UNITED NATIONS

E

Economic and Social Council

Distr.
GENERAL

CES/2004/WP.1
13 May 2004

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-second plenary session
(Paris, 8-10 June 2004)

TASK FORCE ON CONFIDENTIALITY AND MICRODATA – DISCUSSION PAPER

Paper submitted by Dennis Trewin, Chairman, CES Task Force on Confidentiality and Microdata

BACKGROUND

1. So far confidentiality protection has been mainly a national issue. However, in the context of EU and increasing data dissemination over Internet, it is becoming also an international issue. There is a lot of international collaboration among the research community, and the researchers can be very critical towards different access rules in different countries. Often researchers are not allowed to access other countries microdata because of the fear that confidentiality protection cannot be guaranteed but cross-country comparisons can be a very important part of a research project. International agencies are among those who want to use microdata for research purposes. This raises the question of whether it is possible to internationally agree on some common principles for dissemination of microdata. It should also be seen in the context that the 2003 Conference of European Statisticians (CES) agreed that support for research as an important activity of the National Statistical Offices (NSOs). This includes microdata which is the main focus of this report. (Although the reference is to NSOs in this paper, in many countries, particularly those with decentralised systems, there are several statistical producers. The reference to NSOs should be read as incorporating all producers of official statistics.)

2. The planned outputs of this activity are:

- Agreed international principles on the provision of access to microdata, together with explanatory documentation; and

- Case studies of good practice, consistent with the Principles.

These two outputs will be developed following feedback from this Discussion Paper.

3. This discussion paper recognises that the precise arrangements for access to microdata will vary from country to country. They will vary depending on matters such as legislation, public attitudes and the capacity to support the research community. For example, you would expect the arrangements for a well-developed statistical office will be quite different to those in a less well-developed statistical office. It should not be anticipated that each country will come up with precisely the same arrangement.

4. We should also be mindful that not all countries are coming from the same position. Some countries, particularly from Eastern Europe, have not had strong legislation supporting confidentiality. This is being changed in many cases but the cultural change to support the legislation change can take longer.

5. The Terms of Reference for this study are shown at Attachment A. This report does not deal with the security of electronic data transmission which can be a concern in its own right.

6. At the end of the paper, there is a listing of the issues on which we would like feedback. However, we also welcome comments on any other issues relevant to our discussion.

A FRAMEWORK TO SUPPORT DISCUSSION

7. There are various ways a NSO can support research work. These are summarised below. The terminology in the first column is used in the rest of this paper. It is clear from comments received to date that these terms are interpreted in different ways in different countries. It would reduce confusion if there was some agreement on the terms that should be used. The Bureau of the CES agreed on the use of the terms in the following table.

A. Statistical Products for Use Outside the National Statistical Office

Dissemination Stream	**Notes**
Statistical Tables	This can include both standard tables and special tables (or special analyses for that matter) generated at the request of the researcher. Some offices now release very detailed matrices, known as data cubes, which researchers can manipulate to support their own needs. However, if very detailed, the level of confidentiality risk can be similar to microdata.
Anonymised Microdata Files (AMFs)	These are microdata files that are disseminated for use outside the NSO.
- public use files	They have been anonymised and are often released on a medium such as CD ROM

- licensed files (Note: The term anonymised implies that, not only names and addresses removed, but other steps taken to ensure that identification of individuals is highly unlikely.)	Licensed files are distinct from public use files in that use may be restricted to certain researchers and a legal undertaking is signed before files are provided to the researcher. Public Use Files are more generally available. The level of confidentiality protection in Public Use Files should be such that licensing is not necessary. AMFs have been a common way of providing access to researchers in many countries.

B. A Service Window through which researchers can submit data requests

Dissemination Stream	**Notes**
Remote Access Facilities (RAF) (Different type of RAF facilities exist which are discussed in more detail in Section 8.)	Arrangements are now being made in many countries that allow researchers the ability to produce statistical outputs from microdata files through computer networks and without them actually "seeing" the microdata. Because of the additional controls that are available through RAF, and that microdata does not actually leave the NSO, access to more detailed microdata can be provided this way.

