XII. RESPECTING PRIVACY AND PRESERVING CONFIDENTIALITY: HONOURING THE CONTRACT

A. Respondent policies

522. Perhaps the most important issue in developing a respondent policy is that a statistical agency must earn the public trust by treating respondents with respect, not just a means to reach its statistical goals. It is important to remember that, even in the presence of laws that make response to one or more data collections mandatory, participation by the public in statistical agency surveys is a largely voluntary process. Even when the survey process is not voluntary, an agency still has an obligation to treat respondents in an ethical manner: that is, minimizing the burden on their time, respecting their privacy and maintaining the confidentiality they were promised when they provided the information.

523. Respect for privacy is an acknowledgement that it is the individuals who "own" the information about themselves. According to the concept of privacy, it is the individual who decides what information is made available, when it is to be released and to whom it is released. Laws requiring the disclosure of this private information for statistical purposes are enacted only when there is an overriding public need for the information, and they contain provisions to protect from disclosure the identifiability of the data. This concept also to applies legal entities such as corporations. Public corporations, of course, agree to the regular disclosure of information such as sales and profits in order to participate in a regulated stock market. Private or closed corporations are required to make far fewer disclosures.

524. Ensuring confidentiality is the appropriate response from the statistical agency when it obtains private data. It is a pledge to honour the contract between the respondent and the agency when the respondent provides private data. The present chapter considers the elements of this contract and the means to protect the data when it is obtained from respondents.

525. It is critical to develop policies designed to create a cooperative frame of mind on the part of the intended respondents. Listed below are the elements of such policies, followed by an interpretation of each of these elements.

1. Principles of respondent relations

526. There are two broad classes of respondents: businesses and individuals, or households. Certain principles apply to both; others are unique to the business community. Items (a) through (e) are applicable to both cases; (f) and (g) apply only to businesses.

(*a) The purpose of the data collection must be clear and meaningful to the respondent*

Helping respondents understand the reasons for collecting the data is often a difficult task, and the application of this element is particularly difficult in the case of small businesses. The overall framework within which data is collected prevents the statistical agency from taking frequent censuses. Indeed, in the interest of efficiency, it is best to take samples no greater than what is strictly necessary to yield aggregate totals quickly and accurately. However, a small business is typically interested in the locale in which it operates and in its narrowly defined type of activity, neither of which can be adequately represented by most statistical undertakings, with the exception of a census or a comprehensive large-scale survey. For either class of respondent, any explanation of the purpose should be as clear and convincing as possible.

(*b) The statistical agency must be perceived as holding in the strictest confidence all individual records, protecting them from any other party inside or outside Government*

This assurance has to be delivered in a way that makes respondents feel confident.[108] The assurance itself consists of two elements: respondents must be made aware that the information held by the statistical agency cannot be accessed by someone with malicious intent,[109] and that the law recognizes that confidential information held by the statistical agency cannot be shared with the political authorities of the country, with regulatory agencies or with the civil service engaged in policy development. In a number of countries, specific reference is made to the fact that information submitted to the statistical agency cannot be subpoenaed by the judiciary.

(c*) The statistical agency must be seen as willing to accommodate respondents, either by providing additional explanations or by accepting legitimate substitutes for a traditional questionnaire*

In the case of business statistics, the statistical agency should not only be aware of how businesses keep records, but it must also be willing to go out of its way to recover the information it needs with minimal effort on the part of the business. For example, a shareholders' report that includes all or most of the variables sought by the data collector should be an acceptable substitute for a completed statistical questionnaire.

(d) The professionalism and objectivity of the statistical agency, as well as its freedom from political interference, must be established, accepted, and continually advertised

In approaching respondents, a statistical agency ought to request no more

[108] Cases have been recorded in which businesses explicitly stated that they would prefer that even routine information be collected by the national statistical agency, on the grounds that all information provided would be safe from inspection by unauthorized parties.

[109] Of course, this assurance can only be given within limits, but the public must be assured that it would be extraordinarily difficult to penetrate the defenses put up by a statistical agency.

information than is necessary and only after it has determined the least burdensome way of acquiring the information. The request is backed by the agency's reputation, specifically in the sense that the methods it employs to achieve results protect its autonomy and freedom from political interference.

