matter, there was a lot to be learnt and hopefully copied from some countries whose systems, as outlined in the seminar, illustrated the benefits to be gained from improved coordination and control of the States' statistical information.

New technology and data confidentiality. Turning to the impact of new technology on data confidentiality, Mr. Embleton did not fully share the view that one great advantage of the resultant ability to construct data banks was that these provided greater protection of personal data. He readily accepted that such data banks were less accessible in some respects than manual files but the facility to analyse more easily and quickly data in ever

greater detail increased the risk of inadvertent disclosure of confidential information. He believed too that respondents were becoming increasingly alert to this possibility and did not view computerized files as providing greater assurance of protection. The position was exacerbated by the potential to link files, a development which added to the fear of the respondent. This, as was noted in the Greek paper, was giving rise to growing concern on confidentiality. Existing confidentiality laws, if enacted some 20 to 30 years ago, needed to be revised to take account of the new environment and thus provide the necessary assurances. It was his view that guarantees of confidentiality had to be reinforced and extended to include the right to privacy and the right to know if the fears of respondents were to be allayed.

In this context the basic principles set out in Section III-2 appeared to cover adequately the linking of records. The second principle - the assurance to the private individual or firm that no use would be made of their personal data without their personal consent - was interpreted as extending to cover linkages of data sets relating to the same individual or concern. This in turn granted to the respondent the right to know if linkage of records was involved and with this, the right to refuse if he felt his privacy was threatened. For the statistician the danger in such an arrangement was that information might be under-utilized. What happened for example if the possibility of linkage arose at a much later stage than the actual collection (e.g. in the case of certain types of longitudinal analyses) - were respondents to be contacted anew for the necessary consent? If not, did this preclude such linkages or could the statistical authority exercise discretion? The issues raised were not by any means one-sided.

Confidential and non-confidential questions. Finally turning to the restriction of confidentiality to selected sensitive topics in questionnaires with remaining topics being non-confidential, a number of aspects arose. It was hard to imagine a compulsory or statutory statistical inquiry in which the guarantees of confidentiality did not apply to all questions or topics irrespective of how innocuous any or all of these might be. What was deemed confidential by one individual, might not be so to others. Any general rule would inevitably have exceptions which, in the interests of protecting respondents, must be taken into account.

There was also the possibility that a combination of non-confidential characteristics could breach overall confidentiality - this was equally the case for surveys in which all information was treated as confidential. Thus the exemption from confidentiality of particular topics did not imply total freedom of publication.

Further there were the effects on respondents of such practices. The dilution of confidentiality might readily suggest that the State's needs, in this case statistics, transcended the individual's right to privacy. This could not enhance the public standing of and trust in the national statistical service. Once diluted in this way, the question might be legitimately asked - where would it end? Mr. Embleton suggested that this was not a question statisticians could afford to have asked.

Mr. WERNER (*Statistisches Bundesamt*) remarked:

He noted a number of similarities between the Greek and German systems for protecting personal data. At one point in the paper however, because of the context in which a specific reference was made to the German situation, the danger existed that a wrong conclusion would be drawn concerning the German legal basis. It was stated that the principle that certain questions only in a questionnaire might be confidential was applied in Germany also and that name and address of an undertaking would not form part of the confidential data. It was described how, in the case of the last count of enterprises involving nine questions, two questions only were considered as confidential (section IV-2). In fact, Mr. Werner explained, in the Federal Republic of Germany all individual data entering into the federal statistics were confidential. This meant that there was no difference whether an item of data was more or less sensitive, the decisive factor was simply whether or not it was obtained in the course of a statistical inquiry. Normally to carry out a statistical survey of enterprises required knowing the names and addresses of the undertakings in advance so that these were not items of information collected for federal statistics. In this special case one could take the view that these were not subject to the special protection of statistical secrecy.

The author replied to the discussion:

Under Greek data protection legislation, the only person who was entitled to see an item of individual data was the person to whom it related, and that a formal application had

to be made for this purpose. This applied not only to data supplied for statistical purposes but also to legal records, tax records etc. He took note of Mr. Werner's precisions concerning the circumstances under which name and address were not classified as confidential information but remarked that the Greek system of designating only certain sensitive questions on a questionnaire as confidential had not given rise to any difficulties.

RECORD LINKAGE v CONFIDENTIALITY FROM THE PERSPECTIVE OF THE
U.S. BUREAU OF THE CENSUS

DAWN D. NELSON
Bureau of the Census

SUMMARY

Unlike some European Countries, the United States does not have a registration system in which people are identified by a number (1). The absence of a registration number, however, has not seriously limited the operational feasibility of linking individual record data from two or more files. Consequently, the United States, like European countries, is questioning whether such record linkages violate confidentiality guarantees made to respondents. This paper describes the way in which one U.S. Government agency, the Bureau of the Census, views this issue and copes with the problems. The discussion primarily refers to data on individuals. The specific issues and problems associated with data on establishments are not addressed; however, there are many similarities.

0. INTRODUCTION

The Census Bureau readily recognizes the benefits that are gained by linking an individual's survey record with records from other sources. In fact, the Bureau is authorized by law (Title 13, U.S.C.) to obtain statistical or administrative records to improve data quality and to reduce the amount of data collected and the collection costs. For example, we can verify the accuracy of respondent-reported data on the receipt of a government program benefit by comparing it with the program's records. We also can use program data to supplement household survey data, thereby reducing respondent burden and costs. Usually, however, the records from other agencies are used in studies to evaluate and improve the quality of the demographic censuses.

1. CONFIDENTIALITY AT THE CENSUS BUREAU

The records obtained by the Bureau are treated in the same manner as our confidential survey data. That is, the data are used for statistical purposes only and are not released in a form whereby any particular establishment or individual

can be identified. Furthermore, we have interpreted our confidentiality law (Title 13, U.S.C.) so strictly that we will not even return individually identifiable records in their original form to the agency that provided them. We also have implemented procedures regarding the release of Bureau-collected data that are designed to prevent linkages with records from other sources.