C. Arrangements for allowing researchers to work on the premises of the National Statistical Office

Dissemination Stream	**Notes**
Data Laboratories (DL)	On-site access to more identifiable microdata, typically with stringent audit trails and NSO supervision. The access to more detailed data creates some inconvenience to the researcher, because of the requirement of working at the NSO, or at an NSO enclave. The term microdata is used throughout the paper. It can refer to data about an individual person, household, business or other entity. It may be data directly collected by the NSO or obtained through other sources e.g. administrative systems.

OUTLINE OF DISCUSSION PAPER

8. While both NSOs and researchers have a common interest in the dissemination and analysis of high quality data, there are also fundamental differences in the principal motivation of the two groups and very different consequences from a breach of confidentiality. The paper seeks to provide an overview of these different perspectives, and find key elements of an optimal solution that takes account of the tensions.

9. The focus will be on access to anonymised (i.e. names and addresses have been removed) microdata from persons and households. This covers the last three, but not the first, of the four Dissemination Streams mentioned in the previous Section. Data from businesses are discussed separately in section 11.

10. The next two Sections will outline the perspectives of (a) the NSOs and (b) the research community. The following Section tries to bring these perspectives together in the spirit of finding the key elements of an acceptable solution. In this context, the role of the three dissemination streams of interest is discussed. This leads to an outline of the key principles for guiding individual country implementation of arrangements for providing microdata access to the research community.

11. There are a number of special issues that should be discussed. These are discussed particularly the linking of different data sets to support statistical and research uses. The final version of this paper will include some examples of common best practice but they are not included in this version.

THE PERSPECTIVE OF THE NATIONAL STATISTICAL OFFICE

12. NSOs must maintain the trust of respondents. Confidentiality protection is a key element of that trust. If respondents believe or perceive that NSO will not protect the confidentiality of their data, they are less likely to cooperate or provide accurate data. One incident, particularly if it receives strong media attention, could have a significant impact on respondent cooperation and therefore the quality of official statistics.

13. This is the dominant issue from the point of view of NSOs but there are other concerns. A key one is whether there is sufficient authority to support researcher access to microdata, either through a legal mandate or some other form of authorisation.

14. Some NSOs are concerned that the quality of their microdata may not be good enough for further dissemination. Whilst, quality may be sufficiently accurate to support aggregate statistics, this may not be the case for very detailed analysis. In some cases, adjustments are made to aggregate statistics at the output editing stage without amendment to the microdata.

Consequently, there may be inconsistencies between research results based on microdata and published aggregate data.

15. NSOs may also be concerned about the costs. These include not only the costs of creating and documenting microdata files, but the costs of creating access tools, safeguards and supporting and authorising enquiries (e.g. helping new users navigate around complex file structures and variable definitions) made by the research community who are analysing these data files. Although the costs are borne by the NSOs, they are often not provided with budget supplementation to do the additional work. (Also, researchers generally do not have the funding to contribute to these costs.)

16. On the other hand, NSOs are increasingly recognising the importance of supporting the research community - the additional value that is provided to NSO data collection and processing effort through effective use of its data for research. Specifically, it is in the public interest that insights, which can be provided from the data, can be made available to decision makers and the public. Furthermore, if data is used more extensively in this way, it can provide an extra level of protection against cuts to these statistical programs. Nevertheless, NSOs are the custodians of data which has been trusted to them and they are responsible for the legality, consistency and transparency of practice.

THE PERSPECTIVE OF THE RESEARCH COMMUNITY

17. The research community is not those who belong to universities. It may include researchers in other research institutes, government, business, NGOs and international organisations for example. The term is used in a generic sense.

18. From the perspective of the research community, supporting research based on microdata should be an important component of any official statistical system. Julia Lane highlighted many of the benefits in her presentation to the 2003 Conference of European Statisticians.