(e) The statistical agency should be perceived as thoughtful and concerned in matters relating to response burden; that is to say, it should be committed to finding means that will simplify the paperwork burden

Generally, it should be understood that the statistical agency will approach a respondent with a request for data only as a last resort, all alternative avenues having been reviewed and found insufficient. An agency that desires to be known for a reasonable policy on matters of response burden will see to it that data collectors have understood thoroughly the meaning of this last element in the policy. In particular, the agency must demonstrate:

- That at all times the statistical agency closely monitors its information gathering activities;
- That the statistical agency has taken into account the intrusion and the violation of privacy necessarily implied by its efforts and that the result is a careful balance between respect for privacy and the minimum information required to enlighten public discussion;
- That the information collected will not be misused;
- That it can measure the extent of the burden it imposes, and that, through the application of its policies, this burden will be reduced.[110]

527. In addition, there are two principles that apply only to the business community:

(*f) The way in which the information is collected must reflect the ways in which businesses keep records*

Successful collection involves the capacity to transform the records held by businesses into the standard records required by the data compiler. However, it is the statistical agency that must make the effort, not the businesses. The extent to which the information can be transformed without distortion can be gauged if the data collector understands how businesses keep records. In general, it is advisable for a statistical agency to a sufficient number of employees with a thorough understanding of business accounting and bookkeeping. Ultimately, an effort should be made to convince businesses that it is safe to share confidential information with the statistical agency.

[110] In some agencies there is an actual annual calculation of total burden imposed on business. This calculation serves as the basis for an official report to the Government on reduction of the total burden over the past year.

(g) The information collected should employ the same terminology used in daily business operations

It is important to distinguish between the language used by statisticians to communicate among themselves and the language required to communicate with the intended respondents. Such expressions as "kind of activity unit" and "net profits" have little meaning outside the context of a statistical agency. The language to be used in order to elicit information must be the language that is familiar to the respondent, whether a household or business.[111]

528. A statistical agency, however, is not the only agency that collects information. Indeed, the Government collects detailed information from the public in a variety of ways and for many reasons. For example, detailed balance sheet information is collected in order to levy taxes and administer subsidies; and detailed commodity information is collected in order to administer laws and regulations that deal with health and safety.[112] The statistical agency should be among the most vigorous pursuers of paperwork reduction, using the information collected by others whenever there is an opportunity created by law and by overlap in content. It should be able to demonstrate to the public that it communicates with other government agencies regarding the means available under the law to prevent duplicate requirements, particularly from those businesses that are least equipped to fulfil them.

2. Compulsion and voluntary response

529. The nature of the law differs from country to country. In some, compliance with statistical collection is simply obligatory. If respondents - businesses or heads of household - do not provide information in the form in which it is requested and in a timely fashion, they are in violation of the law. In other countries, requests for certain classes of information are supported by legal requirements, whereas others are made on a voluntary basis. Finally, there are cases where the law is ambiguous on the subject. When this is true, the statistical agency may be fearful of demanding too much information: if challenged, the law might rule that no information is to be demanded compulsorily (other than that sought by the census), and the resulting publicity might adversely affect response rates.

530. The situation in most countries is probably representative of the second case, in which the law recognizes a restricted set of compulsory surveys. Whatever the legal basis, all agencies find that the most important objective is to secure a cooperative

[111] The Australian Bureau of Statistics engaged in a comprehensive attempt to rewrite its questionnaires for business surveys in the language "used by the trade". The investment required to do this is considerable. But even if it does not bring about a perceptible increase in response rates, it is likely to help reduce response error.

[112] Two exercises conducted at different times in two different OECD countries revealed that the amount of paperwork imposed by statistical collection when compared to all paperwork imposed by Government did not constitute much more than 5 per cent of the total.

attitude on the part of the respondents, particularly from small businesses and households.[113]

531. If an option is available, the statistical agency may adopt a moderate stance, wherein economic inquiries are compulsory and all others are voluntary. Irrespective of the legal posture, it is important to remember that without a cooperative attitude, no amount of compulsion will alleviate the response problem.