The Bureau releases microdata files from its demographic censuses and surveys and from reimbursable household surveys conducted for other agencies. These microdata files consist of computerized records containing data for individual respondents. In compliance with our confidentiality law, direct identifiers such as name and address are removed from the individual records before release. This step is not sufficient, however, to prevent record linkage by computer matching. The remaining personal characteristics such as sex, race, and date of birth are commonly found on records held by other agencies; for example, records used to administer government programs. If the records in another file contain sufficient data in common with our records, a match can be made. As a result, the individual's identity may be disclosed through directly identifying information in the other agency's file.

To reduce this potential for matching, the Bureau examines files for problems on a case-by-case basis before release. For example, we look for: 1) highly specific characteristics that are also found on an administrative list such as vehicle make, model, and year; 2) information transcribed from an external source that could serve as a link back to that source; and 3) data provided by a respondent that is a matter of public record. The disclosure risk associated with any of these particular items depends upon many factors including: 1) the total number of variables available for matching purposes, 2) the resources needed to perform the match, 3) the age of the data, 4) the accessibility, reliability and completeness of the external file, and 5) the sensitivity or uniqueness of the data. The process used to consider these factors is primarily subjective.

A number of actions, short of withholding the file, are possible if an item presents a disclosure problem. The easiest alternative probably is to delete the item(s), but it is perhaps the most detrimental to users. Another option

is to collapse data values into appropriately broad categories. For example, age might be provided in 5-year intervals instead of single years. Other techniques involve the introduction of ambiguities into the released data so that it will not correspond exactly to data in other files. These techniques include conventional rounding, random rounding, controlled rounding, and random perturbation. (For a discussion on the advantages and disadvantages of each of these techniques, see Cox, et.al., 1984 (2)).

Thus, despite the Bureau's own use of record linkage, we are viewed as an obstacle to efforts involving computer matching by others. We believe, however, that our restrictive confidentiality policies and practices are essential. Our survey respondents usually participate voluntarily and receive no direct benefit from participation. Our ability to ensure confidentiality, therefore, is seen as an inducement for respondents to cooperate and to provide accurate data. We think our consistently high response rates (90-95%) in most household surveys provide confirmation of this belief. Consequently, the Bureau generally has opposed proposals that would weaken our confidentiality law or in any other way affect our ability to protect the confidentiality of respondents' data.

2. CONFIDENTIALITY IN OTHER AGENCIES

Arguments that other agencies could protect our data from disclosure and non-statistical usages have not been convincing to the Bureau. Only a few other agencies within the U.S. Federal statistical system have equally strong confidentiality legislation. Most other agencies rely on internal rules and regulations regarding confidentiality, which do not adequately protect statistical data. Or, they rely on the Privacy Act of 1974, which requires the agency to have the individual's consent to allow information to be disclosed to another agency. This requirement has been weakened, however, by two exceptions. An agency may disclose personally identifiable information if: 1) the disclosure is defined *a priori* as a "routine use" for which the data are being collected; or 2) the disclosure is in response to a request by another agency for information related to the investigation of a specific violation of law. In the absence

of a confidentiality law to the contrary, an agency can establish computer matching as a routine use of personal records by simply publishing a statement to this effect (in a government publication called the Federal Register)(3).

Finally, in all but a few agencies such as the Census Bureau and the Internal Revenue Service, the promise of confidentiality can be overridden by the Freedom of Information Act (FOIA). The FOIA requires the disclosure of existing agency records upon request unless it can be demonstrated that the disclosure would be a "clearly unwarranted invasion of personal privacy". This vague language has resulted in some, although not many, court orders to reveal information collected in confidence.

As a result, if identifiable data were to leave the Census Bureau, we could not be assured of its protection. In many cases, the receiving agency could release the data to another agency or even to someone in the private sector. Credit, banking, insurance, and other kinds of records in the private sector are not subject to Federal protective statutes. Therefore, these record-holders probably would be very interested in obtaining Bureau records for matching projects. As the number of people with access to Bureau data increased, the risk of actual or perceived misuse of the data would also increase.

3. THE ROLE OF PUBLIC OPINION

The lack of uniform government-wide protection may be a factor in the public's perception of the risks involved in providing survey data. Results from a study in 1983 showed that the proportion of the public expressing concern about threats to personal privacy increased from 64% in 1978 to 77% in 1983. This study also found that over a third of the public believed the Internal Revenue Service and the Federal Bureau of Investigation "probably share" information on individuals with others; about half thought the Census Bureau did (4). Other studies during the same time period show similar results. Also, there is some indication that the public's concern is increasingly linked to a general mistrust of government use of computers. Newspaper stories on military surveillance of civilians, wire-tapping, and political espionage help to create a climate that influences such beliefs. And currently, the press is reminding us frequently about the specter of "Big Brother" from George Orwell's book, 1984.

In addition, there are instances of the government using computer matching to identify suspected law violators, or other types of fraud or abuse. Several instances were recently described in an article on this subject (5). The examples included a matching project designed for the Department of Health and Human Services to discover double payments to doctors. Apparently, some doctors submit bills for the same service to two different medical care programs (Medicare and Medicaid). That Department also has recently started to compare records of persons receiving benefits against records of deceased persons. This is an effort to discover cases where the agency is not notified of a death, and the payments are being collected illegally by another person. Matches of this type usually only involve administrative record data, but the public may perceive them as a step towards the creation of a data bank containing information from many sources.