(i) Microdata permits policy makers to pose and analyse complex questions. In economics, for example, analysis of aggregate statistics does not give a sufficiently accurate view of the functioning of the economy to allow analysis of the components of productivity growth.

(ii) Access to microdata permits analysts to calculate marginal, not just average effects. For example, microdata enables analysts to do multivariate regressions whereby the marginal impact of specific variables can be isolated.

(iii) Broadly, widely available access to microdata enables replication of important research.

(iv) Access to microdata for research purposes, and the resulting feedback, can facilitate improvements in data quality. The US Bureau of the Census has actually formalised the documentation it requires from researchers to assist it to improve the quality of its surveys.

(v) It increases the products derived from statistical collections and hence the overall value for money obtained from these collections.

19. Not mentioned by Julia, but also important, is that lack of access to microdata may result in researchers developing and conducting their own statistical collections adding to the reporting burden imposed on the community. As well as the cost involved (to the funder as well as the respondents), the collections will usually be of inferior quality and with smaller samples than official surveys. There are benefits from having an accepted and authoritative, as well as high quality, data source for all analysis.

20. The researchers point out that they are not interested in identifying individuals and the evidence is that this is indeed the case. Given this they feel that NSOs have generally been too conservative in the access they provide to microdata.

21. At a 2003 Workshop on Confidentiality Research hosted by the National Science Foundation, Peter Madsen referred to the Privacy Paradox. He argues that "the rush to ensure complete levels of privacy in the research context paradoxically results in less social benefit, rather than in more". He argues that when you include the concept of utility you may get different outcomes.

> "Perhaps through this additional concept of utility, people will recognise that while they surely have the right to privacy, they may also come to the realisation that they have a duty to share information, if the common good is to be furthered."

22. The research community also sees the importance of research into improved methods of confidentiality protection which increase the usefulness of the underlying data. NSOs would agree with the importance of this research. However, our view is that this research is only likely to lead to a partial answer to the desire for improved access to microdata for research purposes and that researchers would remain frustrated if we relied solely on improved statistical methods for confidentiality protection.

HOW MIGHT BE THE TENSION BETWEEN THE NSO AND RESEARCHER PERSPECTIVES BE RESOLVED?

23. This will most effectively be done by NSOs moving from a Risk Avoidance to a Risk Management strategy. How to do this is discussed in more detail in the following paragraphs. But first, some discussion of the risks of identification with microdata might be useful.

24. The rapid expansion of databases, containing data about identifiable persons, means that it is virtually impossible to completely avoid identification, particularly if household structure is contained on the files. Many of these data bases are held by the private sector where controls on their use are generally less than for the public sector. Furthermore, technology advancements have made data matching easier, whether by exact matching or statistical matching techniques (which can lead to exact matches in unique cases). Risk avoidance in essence means not allowing identifiable microdata to leave the premises of the NSO. (Please note that risks will vary according to the size of country among other things. In smaller countries, the risk will be relatively higher because there are more unique cases. For example, in Australia about 25% of households are unique in the size, age, sex structure of the household even when you combine age into five yearly groupings. These households are potentially identifiable through a statistical matching exercise.)

25. Nevertheless, the microdata access provided by NSOs does not seem to have been an area of public controversy. Implicitly, there seems to be a reasonably high level of public acceptability of current practices although we are not aware of countries where there has been a public debate (possibly because access has been managed carefully by NSOs). But general community concerns about privacy suggests there is a limit to what the public is likely to accept. A debate could be easily triggered (across national boundaries) by one untoward incident.

26. It does raise the question of whether NSOs should be more up front about all the uses of their data collections. That is, we are transparent in outlining that one of the valued uses of the data from certain collections will be to provide researcher access to confidentialised microdata under controlled conditions for specific purposes.