3. Incentives and assurances

532. A problem that has been increasing in many countries is that of low response rates in data collections. One of the potential solutions has been to provide incentives to respondents. However, incentives cannot always solve the problem. In some countries the use of financial incentives is illegal. In other cases, the budgets are too tight to allow for the provision of significant incentives. Moreover, if incentives are offered to respondents (e.g., householders), there must be some relationship between the time required of them and the nature of the compensation. For some inquiries the matter is trivial. For example, in the case of attitudinal surveys or continuous surveys, the cost to the householder is the periodic intrusion into the household's private affairs. In other cases, particularly for surveys of consumer income and expenditure, the effort required from the household is considerable.[114] Finally, the payment of incentives may create expectations on the part of respondents and make it difficult to conduct surveys without incentives or require ever-increasing incentives. The following are points to consider:

- The kind of incentive chosen must be significant but not excessive, so that it is neither treated with scorn by respondents nor looked upon as waste by those who determine the statistical agency's budget. The field organization must play an important role in providing advice on this matter;

- Whatever option is chosen, it is best to keep the householders who are surveyed informed about the reasons why incentives were or were not provided, about the image of the statistical agency, and of course, about the civic duty involved in the provision of information to aid in the discussion of matters of public concern.[115]

533. While incentives are most commonly offered to householders, they can also be useful in surveys of small agricultural operators and even of small, informal businesses. The same considerations do not necessarily apply to larger businesses, for which the provision of relevant information may be a forceful incentive in itself. For example, if the

[113] Generally, big enterprises will comply with government requests and will not request elaborate explanations as to why certain classes of financial and economic information are sought. There are, however, examples of transnational enterprises that are uncooperative in providing information outside the country where they are headquartered.

[114] There are many cases in which the survey of income and expenditure is combined with a survey of fixed and financial assets, and the details required from the household on its balance sheet and current and past transactions are such that not only must the household keep a detailed diary of its transactions but the interviewer, too, must spend a significant amount of time helping to complete the questionnaires.

[115] In invoking civic duty, the provision of private information is treated in the same manner as serving on a jury.

business information compiled in typical cases is comprehensive enough to be used in a regional breakdown, businesses operating on a regional rather than a local level will be interested in how their regional results compare to those of their competitors working on the same scale. Accordingly, the incentive of being provided with special tables making those comparisons - so long as the data have the right amount of detail and are reasonably current - may be sufficient to elicit constructive cooperation between the business community and the statistical agency.[116]

534. In all cases, however, the statistical agency must give uncompromising and explicit assurances that the information supplied will not disclose the identity of respondents. For businesses of any size, the combination of strong assurances and the recognition that the information solicited is necessary for the country to manage itself in an orderly fashion is sufficient. In other words, providing incentives in the case of business surveys is generally unnecessary (and, for budgetary reasons, is not in the statistical agency's best interest).

4. Sharing administrative information

535. The sharing of detailed administrative information with statistical agencies has many precedents. The most noteworthy is the case of customs administrations. However, income tax collectors, who are also collectors of massive amounts of information, are just as mindful of integrity and confidentiality as statisticians. Sharing information, particularly with tax collectors, is a delicate matter. Thus, if sharing is to take place, tax collectors must be assured that statistical agencies are at least as scrupulous as they are in the treatment of individual information. The following are some of the factors that the statistical agency ought to bear in mind:

- Whatever information is shared between statisticians and tax collectors must move in only one direction—from the tax authorities to the statisticians;

- In order to get the tax collectors to share information willingly and cooperatively, the statisticians should render some service to the tax authorities, one that does not sacrifice any of the vital safeguards on confidentiality;

- Information collected by the tax authorities is the single most powerful resource for reducing the response burden for small businesses;[117]

- Tax authorities are interested in classifying the forms submitted by tax filers according to their branch of economic activity but not necessarily in the same way as statistical agencies;

[116] In France, the National Institute of Statistics and Economic Studies (INSEE), has successfully pursued a policy of motivating businesses by the provision of custom-made tabulations, including specific comparisons between the business and its peers.

[117] An exception would be in countries where only a minority of the population pays taxes on income, so that it is not worth considering for statistical purposes.