Also, statistical data are used occasionally in a manner that makes it appear that a match has been performed. For instance, the Internal Revenue Service compares summary census information on small geographic areas with Federal income tax returns for the same area. A person's reported income is compared to the income level in his/her surrounding community. An individual reporting an income that deviates substantially from the expected may become the object of further investigation. Press reports of these activities are easily misunderstood by the general public and create the perception that individual statistical records are being used for enforcement purposes.

The effects of this distrust are not well-known; however, the Bureau has assumed that the consequences are lower response rates and poorer quality data. Concern over public opinion and perception appears to be well-justified, especially in light of recent events. The power of the public was recently exhibited by the postponement of the West German census due to fears over potential misuses of the data by the government. In the U.S., public protests played a large role in defeating a proposal designed to facilitate the sharing of confidential data among Government agencies.

4. THE FATE OF A DATA-SHARING PROPOSAL

This data-sharing proposal was made in 1983 by the Office of Management and Budget, which houses the Government's central statistical policy office. Briefly, the proposal recommended designating "a limited number of statistical agencies as 'protected statistical centers', (which) would be authorized both to create and to receive protected statistical files. Other agencies would be authorized to protect their own statistical files, but would not have any privilege of access to protected statistical files other than their own". Under this proposal, a protected statistical file could not have been released for general law enforcement purposes, for national security investigations, or for FOIA requests. Department heads responsible for protected statistical centers would have the authority to designate protected statistical files, and to voluntarily release them for statistical purposes to another center. Sanctions and penalties would have been imposed for any infractions of the rules and regulations governing the data transfers (6).

One of the main objectives of the proposal was to increase access to data for statistical use, including access to administrative data. Such shared data could be used for matching and to improve sample selection procedures, thus introducing efficiencies and reducing survey costs. The legislation to allow this sharing also would have provided some much needed consistency in confidentiality laws and would have protected data that are not adequately protected now.

One proposal, however, generated considerable negative press coverage. The Washington Post succinctly described such proposals as "dreadful" in an editorial on November 22, 1983. They elaborated on this point of view by explaining that government institutions like the Census Bureau have built confidence in their data safeguards over very long periods of time and that any exceptions to established confidentiality practices would threaten this confidence. Many editorials and articles put the argument in terms of efficiency and cost saving considerations versus damage to the public's trust in guarantees of confidentiality. Most of them agreed with the Providence (Rhode Island) Journal's conclusion (November 22, 1983) that "...it is doubtful that the imagined gains would be worth the price". This adverse reaction by the news media, in part, precipitated a withdrawal of the proposal.

This proposal was supported by many of the statistical agencies; however, the Census Bureau did not support it. The Bureau believed that the safeguards against disclosure were inadequate. Also, a number of operational difficulties were anticipated given the nature of the decentralized statistical system in the U.S. Consequently, we thought the risk of losing public confidence and respondent cooperation would be increased (7).

5. RESOLVING THE CONFLICT

In some ways our position on confidentiality conflicts with our primary mission. This mission is to collect, process, and disseminate statistical information in a manner that serves the needs of the user community, particularly those of other Government agencies. Obviously we are not fulfilling one of those needs; that is, the need for individually identifiable data for statistical and research studies, particularly those involving record linkage. We are, however, attempting to achieve a sensible balance between the competing goals of maintaining respondent confidentiality and meeting the vast informational needs on society.

To meet the demands for information, we release microdata files; to ensure confidentiality, we strip these files of information that is considered identifying. Despite our efforts, some very small risk of disclosure remains because of the overwhelming amount of detail present on released files. For example, detail on a person may include a combination of unique characteristics that could conceivably permit identification by others who "know" or recognize the person. Or, as discussed before, the detail may be used in matching with other files. It is very difficult to reduce this detail for confidentiality reasons and still provide useful data to the users. As a result, some people argue against the practice of releasing microdata. The Bureau believes, however, it can reduce the risk of disclosure to an acceptable level. So far, we have been successful; we do not know of any disclosures that have resulted from the release of our files.

The Bureau has established broad criteria to guide the microdata release process in relation to confidentiality issues. Summarized briefly, there are two basic criteria:

- 1) no geographic area with fewer than 100 000 inhabitants

can be identified, and 2) the subject content of survey design cannot present an unusual risk of individual disclosure. Depending on the file content, the geographic limitation may be raised to decrease the likelihood that an individual could be identified by others. The rationale for this is that an individual is less likely to be unique within a larger population group; that is, more than one individual probably will have the same combination of characteristics provided on the file.

Prior to 1981, the geographic limit was 250 000 persons. This was lowered to be more responsive to users who need detailed geographic information for small area data analyses. The lowering of the geographic criterion made it necessary to reduce the amount of detail that might be available to identify an individual. As a result, the Bureau established a Microdata Review Panel to ensure that appropriate disclosure-avoidance measures are taken before a file is released.

The Microdata Review Panel is comprised of six Bureau employees who review each proposed microdata file to identify and resolve potential confidentiality problems. Before the Panel was established, this function was performed by the manager who was assigned overall responsibility for the survey project. Under this system, the managers of different surveys occasionally made inconsistent decisions regarding problems and solutions. By centralizing this function, these inconsistencies were eliminated for the most part. More research is needed, however, to provide an objective basis for making decisions on how much detail to release. In the absence of an objective disclosure analysis process, the Panel is usually fairly conservative in making its determinations. With the perfection of disclosure-avoidance techniques and tools to assess the effectiveness of these techniques, it may be possible to release even more detail on microdata files.

Finally, it is possible that a data sharing proposal, which is acceptable to the Bureau, may be developed in the future. We are currently supporting the development of a proposal concerning the use of the Bureau's Standard Statistical Establishment List (SSEL) by other statistical agencies. The SSEL is a comprehensive list of economic establishments maintained by the Bureau. Other agencies want access to this list to improve the efficiency of certain statistical activities such as sampling. As the proposal is currently envisioned, we would be allowed to maintain control over the uses of the list by other agencies. This is the primary reason the Bureau is looking favorably on this proposal.