27. How do NSOs manage the risks? Some suggestions are outlined below.

i) Agree on a set of principles (such as those outlined later in the paper) which must be followed in the provision of access to microdata.

ii) Ensure there is a sound legal and ethical base (as well as the technical and methodological tools) for protecting confidentiality through microdata access. This legal and ethical base requires a balanced assessment between the public goods of confidentiality protection on the one hand, and public benefits from research on the other. This will depend to a large extent on the merits of the research proposal and the credibility of the researcher.

iii) To have an arms length process for the balancing of the public good which might be derived from access to confidential data versus the risk of confidentiality violation. Ethics Committees may be able to assist in situations where there is discretion in deciding whether to provide access or not but regardless NSOs must conform with the legislation or other protocols that operate in their country.

iv) Be completely transparent about the specific uses of microdata to avoid suspicions of misuse.

v) Provide more access through remote access facilities and data laboratories as completely unidentifiable microdata for public release may no longer be possible without considerable "distortion" of the data. Explore other opportunities to use technologies to improve access to microdata in ways that adequate confidentiality protection is provided.

vi) Put some of the onus of responsibility to the research community. Ensure researchers are aware of the consequences to them and their institution if there are breaches. Follow through on retribution if there are breaches. Access should be regarded as a privilege not a right.

28. The last point requires some comment. The culture and value system of the research community is very different to that which operates in an NSO. They often regard some of the "controls" inherent in the microdata access arrangements as unnecessary bureaucracy. Whilst we are not aware of researchers using their access to microdata to identify individuals, we are aware of microdata provided to them on an exclusive basis being provided to other researchers

without permission, or of cases where microdata has been statistically matched without permission with other data to produce richer data sets. The researchers in question may feel they have done nothing wrong as they have not tried to identify individuals - this may be due to the different culture under which they are operating and different views of risks from incidents.

29. How can NSOs put some of the risk back on to researchers? This might include:

i) Asking them to prove their bona fides as a researcher. Demonstrating the public benefits of their research.

ii) Signing a legally binding undertaking with similar penalties to those operating for NSO staff if they breach confidentiality provisions.

iii) Ensuring researchers are fully aware of their obligations through appropriate training. Follow-up with effective audit and monitoring procedures.

iv) Where offences occur, withdrawing all current and future services from the researcher and possibly their institution for a period of time (possibly until they have undertaken disciplinary action against the offender). Undertaking legal action where appropriate.

30. The reality is that a combination of legal, administrative and technical measures will be necessary to ensure public confidence in the arrangements. Furthermore, the research community must accept that they have no automatic right of access. The NSOs may be enabled to provide access but researcher access must be at the discretion of the NSO. Also, researchers will have to accept that they will have a responsibility to maintain and uphold the conditions under which they have been provided access. The limitations and safeguards may be more restrictive than exist with other data sets to which they have access but they still must be followed.

31. It is important that NSOs do some contingency planning in the event the microdata access does become an issue for public debate. They should not assume that it will not happen. What are some of the key defences?

(a) NSOs can point to the care they take in providing confidentiality protection through devices such as anonymising the microdata, providing strong physical security protection and our care in devising a process for the assessment of the balance between the conflicting public goods of confidentiality protection and the public benefits of research.

(b) If an offence has occurred and NSOs are questioned, NSOs should be open about the offences and the penalties that have been invoked; they should make it clear that the breach is the responsibility of the researcher.

(c) It should be possible to point to the overall public benefits of providing microdata access, particularly for the situation where the offence has occurred.

(d) Well known people who are prepared to publicly support the arrangements should have been arranged. Senior privacy officials are of particular importance in this regard.

ANONYMISED MICRODATA FILES

32. This is seen as a very valuable service by researchers. However, in light of the increased possibilities for data matching, the trend might be to reduce the amount of data available in AMFs and to put more reliance on facilities such as RAFs and data laboratories for researcher access. The alternative is to put increased reliance on researchers honouring undertakings that they make.