- There is a substantial strategic gain if the industry classification of tax records is aligned with that used for statistical purposes;

- The public perception of statisticians classifying tax records might weaken any assurances given by statisticians about the confidentiality of the information submitted.

536. The challenge for the statistical agency is to find the best possible compromise among these factors without jeopardizing in any way statistical confidentiality. Whatever is found to work vis-à-vis the tax authorities is bound to work, with the occasional variant, for the other suppliers of administratively generated information.

537. In addition to the various guarantees that suppliers of administrative information require and that can be agreed upon by protocol and by verbal assurances, the statistical agency should give signs that the physical holdings of sensitive information are especially well guarded. In many statistical agencies where access to the premises is controlled, access by outsiders to the floors on which tax documents are held requires special authorization.

B. Data protection

1. Physical and electronic security

538. These days statistical agencies have two perimeters that must be protected. The first is physical - the actual location of the documents, computer records, microfiches, photographs and other materials. In this respect, offices with records housed in more than one building not only face a higher cost in protecting the records but also in persuading the public that what they do is consistent with the security of individual records. The second perimeter is virtual - the electronic perimeter traced by the agency's internal communications system, which is presumably connected to its stores of individual data.

539. Both perimeters need to be protected from malicious or unauthorized intrusion. As mentioned above, in many offices, access is strictly controlled (e.g., employees have to wear identification tags), and additional security measures are taken where the most sensitive records are housed.

540. The electronic perimeter must also be defended with password clearances and other types of security, but this is more difficult to accomplish, particularly as technology continues to demand upgraded defences. The popular media has given great publicity to successful attempts by hackers to penetrate even the most secure communications systems and gain access to the most sensitive information, in some cases involving national security files. Since the public is more apprehensive about computer security than that of conventional documents, and since this trend is likely to increase, a few statistical agencies have chosen to completely isolate the internal system of communications from the outside world, including from its own local offices.

541. In such agencies no direct link exists between the stores of data and communications to the outside world. Furthermore, whatever confidential data are

transmitted from local offices to the central office are protected by the best encryption system available. Whether or not this solution will provide sufficient protection in the face of rapidly evolving technologies cannot be foretold. One of the agency's systems analysts should be placed permanently in charge of data security, with responsibility for such matters as encryption; removal of identifying attributes; and protection against malicious tampering with data files. In addition, someone should be assigned responsibility for the physical security of the premises and data holdings.

2. Sanctions

542. Some official regulation must outline the sanctions imposed for security breaches, as a deterrent as well as to demonstrate to the public that the matters of confidentiality and integrity of data holdings are taken seriously. The most natural way of dealing with this is through the law itself. An ideal situation is one in which legislation clearly lays out the various types of infractions. Examples of these, in increasing order of severity, include the following:

- Carelessness**.** An interviewer leaves an envelope containing identifiable completed questionnaires on the bus, where it could be discovered by a third party and picked up by the press as an example of public sector laxity;

- Improper behaviour**.** A subject-matter expert comments in public about the income declared by a particular family when interviewed in the course of a household income and expenditure survey (the possible consequences are the same as above);

- Behaviour with malicious intent**.** An employee wishing to embarrass the management of the statistical agency sends the prime minister's census form to the press;

- Use of confidential information for personal profit**.** An employee offers to provide confidential information on a company to one of its competitors.

543. These, of course, are merely illustrations; the law can be more comprehensive about what constitutes a security crime. However, they illustrate the kinds of infractions that should be listed and for which there ought to be credible penalties.

C. Confidentiality and disclosure

1. Principles

544. The preceding sections have presented the precautions that are to be taken with incoming information, ways to prevent it from falling into the wrong hands and the disciplinary measures that should be taken when there is a breach of security. This section explores how the contract established with respondents must be honoured. It considers ways to prevent publication of aggregate data that would disclose information

revealing the identity of an individual, business or institution?[118] The risk of doing so increases with the publication's degree of detail. For example, it is customary to publish the results of a family expenditure survey by region, demographic characteristics of the respondents and income. Even in a survey incorporating tens of thousands of responses, the identities of respondents whose statistical data fall into extreme ranges (e.g., persons who are affluent and well known) might be easily identifiable.