Also, we assume that it will provoke less reaction from the public than a proposal leading to the widespread exchange of personal information about individuals. This limited proposal might be a first step in reaching agreement on the way to provide greater access to records for statistical purposes without reducing our ability to protect the individuals who provided the data.

Notes and References

- (1) There is a numbering system associated with the U.S. Social Security Program that shares some similarities with a population registration system; however, every individual does not have a Social Security Number.
- (2) Cox, Lawrence H.; Johnson, Bruce; McDonald, Sarah-Kathryn; and Nelson, Dawn D. (1984), "Confidentiality Issues at the Census Bureau". Paper currently in preparation for presentation at the Census Bureau's first Annual Research Conference in 1985.
- (3) Shattuck, John (June, 1984), "Computer Matching is a Serious Threat to Individual Rights", Communications of the ACM, Volume 27, Number 6, 538-541.
- (4) Harris, Louis and Associates, Inc. (1983), The Road After 1984. A survey on the Impact of Technology on Society sponsored by Southern New England Telephone Company.
- (5) Kusserow, Richard P. (June, 1984), "The Government Needs Computer Matching to Root Out Waste and Fraud", Communications of the ACM, Volume 27, Number 6, 542-545.
- (6) Clark, Cynthia Z.F., and Coffey, Jerry L. (1983), "How Many People Can Keep a Secret? Data Interchange Within a Decentralized System" Proceedings of the Survey Research Methods Section - American Statistical Association, Washington D.C., 580-585.
- (7) Kincannon, C.L. (1983) "Discussion", Proceedings of the Survey Research Methods Section - American Statistical Association, Washington, D.C., 608-610.

DISCUSSION

Mr. FORECAST (*Central Statistical Office, London*) opened the discussion:

He remarked on similarities between the situation in the U.S. described in the paper and that existing in the U.K. Both countries had highly decentralized statistical systems and neither had registration systems involving personal identification numbers. Although the paper had suggested that record linkage on the basis of name and address was found by the Bureau of the Census to be relatively easy, subsequent remarks during the seminar had nevertheless identified certain difficulties. This confirmed the U.K. experience which had established that linkage between different statistical records was an expensive operation in the absence of a personal number.

Mr. Forecast described two experiments in record linkage in the U.K. The first was the Office of Population Censuses and Surveys longitudinal study which followed up the experience of a one per cent sample of persons drawing from the 1971 census, using registration data for births, marriages and deaths. The second was a study of a 10% sample of school leavers, following their subsequent experience in higher education. Neither had given rise to objections on ground of confidentiality and, although the matching procedures were very successful, they were expensive. Both had the feature that the work was entirely within one Department, namely the Office of Population Censuses and Surveys in the one case and the Ministry of Education in the other. As in the U.S., proposals for inter-departmental record-linkage had been rejected. Mr. Forecast drew two general conclusions from this experience. Firstly, record linkage should be done on a sample basis only: the public would have much less fear of possible official misuse of such linkage if it were on a one per cent than on a 100% basis. Secondly, administrative data might be used for statistical purposes but not vice-versa.

Mr. BARNES (*Office of Population Censuses and Surveys*) remarked as follows:

There was an apparent contradiction between the remark in Section 1 of the paper to the effect that the good

confidentiality record of the Bureau contributed to its high survey response rates and the statements later in the paper concerning public scepticism regarding confidentiality.

Mr. BEGUE (*Institut National de la Statistique et des Etudes économiques*):

He noted the statement in Section 1 of the paper that the controls applied by the Bureau to eliminate the possibility of matching in microdata supplied to users were essentially subjective. Returning to the question in Section 5, the need was mentioned of finding objective criteria for this purpose. Mr. Bégué doubted whether such objective criteria could be found because there was no way of knowing what other information might be available to the recipients of the microdata. The development of technology risked making out of date very quickly objective methods developed today. In any case, it was inadvisable to draw up restrictive rules based on such methods. The important thing was to firmly establish and guarantee the principle of confidentiality and leave statisticians a certain flexibility in its application.

Mr. HARRIS (*EUROSTAT*) remarked as follows:

The impact on the compilation of statistics of public concern about confidentiality had not been nearly so great in the U.S. as in Europe. This was, in part, due to the greater size of the country with a greater openness towards information. This was well illustrated by the willingness on the part of firms to provide for statistical purposes the type of information which it was impossible to obtain in Europe. He commented that the opening paragraphs of section 5 summarized the entire problem of statisticians in striking a balance between respondent confidentiality and the informational needs of society.

Mr. COX (*US Bureau of the Census*) remarked as follows:

Regarding the points raised in discussion to the effect that the development of generalized disclosure avoidance methods and software, such as were being pursued in the United States, would lead to rigid disclosure avoidance procedures, with negative consequences, he believed that this was a potential problem at worst and that it could be managed. It was considered that the gains achieved by developing superior disclosure avoidance methods - improved confidentiality protection, more uniform criteria applied uniformly, and due to their repeatability, the ability to evaluate these methods scientifically (to

name but three) - far outweighed potential problems in their use. Having developed such methods, a statistical agency knew rather than speculated about their efficiency and effectiveness and effects upon data quality and completeness. In their absence, this was not possible. The ideal situation would be to have automated methods invoked interactively by a data analyst, thereby combining statistical methodology and the power of the computer with human expertise.

The author replied to the discussion:

She said that in the Census Bureau they had not used matching as a source of data but rather as an additional check for evaluating the quality of data or supplementing the collection effort. Matching did not, therefore, represent any cost saving but was an additional cost. For this reason the amount of matching done was fairly limited. She agreed that there was a contradiction in the paper regarding the influence of confidentiality on response rates. She felt there was a need for further studies on this point to analyse the difference between what the public said and what it did. It was not clear that people had fully understood the questions put to them in the survey referred to in Section 3 where they expressed the view that census information was shared with other departments.