33. Although NSOs generally provide equality of access to all users of their statistics, this may not be appropriate for AMFs. A different attitude may be taken to users who do not have strong bona fide research credentials or if they have access to data bases where it would be easy to match AMFs. Ethics Committees may be able to assist NSO heads make decisions where discretion has to be exercised.

34. The exception is Public Use Files where access is deliberately intended to be broader. A question for debate is whether legally binding undertakings should be part of the arrangements for providing access, even for public use data files, unless public use files are clearly unidentifiable even if statistically matched with other data sets. (Except for the largest countries, this may be very difficult to achieve for files that contain household structure given the relative ease of matching with external data bases to identify unique cases.)

35. Users have raised the importance of Public Use Files. They are greatly appreciated in those countries where they exist and they are used extensively for research and teaching purposes. Yet it may not be difficult for someone who is so inclined to publicly identify some individuals through statistical matching with other data bases. We are not advocating the discontinuation of Public Use Files but a close examination of the conditions under which they are released to better manage the risks of a confidentiality violation. For example, a legally enforceable undertaking may be one of the requirements of access.

REMOTE ACCESS FACILITIES

36. These are important but the way Remote Access Facilities are implemented varies from country to country. They are characterised by the fact that researchers do not have access to the microdata itself. Often there is a contractual arrangement between the NSO and the researcher, and often with the institution of the researcher. By way of illustration, Statistics Canada provides researchers with dummy microdata files and allows researchers to submit runs via computer networks. Statistics Canada runs them off line and sends the results back via computer networks after checking for confidentiality. Although similar arrangements exist at the Australian Bureau of Statistics, there are some important differences. The microdata files are confidentialised before becoming accessible through the RAF to prevent spontaneous identification. However, trial runs are permitted against the RAF files and small numbers of unidentifiable unit records are allowed to be downloaded to explore outliers and the like. Output is checked before being sent to the researcher. It operates in batch mode but an interactive version is being developed. The arrangements in Statistics Denmark are different. It is an on-line system where researchers can run analyses against the full anonymised microdata file. Arrangements are such that they cannot download the microdata itself. To further manage risks, they put greater reliance on the undertakings made by institutions and the retribution (particularly denial of future access) if there are breaches of the rules.

37. In reality, there are two types of Remote Access Facilities.

(a) Remote execution where a researcher submits a program and receives the output later by email.

(b) Remote facilities where the researcher performs the analysis and can immediately see the answer on the screen.

Many countries have facilities along the lines of (a) but, apart from the Danish system, facilities along the lines of (b) are still under development.

38. Although only available so far in a few countries, and the models and approaches are somewhat different as illustrated above, the experience to date has generally been positive. There would be value in sharing experiences with each other and with countries who are contemplating similar developments.

39. From the cost perspective, RAFs are preferable to Data Laboratories as the supervised access in RAF is less expensive than the supervised use involved in Data Laboratories.

40. If these facilities do not remove identification risk entirely, there should still be some undertaking made by researchers to ensure they are fully aware of their obligations. Education is also important, together with regular monitoring and checking of the use of these facilities.

DATA LABORATORIES

41. They have been in use for many years in some NSOs and have been effective in controlling identification risk particularly for data sets where a confidentialised microdata file is not possible. They still require conditions of access to provide an adequate level of protection. The main criticism of DLs has been the lack of convenience to the researcher, including sometimes being forced to use unfamiliar data analysis software. They are also expensive for the NSO to manage compared with other options.

42. Some NSOs have established new premises for data laboratories in locations that are more convenient to researchers, (sometimes known as Research Data Centres) but this can be an expensive option.

43. What are some of the key conditions of access to Data Laboratories? These might include (a) documentation of the public good that the research will provide, (b) outlining how the results will be accessible to the public, (c) evidence of the bona fides of researchers, (d) a legally binding undertaking, and (e) requirements for supervision by NSO staff.