545. The standard solution is to conduct initial research determining those cells in which it will be possible to identify respondents. In subsequent research, the minimum degree of aggregation required to suppress[119] respondents' identity is decided. Usually, this kind of research takes place at two levels. First, there is a purely practical, rudimentary attempt to eliminate the offending cases at minimum cost, carried out by trial and error. For those agencies that can afford it, there is also a more theoretical approach to identifying the offending cells and showing that the solution proposed is the least costly in terms of information suppressed. Since few statistical agencies are equipped to conduct such research, an alternative solution is to commission interested academics to devise minimum data suppression models. Of course, research results can be applied elsewhere, and smaller agencies are encouraged to maintain communication with large statistical centres, particularly on this subject.

2. Dominance and residual disclosure

546. A number of issues relating to confidentiality have been the objects of scrutiny over the past fifty years, and thus have become more prevalent with the diffusion of computer technology and the resultant expansion of the community of users. The increased capacity to store, cross-tabulate and publish statistical data has inadvertently created more opportunities for accidental disclosure. Moreover, users' ability to manipulate data has also increased and with it their ability to uncover more information than is meant for them.[120]

547. For each of its tabulations, the statistical agency must track how many respondents are included in each cell and how many are in each of the cells that can be calculated as a residual. Measures must be taken to disguise all those that have less than the threshold (three is generally the cut-off point). In those cells with a number of respondents equal or greater than the threshold, but with one or two respondents accounting for more than 90 per cent of the value of the displayed variable, measures

[118] There is an interesting question as to whether these are the only categories that must be protected. If we disclosed the average income of cardiologists, for instance, or the crime rate by ethnicity of the perpetrator, would that be a breach of contract? The matter of exactly what information ought to be protected has not yet been the object of careful examination.

[119] The term "suppress" is used to suggest, not that the basic information should be excluded, but rather that it should be kept anonymous. The solution is to combine cells.

[120] Suppose a statistical agency runs a tabulation service on demand, the object of which is to provide, in machine-readable format, tables derived from a complex survey. The general principle is that whatever is requested by one user is available to all. Let us also suppose that there are **n** users, and that each wants different cross-tabulations. It follows that a statistical agency would have to track the impact of the **n**th demand on the number of possible residual disclosures, assuming the **n**th user had access to the previous **n-1** requests.

should be taken to disguise their activity as well.

548. For offices with a significant publishing programme, such the work involved in tracking dominance and residual cases[121] justifies the creation of a special unit in charge of confidentiality research. Such a unit is usually a branch of (or is at least closely tied to) the department in charge of applications of mathematical statistics.[122]

3. Confidentiality and household statistics

549. While one person or household may be much like another, such large corporations as Sony, the Royal Dutch Shell Group, International Business Machines Corporation (IBM), the De Beers Group, Gasprom or General Motors are quite unlike small businesses. For this reason, a number of differences between social and economic statistics must be taken into account. In social statistics, the notion of dominance is relevant only in some surveys of household assets, income and expenditure. Furthermore, in the domain of economic statistics, any survey sample will typically include a certain class of units because they make up a large proportion of the variable, but families and individuals need only selection with non-zero probability. Lastly, with the exception of special surveys (e.g., the financial balance sheet of households), random perturbations can be introduced in the results of household surveys to mask individual cases without affecting the estimates of any of the moments of distribution. However, this is rarely the case with business surveys.

550. Recently, longitudinal surveys have raised a number of interesting issues connected with confidentiality research. These surveys, in which a selected unit (generally an individual or family) is tracked over time so that changes in its attributes can be observed, are becoming popular among some of the more advanced statistical agencies. The temporal aspect of this type of survey greatly increases the likelihood that the sample will be identifiable. However, minimum necessary suppression may render useless the delicate data analysis and presentation that such surveys demand. Clearly, research into the confidentiality of household statistics is vastly different from research on businesses, and should be treated as a separate item on the research agenda.