Mrs. Nelson thought it was worthwhile pursuing the search for objective standards for preventing the inadvertant disclosure of confidential data. The main problem was the cost of such a study or of writing computer programs for the job.

THE SITUATION IN SWITZERLAND

H. BRUENGGER
Office statistique fédérale, Berne

Switzerland is not regarded as one of the most progressive countries as far as statistics are concerned. The malicious comment made by some of our compatriots is that Switzerland is prosperous precisely because there are so few statistics. These people, however, do not realize that agriculture, for example, is a thriving industry in Switzerland even though (or because) it is in the exceptional position of being the subject of a great many, high quality statistics.

Nevertheless, Switzerland clearly lags behind other European countries in providing legislation for statistics and data protection. At national level, the legal basis for official statistics is fragmentary, heterogeneous and in most cases dates back to a period in which concepts such as "data protection" or "personal data" were completely unknown. Two years ago, we therefore made public the draft of a new Statistics Act and this has been accorded a surprisingly large measure of approval. This bill could have been put before Parliament had the Government, because of pressure of parliamentary business, not refused to give it urgent status.

A similar stage has been reached in data protection legislation at national level. Here too, the bill was made public a year ago and met with a very varied response. The major bones of contention were, however, legislation covering data protection in the private sector, i.e. in companies, and the constitution and jurisdiction of a data protection commission to ensure proper protection of data in both the private and public sectors. There has, however, been less controversy about the data protection provisions in the actual Federal order. Again, it will be some years before this Act can come into force.

The fact that work on two bills was carried on almost in parallel necessitated closer coordination and, looking back, it is interesting to see how close and effective the cooperation became between the National Statistical Office and the authorities responsible for data protection. This personal relationship has made it possible for us to influence the draft Data Protection Act so that matters of particular statistical interest can be taken into account. These include, in particular:

- 1) the option of using administrative data for statistical analyses even where no express mention was made of statistical use at the time that the data were obtained;
- 2) the possibility, while respecting certain provisos and conditions, to provide third parties conducting their own statistical work with individual data from statistical studies - these data being shorn of names and addresses but not rendered completely anonymous. Cantonal and municipal statistical offices comprise some of these third parties, others being scientists and researchers;
- 3) the decision not to grant individuals the right of access to data about them or the right of correction where these data are unsatisfactory. This provision does not apply to the company register but I would like to point out that, in the field of personal rights, Swiss law does not distinguish between physical and legal persons so that data protection regulations are just as applicable to data about individual companies.

The data protection provisions of this bill do, even though they still have no legal force, already form part of current practice. In particular, where data of the type described above are passed onto cantonal statistical offices or to researchers, those receiving them must enter into a contractual obligation to observe the conditions laid down in the future laws. This is, however, done only by the National Statistical Office and is not true of the whole statistical sector at national level. Since other Government agencies do not draw a clear line between statistical and administrative activities, we also refuse to provide them with any data that have not been rendered completely anonymous, even though their declaration that these are to be used only for statistical purposes would be sufficient legal justification to do so.

The above analysis shows that we in Switzerland are moving towards legislation that strikes a balance between the needs of the data protection and statistics.

CLOSING REMARKS

G. ALS

Service Central de la Statistique et des Etudes Economiques,
Luxembourg

It is difficult to distil the results of some two and a half days of wide-ranging discussion into a reasonable compass. These closing remarks will therefore not be comprehensive but will rather concentrate on a number of different aspects. The task of making a selection has not been easy: the seminar aroused great interest and opinions based on quite different viewpoints were expressed by the different groups of participants. Inevitably there is a risk of being somewhat subjective and these remarks will indeed be the observations of an official statistician very much concerned with the dilemma in which official statisticians now find themselves.

In order to appreciate the present situation, it is necessary to cast a backward glance at how the situation has developed over the last decade. The concern of statisticians to maintain the secrecy of information about individuals and individual businesses is traditional and well known. Historically, official statisticians have jealously guarded the confidentiality of the individual information entrusted to them. They have developed techniques to avoid not only direct disclosure of information, but also to avoid risk of indirect disclosure, say by the calculation of differences between tabulations which may cover slightly different areas. Conventions have been developed about the minimum number of observations to be included in published aggregates and legislation has laid strict obligations on enumerators so as to prohibit voluntary disclosure or misuse. In some countries indeed, the statistical services have been given a certain measure of independence from Government so as to demonstrate their impartiality and integrity. In general, high standards of professional secrecy have been set and maintained. This concern with confidentiality has indeed been seen by statisticians as an essential prerequisite for carrying out their statistical work. The success of any statistical survey depends on the cooperation of respondents and an important element in securing such cooperation is the guarantee that the replies will be kept confidential.

It could seem at least possible that the general satisfaction of statisticians with their achievements in respect of confidentiality may have led to a certain smugness. Be that as it may, the concern over privacy which developed suddenly in some countries during the 1970s caught many official statisticians by surprise. The composition of the seminar ensured some airing of all points of view.

Academics, data protection officials and statisticians were all present. Inevitably, the statisticians predominated in a seminar organized by the Statistical Office of the European Communities and this meant perhaps that the dialogue with the data protectionists was not taken as far as it could have been. Nevertheless, there was a general agreement that the papers produced by countries setting out the position in individual countries with regard to data protection legislation and the general account of the state of the art in reaching accommodation with the legislation would provide an invaluable reference library.

Six general conclusions emerge from the seminar.