ENGAGING AS TEMPORARY NSO STAFF MEMBER

44. Another way that researchers may access microdata is through engaging as a temporary NSO staff member. This should really be seen as a work program issue rather than a data access solution. It should not be done unless the researcher is assisting with the work of the NSO - otherwise it could be seen as a sham. If this type of pretence was occurring and became

public, confidence in the NSO would diminish.

45. The involvement of the researcher may be at the initiation of the NSO - they are seen as someone who can bring special skills to the work of the NSO. On the other hand the proposal may come at the initiation of the researcher. But the NSO must accept the merit of the proposal and decide to incorporate in its work program activities.

BUSINESS DATA

46. Another special issue is Business Data including agricultural businesses. This is more easily identifiable than household or personal data, especially on a spontaneous basis, particularly for large businesses, because the distribution of their characteristics is much more skewed. In many countries, data bases of business data are often more accessible thereby enabling matching.

47. From the point of view of researcher access, the main differences between household/personal data and business data are the dissemination streams that are relevant.

- Statistical Tables remain relevant although the higher level of confidentiality risk means that less detailed data will generally not be available.

- Anonymised Microdata Files will only be relevant for the smallest businesses where that may be a group of particular interest for researchers. Even then there will need to be "distortion" of some data (e.g. financial data) to avoid matching with other data bases (e.g. taxation data). Reliance on undertakings would be extremely risky.

- For similar reasons, Remote Access Facilities may only be relevant for microdata files of the smallest businesses. At least, use of these facilities will enable NSOs to control the matching risk so it may not be necessary to "distort" the data to protect confidentiality.

- Data Laboratory arrangements are likely to be most pertinent for access to microdata files covering businesses of varying sizes. Such arrangements exist in Statistics Netherlands for example.

SPECIAL ISSUES

48. One special issue is linked data sets, whether using either exact or statistical matching techniques (although exact matching will generally be more sensitive). Sometimes at least one of the linked data sets will come from outside the NSO. Data linking can add considerable value to data sets for analysis purposes but it does increase identification risk. Also there is public suspicion about data linking. As a result special arrangements are often used in privacy legislation to cover the linking of data bases. The issue of data linking is dealt with in more detail in Section 14.

49. NSOs often have a principle of "equality of access" underlying their dissemination activities. This has the support of the general public. Should the same principle apply to microdata access? It could be expected that the general public would be more positive about providing access to those who will be undertaking research for the public good than those who are using it for commercial purposes. There are credible arguments for providing some restrictions on access to microdata and treating it differently to aggregate statistical data. The exception to this would be microdata files that are deliberately designed as Public Use Files.

50. Privacy principles often argue for informed consent as a condition of use of individual data. For statistical collections which are generally quite large it would be impractical to manage the documentation involved in having specific consent from each subject. However, transparency is important and even though informed consent is not being sought NSOs should be open about how the data they collect from individuals is being used. If NSOs are not transparent, "discovery" may become a big story even when practices are defensible.

51. The last special issue dealt with is access by researchers from other countries. This is discussed in Section 15.

PRINCIPLES

52. The UN Fundamental Principles are very clear on statistical confidentiality.

> "Individual data collected by statistical agencies for statistical compilation, whether or not they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes."

Any principles for microdata access must be consistent with this Fundamental Principle.

53. The following principles are proposed to start debate. Each are discussed in the following paragraphs.

Principle 1: It is appropriate for microdata collected for official statistical purposes to be used for secondary data analysis to support research as long as there are prescribed conditions that protect confidentiality.

Principle 2: There should be a legal or other arrangement to support use of microdata in order to increase public confidence that microdata will be used appropriately. Provision of microdata should then be consistent with these legal and other arrangements.

Principle 3: Microdata should only be made available for research or statistical purposes.

Principle 4: The processes for researcher access to microdata as well as the uses and users of microdata (except for public use files) should be transparent, and publicly available, again to increase public confidence that microdata is being used appropriately.