4. Disclosure with consent

551. In terms of public image, it is in the best interest of the statistical agency to position itself as the guardian of response burden, continually demonstrating its desire to minimize paperwork when possible. Accordingly, the statistical agency should aggressively pursue opportunities to use the regulatory powers of other parties to gather information, rather than duplicate the queries administered by other government offices. However, eliminating duplication may imply the sharing of information, which can work in two ways: the statistical agency obtains its information from the data that is collected

[121] For an overview of a tracking and suppression system, see Gordon Sande, "Automated cell suppression to preserve confidentiality of business statistics", in *Proceedings of the Second International Workshop on Statistical Database Management*, Los Altos California, 27-29 September 1983, Roy Hammond and John L. McCarthy, eds. (Lawrence Berkeley Laboratory, 1983).

[122] Such a department will normally be responsible for encryption research, among other tasks.

by the other government body, or vice-versa. When given a choice, respondents usually prefer to report to the statistical agency for a variety of reasons, the most important being its reputation for discretion and attention to detail. If it is possible that information will be shared, respondents must give consent (preferably in writing) to have their forms reviewed by a third party, and be made to understand why such sharing might occur. Respondents should also be made aware that while their refusal is absolute, the result might be that they have to provide the same information twice.

552. There are other circumstances in which written consent for disclosure may be sought. In highly industrialized countries, the analytical tables for industry (the input-output tables, for example) can lose their value if disclosure rules are interpreted to the letter. Before accepting a sacrifice in the analytical value of its compilations, the statistical agency may wish to seek consent from the businesses that run the risk of having their operations disclosed. These businesses would be asked to agree to a form of publication that affords imperfect disclosure protection. Of course, taking this step on any significant scale might jeopardize the statistical agency's reputation.

5. Forced disclosure

553. In certain instances, there is no feasible choice but to disclose the activities of a particular business enterprise. For the most part, this occurs with public utilities, State monopolies and industries that are dominated by a single firm, as is often the case in smaller countries. When the situation is well known and in fact precedes the creation of the statistical agency, a special clause dealing with it is written into the law. In annual statistics this is less of an issue, because State enterprises and large public monopolies are usually requested to report their activities and financial circumstances in great detail, and these reports are made public.

6. Passive confidentiality

554. It is not always possible to prevent all breaches of confidentiality. The most obvious example is that of international trade statistics. In this case, the data collected is pertinent to recorded transactions rather than the individuals or businesses responsible for them. As these statistics constitute a census of transactions over the period of reference, are published in extremely fine detail and include several attributes (what is exchanged, its destination, means of transport, points of entry and departure, and so on) it is not realistic to assume that total confidentiality can be maintained without having serious effects on the timeliness and utility of the information provided. The customary approach is to opt for a mixture of active and passive disclosure protection. For all cases where it is known that there is only one importer or exporter,[123] or a case of dominance, suppression measures are devised and implemented in advance. For all other cases, suppression measures are taken only if a business or individual takes the initiative and complains to the statistical agency. If the agency is accused of failing to act on the matter in a timely

[123] This approach works asymmetrically. In most countries there are many more importers than exporters. The latter can be more easily identified ahead of time, and pre-emptive measures can be taken to suppress identifiable information. However, any business can be deemed an importer, and therefore a priori decisions are more difficult to make.

fashion, the respondent should be made aware that, unlike situations in which the statistical agency designs and administers its own surveys with the means to detect disclosure a priori, international trade statistics employ records compiled through an external administrative process and are not the direct responsibility of the agency. Therefore, it can only act when notified.

7. Arrangements for research

555. Under certain conditions and by their own decree, statistical agencies may order the publication of the names, addresses, industrial activity (expressed in the form of an industrial code) and size (expressed by a code denoting an employment size class or some other agreed variable, also by size class) of selected business respondents. Generally, such an exception is made for research purposes.

556. Increasingly, there are perfectly legitimate applications that not only employ statistical aggregates, but also depend on individual data for their success. For example, the study of complicated interactions between consumer income, expenditure, household savings and taxation through the examination of published statistics tends to suffer greatly from aggregation bias. The testing of formal hypotheses can be more readily conducted on the basis of individual longitudinal records. Few offices are equipped to conduct this kind of research themselves. It follows that in this situation, the common interest is best served if the statistical agency provides researchers with the information they need and allows them to formulate hypotheses and conduct the necessary research effort.