1. The first basic conclusion was that statisticians and data protection officials needed to cooperate in order to fulfil their respective tasks and in order to clear up misunderstandings and reduce tension. Confrontation would be a recipe for disaster. It had to be recognized that the conflict between several elements in current thinking had yet to be adequately resolved. These elements included the political need to protect individual privacy, the need to maintain the flow of information for government decision making, and the need for the public to be aware of how government decisions are made. Beyond the needs of government, there were the reasonable needs of social and other researchers to be taken into account. If the needs of government and the wishes of the public generally were to be met, it was necessary to strike a balance between requirements which were in some sense in conflict. It was comforting to note that recent laws on the protection of privacy have generally - though not always - involved consultation with statisticians prior to enactment. It was particularly reassuring and encouraging to note that in France it was the data protectionists themselves who began such discussions with a view to having the existing legislation modified in order to take greater account of statisticians' needs. Such legislation was referred to during discussion as second-generation laws. Statisticians can perhaps take some encouragement from such developments.
2. Statisticians recognize that - like true scientists - they have been living for far too long in an ivory tower from which they need to emerge. Speakers suggested that they should become politically expert or if that were too much to ask, they should at least acquire a commercial consciousness and play an educational role. A

paradoxical situation has come about, namely that while all Governments affirm that statistics are indispensable for helping them draw up rational policies, statisticians themselves experience great difficulty in convincing the public of the usefulness of statistics. Undoubtedly statistics provide an infrastructure for decisionmaking but it is difficult to explain this to the man in the street.

3. It is important to make clear both to politicians and to the general public that statistical offices are not just administrative institutions. Certainly they were created by the politicians but they have since demonstrated their integrity and independence, perhaps particularly where methods are concerned, and have, in many instances, now become scientific rather than administrative institutions. This special status was lent extra weight by the principle of statistical confidentiality, which statisticians guarantee to all suppliers of data.
4. Statisticians are at present in a critical situation in some countries and potentially in all countries. On the one hand, they have to contend with a lack of funds, people's reluctance to provide information, and pressing demands to process administrative data rather than organize surveys and censuses. On the other hand, they see new barriers going up against them when they try to get access to administrative data sources. There is here a dialectical contradiction which can be resolved only by a revolution in attitudes.
5. Karl Marx taught that there is a dialectic of history: thesis - antithesis - synthesis. Statistics are now at the antithesis stage. During the seminar the position was stated more simply, it being described as a swing of the pendulum. The pendulum has swung too far. It will return sooner or later, but not under its own momentum. It needs a push from the statisticians. From this point of view, it is reassuring to note that the formulation of recent laws on the protection of privacy have involved consultations with statisticians.
6. The final general conclusion concerns the relative interpretations of constitutional rights guaranteeing privacy and individual freedom. In Germany, it has taken more than 30 years to arrive at the present interpretation of the constitution. In many other countries the same constitutional rights are interpreted in a way which is more favourable towards statistical work.

SPECIFIC CONCLUSIONS

The seminar also produced a number of specific conclusions:

7. Advantages of the new legislation

Although in many cases the new laws on the protection of privacy have made the statistician's work more difficult, they have also brought certain advantages: they have obliged statisticians to give even deeper thought to the question of statistical confidentiality, induced them to bring in even stricter rules concerning enumerators, and made them re-examine their arrangements for working with local administrations in the collection of statistics.

8. Statistical confidentiality

The seminar made very clear that there is no general agreement among statisticians on the importance of statistical confidentiality. Many countries spoke on this subject - the United Kingdom, Belgium, Greece, the Netherlands, France, etc. It is firmly established that statisticians interpret confidentiality as meaning they must on no account disclose personal data to third parties, including Ministers of State.

9. Depersonalizing files and historical needs

Several seminar participants stated that depersonalizing computer files went against the interests of historical research and the needs of future generations. However, two separate points emerged from the discussions:

- original documents containing personal details remain in the State archives. As for computer files - assuming they still contain personal details - it was wondered whether the technology was not developing at such a pace that future generations would no longer have the outmoded equipment necessary for reading them.
- any advantage there might be in keeping computer files containing personal data would tend to be of a short-term nature: it would mean that longitudinal analyses and other studies of the data could still be carried out in the more immediate future.

10. Unscheduled statistics

It was pointed out that when a major statistical census is undertaken it is impossible to predict every single table which the results may yield. Statisticians therefore need to be able to compile and extract from the tapes statistical tables which were not scheduled at the outset without contravening the principle of finality laid down in the Strasbourg Convention.

11. Samples taken from the population census

In general, problems of data protection do not arise in connection with surveys where the national statistical offices themselves make use of samples taken from the population censuses. The question is whether such samples may be supplied to third parties. In principle, there is obviously a problem of confidentiality here, because the samples contain personal details. But things are not quite so simple as that, because the details involved are fairly innocuous: name and address, sex and age, all details which are more or less public. In addition, the samples would be used for purely voluntary surveys. The actual situation differs from country to country: some see no objection to supplying such samples for scientific purposes, while others at present feel unable to do so. In France the CNIL (Commission nationale de l'informatique et des libertés) is looking at ways of changing the law on this particular issue.

12. One-way transmission of personal data

On this subject there is agreement between statisticians, but not always between statisticians and the data protection institutions. The former consider that the specific nature of statistics justifies transmission of personal administrative data to statistical offices, on the understanding that this is strictly one-way with no question of the inverse procedure being allowed. The situation differs from country to country, but there is a marked tendency towards allowing these one-way data transfers.

- in Denmark they are at the very heart of the statistical system
- in Germany one-way transmission is possible provided it is authorized by a law, which is a general rule of the German statistical system
- in France it is either prohibited or limited for the moment, but the CNIL is considering ways of changing the legislation on this particular issue.

13. The "forbidden fruit" of statisticians

Generally it was agreed that official statisticians should not ask questions on people's private lives nor on their religious, political or philosophical beliefs. The Strasbourg Convention allows for exceptions to this prohibition in cases where the national legislation concerned provides appropriate safeguards. Several countries stated that they do ask some such questions but leave the reply optional.