54. Principle 1 should be seen as an enabling provision. The NSO should have the discretion as to whether the microdata should be provided or not. For example, there may be quality concerns which make it inappropriate to provide access to microdata. Or there may be specific persons or institutions to whom it would be inappropriate to provide microdata. By design, such a limitation would not apply to Public Use Files.

55. The arrangements under Principle 2 should be cleared with the privacy authorities of countries where they exist. There may also be NGOs who have a "watchdog" role on privacy matters. If possible, it would be sensible to get their support for any arrangements or at least address any serious concerns they might have.

56. For Principle 3, it may make sense to apply a "compatibility" test. If the use of the microdata is incompatible with the original data collection, then microdata access should not be provided. Also, some requests for data may be legal (e.g. a court order) but inconsistent with this principle. It is in the long-term interest of public confidence in the official statistical system that these requests are refused.

57. The NSO web site is an effective way of complying with Principle 4. This requirement would not apply to Public Use Files.

DATA LINKING

58. The linking of data sets, whether by exact or statistical matching, can add considerable value to data sets. It can facilitate a much greater range of analyses. Health research, in particular, is an area where linked data sets can be of particular value. It is an appropriate function for NSOs to be involved in the linking of data sets for statistical purposes.

59. Increasingly, researchers are looking to link data sets with the data sets of the national statistical office or other statistical agencies (including the population census in some countries). The statistical agency has to be the custodian for these linked data sets. There may also be situations where it is the preferred custodian of linked data sets even when they come from outside the statistical agencies, because of the safeguards and public trust that already exist.

60. While, there are clear benefits in data linking, there are also risks. Identification risks are increased. Perceptions are also important. Studies in many countries show much public concern about linking data bases. It is particularly important that the four Principles outlined in the previous Section are followed for linked data sets.

61. For those countries with Privacy Commissions, the arrangements for data linking should

be supported by the Privacy Commission.

ACCESS BY INTERNATIONAL RESEARCHERS AND INTERNATIONAL AGENCIES

62. Cross country comparisons are important for understanding the effectiveness of policies and programs, for example. The benefits of access by international researchers and international agencies are clear but so are the risks. Some care has to be taken. The scope for retribution against breaches is much more limited for those living in other countries.

63. How can researchers access data sets from other countries? How can international agencies obtain access to microdata for statistical and research purposes?

(i) Public Use Files where they exist,

(ii) Remote Access Facilities with appropriate safeguards, or

(iii) Anonymised Microdata Files.

64. Public Use Files are only available for some countries. We would argue that (iii), even when enabled by the legislation of the NSO, should only be an option where the NSO of the home country of the researcher or the international institution has adequate legislation to protect the confidentiality of the microdata. The data could then be released through the NSO of the country of the researcher. The NSO "owning" the microdata still has the choice as to whether it feels sufficiently comfortable to release their AMFs under such an arrangement. An exception may be made for Eurostat where specific legislation has been established to protect the confidentiality of microdata provided by member countries.

65. The use of Remote Access Facilities may be the preferred route to provide access to international researchers. There are more controls and position of NSOs, on access to microdata, is more easily defended if challenged. However, the usability of these arrangements needs to be tested.

SHARING OF MICRODATA BETWEEN STATISTICAL AGENCIES

66. In some countries, particularly those with decentralised systems, it may be a legitimate activity to share microdata between statistical agencies to support statistical work. For example, this may be done to reduce cost and reporting burden through duplicated activities. Nevertheless, this should be undertaken in accordance with agreed protocols to maintain public confidence in the appropriateness of these activities. The four Principles announced in Section 13 would also be appropriate for these types of activities, particularly Principles 2 and 4.

ISSUES FOR DISCUSSION

67. The underlying premise in this paper is that there are two public goods.

- The public good of confidentiality protection because it is a cornerstone of a viable national statistical system;

- The public good of researcher access that provides for research findings that will benefit public policy, government programs, etc.