557. Today, circumstances exist in which a statistical agency cannot avoid providing microdata. However, an accessible database with individual information should be constructed in such a way that the probabilities of identification with certainty are extremely low or non-existent. Of course, all records must be made anonymous, and in addition, there may be grounds for introducing random perturbations so long as the moments of the original distribution are preserved.[124] Even after all these precautions, it is necessary to have an officer oversee access to the database and the work derived from it. This means that at all times an agency official knows who is using the database and why, what they are looking for, and what results, if any, have been achieved so far. The agency must coordinate with users whose work is dependent upon particularly detailed sets of data, so that the general public is confident that such detail is provided only in support of meaningful advances in the social sciences.

558. It is difficult to envisage a publicly accessible database for business information, because each industry generally has an establishment or company big enough to dominate the results and be recognized by the extent to which it affects the data. However, with small businesses the situation may be different and, if there are questions relating exclusively to small business, the assembly of a publicly accessible database is justified.

[124] The subject of how to provide data that is non-identifiable but contains the same information as the original set is under active research. See, for example D.B. Rubin, "Multiple imputation after 18+years", *Journal of the American Statistical Association*, volume 91, pp. 473-489.

Note that while numerous examples of successful public-use databases exist for households, few have been constructed for businesses.

559. The research community places special demands on data that cannot be ignored, particularly when those demands are deemed to serve the public interest. Many of these legitimately require access to individual records or to cells in tabulations where the number of respondents is less than the threshold. Since such research is not carried out for personal gain or with the intent to secure commercial intelligence that might hurt established interests, it is difficult for the statistical agency to disallow it. Conversely, it is just as difficult for the agency to make exceptions for certain classes of users despite all the assurances given to respondents. Rarely does a statistical agency achieve an effective balance between the two pressures. The following section describes a solution that has worked in the context of a few agencies.

560. One method that has met with limited success in several countries is the formal recruitment of academic researchers who require access to microdata, even though their remuneration is merely symbolic. This procedure ensures that the researcher agrees to abide by the rules of the statistical agency and understands that failure to do so will result in official sanction. The following conditions would warrant this type of formal offer:

- The researcher has proper credentials;
- The research plan concerns matters that the statistical agency considers to be in the public interest or for the advancement of social science, and no alternative source of data is available;
- The researcher agrees to uphold the protocol and posture of the statistical agency for the duration of his/her employment.

561. Prior to making an offer of employment the statistical agency should ensure that the researcher meets the three conditions stipulated above, and may wish to consult its legal adviser. It is also important not to use this practice to circumvent the rules of confidentiality. Clearly, overuse, or use for projects that would appear trivial, could compromise the agency's credibility. A related issue, concerning control of the intellectual property resulting from the researcher's work, is beyond the scope of this handbook.

562. Another method, tested at the United States Bureau of the Census, consists in creating a "sterile chamber" where researchers can work. While they do not acquire the status of employees, researchers act in full knowledge of the statistical agency's confidentiality policies and accept the statutory sanctions should these be broken.

Conclusions

Statistical agencies are committed to safeguarding information that plainly reveals the operations, belongings, attitudes or any other characteristics of individual respondents. The principles behind this practice are fairly clear, but specific applications require ongoing research in order to certify that there is no breach of confidentiality. Theoretical research, while invaluable in that it can be shared and used repeatedly, is costly and is conducted in only a few agencies. However, even when such research is not possible, the more practical work of suppressing revealing information without compromising the integrity of statistical aggregates should go on at all times. These activities should be carried out openly as a means of reassuring the public.

Exceptions to the rule of absolute confidentiality do exist; when these occur, the statistical agency should be candid with the public and explicit about the conditions that warrant exceptional treatment. Household and business records are not treated in the same way, since the former are easier to disguise, whereas circumstances often arise in which the latter cannot be disguised without deleterious effects on the quality of published data. In such cases (e.g., in the case of natural monopolies or dominant enterprises), either a specific provision for disclosure is included in the statistical legislation, or else the statistical agency should not act without the explicit consent of the potentially identifiable party.