The final conclusion to be drawn from the seminar is that it has been an extremely worthwhile exercise, clarifying many urgent present-day problems and sometimes indicating paths towards their solution.

LIST OF PARTICIPANTS / LISTE DES PARTICIPANTS

Chairman/Président: G. ALS, Directeur général, Service central de la statistique et des études économiques - Luxembourg

Rapporteur: J. KNAGGS, Statistical Office of the European Communities

1. Member States / Pays membres

FEDERAL REPUBLIC OF GERMANY / REPUBLIQUE FEDERALE D'ALLEMAGNE

Statistisches Bundesamt	G. Bürgin E. Südfeld J. Werner
Bundesbeauftragten für Datenschutz	R. Baumann U. Dammann

FRANCE

Institut National de la Statistique et des Etudes économiques (INSEE)	M. Bégué C. Moreau
Conseil d'Etat	N. Lenoir J.E. Schoettl
Commission Nationale de l'Informatique et des Libertés	J.E. Vié
Université Paris X Nanterre	J. Zighera

ITALY / ITALIE

Istituto Centrale di Statistica (ISTAT)	G. Cariani G. Giacummo R. Tomei
Ministero Lavoro	S. Filipello T. Zeuli
Ministero Industria	E. Malcangio
Ministero Giustizia	M. Guardata
Università di Milano	M. Losano

NETHERLANDS / PAYS-BAS

Centraal Bureau voor de Statistiek
(CBS)

W. Begeer
H.D. Dukker
W.F.M. de Vries

European Society for Opinion and
Marketing Research (ESOMAR)

J.H.M.J. Quatresooz

Universiteit van Leiden

P. van Hoesel

BELGIUM / BELGIQUE

Institut National de Statistique (INS)

E. Van Langendonck

LUXEMBOURG

Service Central de la Statistique
et des Etudes économiques (STATEC)

J. Langers
R. Weides

Commission Consultative pour la
protection des données nominatives

R. Biever
G. Schleder

UNITED KINGDOM / ROYAUME-UNI

Central Statistical Office

K.G. Forecast
T.J. Orchard

Office of Population Censuses
and Surveys

G. Barnes
R. Barnes

Social and Community Planning Research

R. Jowell

IRELAND / IRLANDE

Central Statistics Office

E. Embleton

Economic and Social Research Institute

B. Whelan

DENMARK / DANEMARK

Danmarks Statistik

P. Jensen
K. Kjaersgaard-Briksen
L. Thygesen

University of Copenhagen

N. Keiding

GREECE / GRECE

National Statistical Service of Greece

E. Kalambokidis
E. Stiliaris

2. Other countries / Autres pays

SPAIN / ESPAGNE

Instituto Nacional de Estadística V.I. Vasquez Sierra

PORTUGAL

Instituto Nacional de Estatística A. Simoes Ferreira da Cunha

NORWAY / NORVEGE

Central Bureau of Statistics J.K. Tönder

SWEDEN / SUEDE

Statistics Sweden E. Rapaport

SWITZERLAND / SUISSE

Office fédéral de la statistique H. Brüngger

CANADA

Statistics Canada G. Labossière

University of Western Ontario D.H. Flaherty

UNITED STATES OF AMERICA /
ETATS-UNIS DE L'AMERIQUE

U.S. Bureau of the Census L. Cox
D.D. Nelson

3. International organisations / Organisations internationaux

United Nations, New York R. Johnston

United Nations, Economic Commission
for Europe, Genève R. Gentile

4. Commission of the European Communities /
Commission de la Communauté Européenne

Commissioner R. Burke

Member of Cabinet C. Day
Membre du Cabinet

DG IX A. Hoccart

5. Statistical Office of the European Communities - Eurostat
Office Statistique des Communautés Européennes - Eurostat

Director / Directeur	D. Harris
Head of Division / Chef de Division	N. Ahrendt H. Fürst
Head of Section / Chef de Service	K. Löning C. Peare
Administrator / Administrateur	M. Bezikofer H. Schneider M. Skaliotis

European Communities — Commission

Eurostat News — Special edition

Protection of privacy, automatic data processing and progress in statistical documentation

Luxembourg: Office for Official Publications of the European Communities

1986 — 348 pp. — 16.2 x 22.9 cm

Theme 9: Miscellaneous (brown covers)

Series C: Accounts, surveys and statistics

EN, FR

Cat.: CA-AB-85-006-EN-C

Price (excluding VAT) in Luxembourg

ECU 11.37

BFR 500

IRL 8.10

UKL 7.70

USD 11

The papers reproduced in this volume were first presented at an international seminar 'Protection of privacy, automatic data processing and progress in statistical documentation' held by the Statistical Office of the European Communities (Eurostat) at Luxembourg in December 1984. The aim of the seminar was to provide an exchange of views and experience on the range of important problems faced by statisticians and others arising from considerations of privacy and confidentiality, and to encourage an informed discussion on possible strategies for maintaining the flow of statistical information in the face of these problems.