68. The majority of NSOs have taken a very cautious approach on confidentiality to the extent of virtually avoiding all risks. Developments in technology, and the increasing availability of public and private data bases on individuals, suggest we take an even more cautious approach to avoid the release of unidentifiable data. However, the general feeling of the 2003 CES suggested that this may not be the sensible approach to take if you balance these two public goods. As NSOs should not go beyond their legal and other obligations, this may require countries to change their legal and other arrangements for providing access to microdata in order to provide a more appropriate balance. Is it accepted that, in making arrangements for access to microdata, these two public goods need to be balanced?

69. Section 6 of the paper discusses how the tensions between the NSO and Researcher perspectives might be resolved? Is this supported?

70. Section 11 suggests that the access arrangements should be different for data about businesses but the underlying principles remain the same. Is that supported?

71. Some draft principles are outlined in Section 13. Are they supported? Should they be extended?

72. Section 15 outlines some preliminary thoughts on access by international researchers. Any comments?

73. The paper suggests Remote Access Facilities of various forms have an increasingly important role. Is this generally supported?

74. Are the proposed principles appropriate for linked data sets? Should any additional principles apply?

75. Some issues have been identified where there are differences of view between Task Force members or other commentators.

(a) Should the provision of Public Use Files be discouraged unless there is some form of undertaking by the person accessing the file? A change in past practice might be justified by the extreme difficulty in producing Public Use Files where only a small number of subjects are identifiable through Statistical Matching exercises? (The identification risk through Statistical Matching exercises does not seem to be well understood.)

(b) Apart from Public Use Files should equality of access be a principle for the

provision of microdata? Alternatively, should discretion be provided to the NSO?

(c) Should the use of microdata files for non-statistical purposes be banned?

(d) If public good is the main reason for providing microdata services how important is it that research based on microdata files be put in the public domain? Public Use Files would be an exception.

76. One element that is missing from the paper is the "respondent perspective". We would be interested in the results of any studies undertaken on this aspect. In my own country, we have undertaken some focus group studies. They confirm that confidentiality is important to their participation in the survey. However, if there are strong public benefits they are prepared for their data to be used for research purposes (where there are clear public benefits) if appropriate safeguards are put in place.

ATTACHMENT A

TASK FORCE ON CONFIDENTIALITY AND MICRODATA - TERMS OF REFERENCE

(As revised following discussions with the Bureau of the Conference of European Statisticians.)

Background

So far confidentiality protection has been mainly a national issue. However, in the context of EU and increasing data dissemination over Internet, it is becoming also an international issue. There is a lot of international collaboration among the research community, and the researchers can be very critical towards different access rules in different countries. Often researchers are not allowed to access other countries microdata because of the fear that confidentiality protection cannot be guaranteed but cross country comparisons can be a very important part of a research project. This raises the need to unify approaches internationally, and to agree on some core principles for dissemination of microdata. It should be seen in the context that the 2003 Conference of European Statisticians saw support for research as an important activity of the National Statistical Offices.

Task

(i) To produce a draft set of (a) principles and (b) associated guidelines that are general enough to be able to be adapted by the countries that participate in the Conference of European Statisticians.

(ii) To submit a draft for the consideration of the 2004 Conference of European Statisticians.

(iii) If accepted, compile a set of current best practices to guide implementation.

The United Nations Fundamental Principles of Official Statistics should be used as a guide on length and style for the presentation of the draft principles and guidelines.

Task Force Membership

The Task Force membership has been chosen to represent the varying interests of the members of the Conference of European Statisticians. The membership will be Dennis Trewin (Australia), Ivan Fellegi (Canada), Otto Andersen (Denmark), Teimuraz Beridze (Georgia), Luigi Biggeri (Italy) and Tadeusz Toczynski (Poland). Mr Trewin will act as Chairman of the Task Force.

Modus Operandi

During the development of the draft principles and guidelines the Task Force will consult with member countries, representatives of the research community, and selected bodies concerned with privacy issues. This does not preclude the Task Force from collaborating others who may have an interest in the topic.

* * *