Salg og abonnement · Verkauf und Abonnement · Πωλήσεις και συνδρομές · Sales and subscriptions
Venta y abonos · Vente et abonnements · Vendita e abbonamenti
Verkoop en abonnementen · Venda e assinaturas

BELGIQUE/BELGIË

Moniteur belge/Belgisch Staatsblad
Rue de Louvain 40-42/Leuvensestraat 40-42
1000 Bruxelles/1000 Brussel
Tél. 512 00 26
CCP/Postrekening 000-2005502-27

Sous-dépôts/Agentschappen:

**Librairie européenne/
Europese Boekhandel**
Rue de la Loi 244/Wetstraat 244
1040 Bruxelles/1040 Brussel

CREDOC
Rue de la Montagne 34/Bergstraat 34
3te 11/Bus 11
1000 Bruxelles/1000 Brussel

DANMARK

Schultz EF-publikationer

Væntergade 19
1116 København K
Tlf: (01) 14 11 95
Girokonto 200 11 95

3R DEUTSCHLAND

verlag Bundesanzeiger
Ereite Straße
Postfach 01 80 06
5000 Köln 1
Tel. (02 21) 20 29-0
Fernschreiber:
ANZEIGER BONN 8 882 595
Telecopierer:
20 29 27B

GREECE

G.C. Eleftheroudakis SA

International Bookstore
4 Nikis Street
105 63 Athens
Tel. 322 22 55
Telex 219410 ELEF

Sub-agent for Northern Greece:

Molho's Bookstore

The Business Bookshop
10 Tsimiski Street
Thessaloniki
Tel. 275 271
Telex 412885 LIMO

ESPAÑA

Boletín Oficial del Estado

Trafalgar 27
E-28010 Madrid
Tel. (91) 76 06 11

Mundi-Prensa Libros, S.A.

Castelló 37
E-28001 Madrid
Tel. (91) 431 33 99 (Libros)
431 32 22 (Abonos)
435 36 37 (Dirección)
Télex 49370-MPLI-E

FRANCE

**Service de vente en France des publications
des Communautés européennes**

Journal officiel
26, rue Desaix
75732 Paris Cedex 15
Tél. (1) 45 78 61 39

IRELAND

Government Publications Sales Office

Sun Alliance House
Molesworth Street
Dublin 2
Tel. 71 03 09

or by post

Stationery Office

St Martin's House
Waterloo Road
Dublin 4
Tel. 68 90 66

ITALIA

Licosa Spa

Via Lamarmora, 45
Casella postale 552
50 121 Firenze
Tel. 57 97 51
Telex 570466 LICOSA I
CCP 343 509

Subagenti:

Libreria scientifica Lucio de Biasio - AEIYOU

Via Meravigli, 16
20 123 Milano
Tel. 80 76 79

Libreria Tassi

Via A. Farnese, 28
00 192 Roma
Tel. 31 05 90

Libreria giuridica

Via 12 Ottobre, 172/R
16 121 Genova
Tel. 59 56 93

GRAND-DUCHÉ DE LUXEMBOURG

**Office des publications officielles
des Communautés européennes**

2, rue Mercier
L-2985 Luxembourg
Tél. 49 92 81
Télex PUBOF LU 1324 b
CCP 19190-81
CC bancaire BIL 8-109/6003/200

Messageries Paul Kraus

11, rue Christophe Plantin
L-2339 Luxembourg
Tél. 48 21 31
Télex 2515
CCP 49242-63

NEDERLAND

Staatsdrukkerij- en uitgeverijbedrijf

Christoffel Plantijnstraat
Postbus 20014
2500 EA 's-Gravenhage
Tel. (070) 78 99 11

PORTUGAL

Imprensa Nacional

Av. Francisco Manuel de Melo, 5
P-1000 Lisboa
Tel. 65 39 96

Grupo Bertrand, SARL

Distribuidora de Livros Bertrand Lda.

Rua das Terras dos Vales, 4-A
Apart. 37
P-2701 Amadora CODEX
Tel. 493 90 50 - 494 87 88
Telex 15798 BERDIS

UNITED KINGDOM

HM Stationery Office

HMSO Publications Centre
51 Nine Elms Lane
London SW8 5DR
Tel. (01) 211 56 56

Sub-agent:

Alan Armstrong & Associates Ltd

72 Park Road
London NW1 4SH
Tel. (01) 723 39 02
Telex 297635 AAALTD G

SCHWEIZ/SUISSE/SVIZZERA

Librairie Payot

6, rue Grenus
1211 Genève
Tél. 31 89 50
CCP 12-236

UNITED STATES OF AMERICA

**European Community Information
Service**

2100 M Street, NW
Suite 707
Washington, DC 20037
Tel. (202) 862 9500

CANADA

Renouv Publishing Co., Ltd

61 Sparks Street
Ottawa
Ontario K1P 5R1
Tel. Toll Free 1 (800) 267 4164
Ottawa Region (613) 238 8985-6
Telex 053-4936

JAPAN

Kinokuniya Company Ltd

17-7 Shinjuku 3-Chome
Shinjuku-ku
Tokyo 160-91
Tel. (03) 354 0131

Journal Department

PO Box 55 Chitose
Tokyo 156
Tel. (03) 439 0124

Price (excluding VAT) in Luxembourg

ECU 11.37

BFR 500

IRL 8.10

UKL 7.70

USD 11



OFICINA DE PUBLICACIONES OFICIALES DE LAS COMUNIDADES EUROPEAS
KONTORET FOR DE EUROPÆISKE FÆLLESSKABERS OFFICIELLE PUBLIKATIONER
AMT FÜR AMTLICHE VERÖFFENTLICHUNGEN DER EUROPÄISCHEN GEMEINSCHAFTEN
ΥΠΗΡΕΣΙΑ ΕΠΙΣΗΜΩΝ ΕΚΔΟΣΕΩΝ ΤΩΝ ΕΥΡΩΠΑΪΚΩΝ ΚΟΙΝΟΤΗΤΩΝ
OFFICE FOR OFFICIAL PUBLICATIONS OF THE EUROPEAN COMMUNITIES
OFFICE DES PUBLICATIONS OFFICIELLES DES COMMUNAUTÉS EUROPÉENNES
UFFICIO DELLE PUBBLICAZIONI UFFICIALI DELLE COMUNITÀ EUROPEE
BUREAU VOOR OFFICIËLE PUBLIKATIES DER EUROPESE GEMEENSCHAPPEN
SERVIÇO DAS PUBLICAÇÕES OFICIAIS DAS COMUNIDADES EUROPEIAS